

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Post-transcriptional regulation of pluripotency by microRNAs and RNA binding proteins

Permalink

<https://escholarship.org/uc/item/8p1505kj>

Author

Ye, Julia

Publication Date

2017

Peer reviewed|Thesis/dissertation

Post-transcriptional regulation of pluripotency by microRNAs and
RNA binding proteins

by

Julia Ye

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

Dedication

For my family

Acknowledgements

I am thankful for the support and guidance of the many people who have helped make this work possible. First, thank you to my adviser, Robert Blelloch, for taking me into your lab and always challenging me to think creatively, work harder and smarter, and be fearless in all of my scientific pursuits. You are the model of the scientist I hope to one day become.

Many thanks to my thesis committee, Noelle L'Etoile, Mark Ansel, and Hiten Madhani, for helping me grow not only professionally but also personally. I deeply appreciate the practical guidance you gave me during my experimental struggles, as well as the generous emotional support you showed me through the process. I feel so blessed to be able to call you my mentors.

I was fortunate to work with many talented people, including Rob Judson and Ron Parchem, who taught me everything I know and whose mentorship and friendship I cherish deeply. Also, thank you, Hu Jin, for being the best collaborator anyone could ever hope to have.

The BMS and MSTP administrators have impacted my life here enormously. Thank you, Jana Toutolmin, Geri Ehle, Catherine Norton, Demian Sainz, Lisa Magargal, Nathan Jew, Monique Piazza, and Amanda Andonian for ensuring that all I ever had to worry about was my research and for giving me advice, perspective, and guidance in challenging situations.

Beyond the lab, I owe so much to Raga Ramachandran for helping me develop my interests in medicine. I cannot thank Richard Shapiro enough for the opportunity to build the UCSF-Lowell Science Research Program—you helped me discover my passion for teaching, and it has been such a joy to evolve from being your student to your collaborator to your friend.

Finally, my family and friends have been a deep well of support. I am forever indebted to my parents, Jianqin Ye and Hua Su, for giving me the courage and strength to always strive to achieve my goals and to my partner, Ethan Crumlin, for his constant love and encouragement.

Contributions

The research presented here was conducted under the supervision of Dr. Robert Blelloch at the Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research at the University of California, San Francisco. This work was supported by fellowships from the Eunice Kennedy Shriver National Institute of Child Health and Human Development at the National Institutes of Health (grant number 1F30HD084120) and the UCSF Discovery Fellows Program. All of the writing in this manuscript is my own work.

Chapter 2 describes a work published in *Cell Stem Cell* in 2014 with the following authors: Ronald J. Parchem, Julia Ye, Robert L. Judson, Marie F. LaRussa, Raga Krishnakumar, Amy Blelloch, Michael C. Oldham, and Robert Blelloch. Ron Parchem created the dual reporter system and led all experiments. Robert Judson and I performed and analyzed the reprogramming experiments. Marie LaRussa performed the ESC characterizations, Raga Krishnakumar and Michael Oldham conducted the bioinformatics analyses, and Amy Blelloch generated the chimeric mice. Robert Blelloch supervised the research. Together, we wrote the manuscript.

Chapter 3 describes a work in preparation for submission to *Stem Cell Reports* with the following authors: Julia Ye, Ronald J. Parchem, and Robert Blelloch. I designed, performed, and analyzed the experiments. Robert Blelloch and Ronald Parchem supervised the research, and together, we wrote the manuscript.

Chapter 4 describes a work recently submitted to *Nucleic Acids Research* with the following authors: Julia Ye, Hu Jin, Aleksandr Pankov, Jun S. Song, and Robert Blelloch. I designed, performed, and analyzed the experiments. Hu Jin and Aleksandr Pankov conducted the bioinformatics analyses. Jun Song and Robert Blelloch supervised the research. Together, we wrote the manuscript.

Post-transcriptional regulation of pluripotency by microRNAs and RNA binding proteins

Julia Ye

Early mammalian development is driven by precise molecular changes that enable cells to establish, maintain, and exit from the pluripotent state, which is characterized by ability to generate all of the cells that comprise the adult mammal. Embryonic stem cells (ESCs) can be used to model the developmental progression through which pluripotent cells become differentiated somatic cells, and remarkably, ESC-like cells called induced pluripotent stem cells (iPSCs) can be generated from somatic cells like fibroblasts through a process called reprogramming or de-differentiation. Differentiation and de-differentiation are both highly regulated by post-transcriptional mechanisms mediated in part by non-coding RNAs and RNA binding proteins (RBPs). Here, we investigated the roles of microRNAs (miRNAs) and RBPs in the regulation of pluripotency.

Using a fluorescent reporter system, we found that the miR-290 and miR-302 miRNA families are expressed in that order during embryonic development and ESC differentiation but that the loci are transcriptionally activated in a stochastic manner during reprogramming. Addition of the reprogramming enhancer Sall4, however, promoted the transcription of the miR-302 locus before that of miR-290, suggesting that reprogramming does not necessarily represent development in reverse and that the sequence of regulatory changes that occur during the process depends on the specific cocktail of reprogramming factors used. Surprisingly, we found that, despite their expression in pluripotent cells, miR-290 and miR-302 are also both dispensable for the generation of iPSCs.

In a separate set of studies investigating the role of RBPs in pluripotency, we performed an RNAi screen in an ESC differentiation system that allowed us to uncover NF45 and

NF90/NF110 as factors important for maintaining pluripotency. Our molecular, cellular, and genomic analyses of NF45 and NF90/NF110 knockdown and knockout ESCs suggested that NF45 and NF90/NF110 regulate the expression of one another and function both independently and in complexes to promote ESC proliferation and proper differentiation.

The studies described here demonstrate the importance and complexity of miRNA and RBP control over the pluripotent state. Developing a more nuanced understanding of these post-transcriptional regulatory processes will enable the more effective and informed use of pluripotent cells to address clinical problems in the future.

Table of contents

Chapter 1. Introduction	1
Cellular potency	1
Pluripotent stem cells	3
Induced pluripotency	4
Post-transcriptional regulation of pluripotency	6
The role of microRNAs	7
The role of RNA binding proteins	10
Premise of studies	13
Chapter 2. Two miRNA clusters reveal alternative paths in late-stage reprogramming	14
Summary	14
Introduction	14
Results	15
The <i>mir-290</i> and <i>mir-302</i> clusters define transient and distinct states of pluripotency	15
The <i>mir-290</i> and <i>mir-302</i> clusters are stochastically activated in OSK-driven reprogramming	17
Sall4 enhances reprogramming and biases the path taken to the iPSC state	21
Discussion	26
Materials and Methods	28
Chapter 3. miR-290 and miR-302 are dispensable for somatic cell reprogramming	34
Summary	34
Introduction	34
Results	35
Expression of miR-302 is not required for reprogramming	35
Expression of miR-290 is not required for reprogramming	38
Depletion of miR-302 in <i>mir-290</i> KO cells does not affect reprogramming	41
Discussion	49
Materials and Methods	53
Chapter 4. NF45 and NF90/NF110 coordinately regulate ESC pluripotency and differentiation	60
Summary	60
Introduction	60
Results	62
RNAi screen identifies NF45, NF90, and NF110 as promoters of pluripotency	62
NF45 and NF90 regulate ESC proliferation and differentiation	65
NF45 and NF90/NF110 form complexes and influence expression of each other	71
Genome-wide expression analysis support functional interactions between NF45 and NF90/NF110	74

Identification of NF90/NF110 RNA targets	76
Discussion	80
Materials and Methods	84
Chapter 5. Concluding remarks and future directions	98
Are miR-290 and miR-302 absolutely dispensable in reprogramming?	98
A binding motif for NF90/NF110?	100
NF45 and NF90/NF110 interactions with the larger regulatory network?	101
Lessons for the future	102
References	104

List of Figures

Figure 1. Cellular potency through development	2
Figure 2. RBPs involved in pluripotency act at many different regulatory steps	11
Figure 3. The <i>mir-290</i> and <i>mir-302</i> clusters define sequential stages of pluripotency <i>in vivo</i> and <i>in vitro</i>	16
Figure 4. OSK reprogramming leads to late-stage stochastic activation of <i>mir-290</i> and <i>mir-302</i>	18
Figure 5. <i>mir-290</i> -mCherry+/m <i>ir-302</i> -eGFP+ double positive cells show the greatest potential for generating iPSC colonies	20
Figure 6. <i>Sall4</i> enhances reprogramming and biases intermediates to activate <i>mir-302</i>	21
Figure 7. <i>Sall4</i> biases the path taken during reprogramming	23
Figure 8. High iPSC potential is associated with a naïve pluripotent gene signature	25
Figure 9. <i>mir-302</i> KO MEFs can reprogram to iPSCs	36
Figure 10. <i>mir-302</i> KO iPSCs have no detectable molecular or functional deficits	37
Figure 11. <i>mir-290</i> KO MEFs can reprogram to iPSCs	38
Figure 12. <i>mir-290</i> KO iPSCs exhibit minor deficits	39
Figure 13. Construction and use of <i>eGFP-miR-302-sponge</i> construct	42
Figure 14. Reducing miR-302 levels in the context of miR-290 loss does not affect somatic cell reprogramming	43
Figure 15. <i>eGFP-miR-302-sponge</i> -treated <i>mir-290</i> KO iPSCs do not exhibit additional detectable deficits	44
Figure 16. eGFP expression levels in iPSCs is consistent with miR-302 sponge function	46
Figure 17. The <i>eGFP-miR-302-sponge</i> construct suppresses miR-290 and miR-302 family miRNAs	48
Figure 18. An siRNA screen identifies NF45, NF90, and NF110 as promoters of pluripotency	64

Figure 19. Validation of siRNA screen	66
Figure 20. NF45 and NF90 promote ESC proliferation	67
Figure 21. Loss of NF45 and NF90/NF100 dysregulate differentiation to EpiCs	68
Figure 22. Loss of NF45 and NF90/NF110 dysregulate differentiation down embryonic lineages	70
Figure 23. NF45 and NF90/NF110 are largely localized to the nucleus and physically interact	71
Figure 24. NF45 and NF90/NF110 influence the expression of each other at different levels of regulation	72
Figure 25. Global expression analysis is consistent with functional interactions among NF45, NF90, and NF110	75
Figure 26. Transcriptomic changes in NF45 KO correlate with NF90+NF110 KO but not NF110 KO	76
Figure 27. Identification of NF90/NF110 RNA targets	79
Figure 28. Models of NF45, NF90, and NF110 regulation and function	82

List of Tables

Table 1. Mature miRNA sequences of the four miRNA clusters most highly expressed in pluripotent stem cells	9
--	---

Chapter 1—Introduction

Cellular potency

Mammalian development involves the highly coordinated growth, division, and organization of cells that are increasingly restricted in their potential to form the many tissue types of the adult organism as they become committed to specific cell lineages through a process called differentiation. Consequently, cells obtained from different stages of development have varying degrees of potency, or ability to differentiate into other cell types (Figure 1).

Embryogenesis is initiated when a sperm and an egg join to form a fertilized zygote. The zygote has the potential to form all the cell types of the embryo proper as well as the extraembryonic tissues that comprise the yolk sac and placenta; thus, it is considered totipotent. Through a defined and stereotyped set of cell divisions, which, in mice, occurs by approximately embryonic day 3.5 (E3.5), the zygote morphs into a structure called a blastocyst, which contains a compacted group of cells called the inner cell mass (ICM). Some cells in the ICM then go on to form the epiblast, a structure that exists from about E4.5 until E6.5. Epiblast cells are considered pluripotent, because they are capable of differentiating into the three embryonic germ layers comprising the embryo proper. This differentiation process occurs at gastrulation, which begins at E6-E6.5 with the formation of the primitive streak. From these three germ layers arise the tissues and organs of the fully-formed organism: specifically, ectoderm gives rise to the nervous system, neural crest, and epidermis; mesoderm generates connective tissues like bone, muscle, blood, and blood vessels; and endoderm forms the digestive system and respiratory system (Nagy et al., 2003). Although these adult tissues are largely composed of terminally differentiated cells that are committed to their cellular identity and do not divide, they also possess small cellular populations that are capable of replacing aged or damaged cells by

differentiating into cells specific to their particular tissue lineage of origin. These multipotent progenitor or stem cells have been especially well studied in tissues that are characterized by rapid turnover, like those of the GI tract, skin, and hematopoietic system (Barker, et al., 2010; Eaves, 2015).

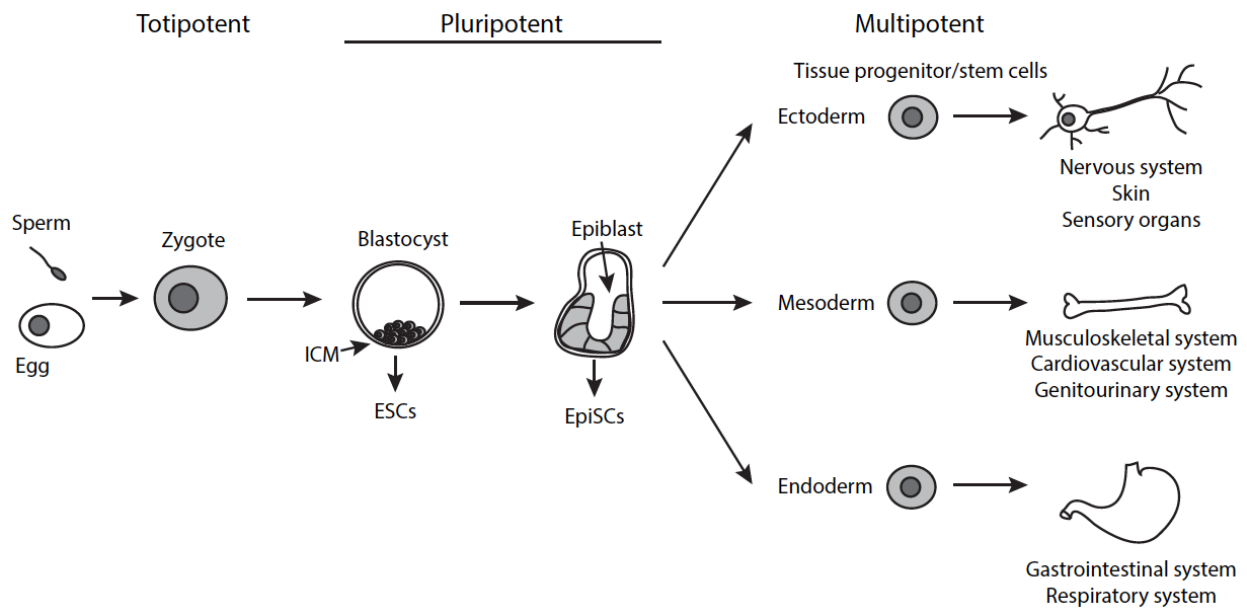


Figure 1. Cellular potency through development. The potential of cells to produce different cells in the body decreases over developmental time as cells differentiate to become more specialized. Pluripotent stem cells such as ESCs and EpiSCs can be derived from the early embryo at the blastocyst and epiblast stages.

The work reported in this dissertation focuses on the pluripotent state. Pluripotency is a property not just of cells in the early mammalian embryo but also of stem cell populations derived from these embryonic cells. As described above, pluripotency *in vivo* is a transient developmental characteristic that is lost as cells in the embryo become more specialized. *In vitro*, however, pluripotent stem cells can be maintained for extended periods of time under specific culture conditions and can even be artificially produced from differentiated cells through experimental manipulations capable of re-establishing the epigenetic status of embryonic cells (reviewed in Rodolfa and Eggan, 2006). In the paragraphs below, we will discuss the nature of

pluripotent stem cells and their regulation by two classes of molecules called microRNAs (miRNAs) and RNA binding proteins (RBPs).

Pluripotent stem cells

Stem cells are characterized by their ability to self-renew (divide continuously while maintaining the same cellular state) and differentiate into more specialized cell types. As mentioned above and diagrammed in Figure 1, they can be derived from multiple stages of embryonic development and can also be found resident in adult tissues. Pluripotent stem cells are cells that are extracted from the developing mouse embryo from E3.5 to as late as E8.0 (Osorno et al., 2012) and grown in culture *in vitro*. Embryonic stem cells (ESCs) and epiblast stem cells (EpiSCs) are the two main types of pluripotent stem cells, and they represent different degrees of embryonic pluripotency (reviewed in Nichols and Smith, 2009; Weinberger et al., 2016), as they possess distinct transcriptional, epigenetic, metabolic, signaling, and functional characteristics (reviewed in De Los Angeles et al., 2015).

Although mouse ESCs are derived from the E3.5 ICM (Evans and Kaufman, 1981, Martin, 1981), molecular studies have demonstrated that they most closely represent the cells of the ~E4.5 pre-implantation epiblast (Boroviak et al., 2014). Some of the characteristics that define ESCs include the following: the high expression of pluripotency markers like Oct4, Nanog, Sox2, and Klf4; stable propagation in culture in the presence of leukemia inhibitory factor (LIF), which activates Stat3 signaling, as well as Mek and Erk inhibitors (2i); activated X chromosomes; and the ability to contribute to teratomas and chimeras. For their ability to incorporate into recipient blastocysts and contribute to all the tissues of resulting chimeric animals, ESCs are often described as representing “ground state” or “naïve” pluripotency.

In contrast, mouse EpiSCs are derived from the E5.5 post-implantation epiblast (Tesar et al., 2007; Brons et al., 2007) and are described as being “primed” pluripotent. EpiSCs express a different set of marker genes compared to ESCs: although Oct4 and Sox2 are expressed as in ESCs, Nanog and Klf4 are low, and certain early differentiation markers like Fgf5 and T/Brachyury are elevated. Moreover, their growth in culture relies on the presence of Fgf2 and Activin A, they die when treated with 2i (Guo et al., 2009), and they have undergone X chromosome inactivation. Although they can contribute to all three germ layers in teratoma assays, EpiSCs cannot contribute to blastocyst chimeras (Guo et al., 2009; Tesar et al., 2007), suggesting that they represent a developmental and functional state different from that of ESCs. Interestingly, it has been shown that mouse EpiSCs resemble human ESCs in morphology, culture requirements, gene expression, and functional characteristics (Brons et al., 2007), which demonstrates that the nature of the pluripotent state also varies among species.

Because of their functional parallels with embryonic cells, pluripotent stem cells serve as a biologically relevant and experimentally tractable *in vitro* platform not only for studying early embryogenesis but also for developing advances in disease modeling, drug discovery, and tissue regeneration.

Induced pluripotency

Historically, stem cell research faced major challenges—not just technical but also ethical—because for a long time, acquiring pluripotent stem cells necessarily involved the destruction of developing embryos. Human stem cell research was especially controversial for this reason. The demonstration that ESC-like “induced” pluripotent stem cells (iPSCs) could be generated from differentiated fibroblasts through the simple overexpression of some key

pluripotency factors was therefore a landmark event for the field (Takahashi and Yamanaka, 2006; Takahashi and Yamanaka, 2007). These scientists showed that by using retroviruses to introduce Oct4, Sox2, Klf4, and cMyc (the “Yamanaka factors”) into mouse embryonic fibroblasts (MEFs) or human foreskin fibroblasts (HFFs), they could alter the epigenetic marks and gene expression patterns in these differentiated somatic cells such that they became virtually identical to those of mouse or human ESCs, respectively. Since then, advances in somatic cell reprogramming, also called directed de-differentiation, have come at lightning pace. Other transcription factor cocktails have been shown to also be capable of producing iPSCs (Yu et al., 2007; Nakagawa et al., 2008; Buganim, et al., 2014), and a number of new techniques have been developed to avoid the use of integrating retro- and lentiviruses that could lead to cancer-causing genetic mutations in the resulting iPSCs (reviewed in Malik and Rao, 2013).

Although others had previously demonstrated the ability of the epigenome to be reset through various techniques such as somatic cell nuclear transfer (SCNT), which was first performed in *Xenopus* (Gurdon, 1962), and somatic cell-ESC fusion (Cowan et al., 2005; Tada et al., 2001), this method of ectopically expressing pluripotency transcription factors was groundbreaking. Somatic cell reprogramming not only allowed researchers to completely sidestep ethical concerns surrounding human stem cell research, but also dramatically increased the impact of the stem cell field. Now pluripotent cells can be generated from skin cells non-invasively acquired from human patients, which allows scientists and clinicians to study the specific patterns of disease manifestation in individual patients and design therapies uniquely targeted to those patients (reviewed in Bellin et al., 2012; Takahashi and Yamanaka, 2012).

Post-transcriptional regulation of pluripotency

As can be surmised from the above discussion of developmental pluripotency and induced pluripotency, the pluripotent state is highly dynamic and requires the coordinated action of numerous molecular factors operating at every regulatory layer. The transcriptional, signaling, and epigenetic control of pluripotent stem cells has been characterized in great depth and detail at this point (reviewed in Young, 2011; Ng and Surani, 2011; Watanabe et al., 2012). A group of transcription factors, which include Oct4, Sox2, and Nanog, forms the core of the pluripotency network, as they, in concert with various other transcription factors and cofactors, activate the expression of genes important for ESC identity and repress the expression of lineage-specific genes. Oct4, Sox2, and Nanog also bind to the promoters of one another, thus forming an auto-regulatory, self-reinforcing positive feedback loop. Augmenting the activities of these core transcription factors are the signaling pathways that are triggered by extracellular molecules like LIF, Wnt, and BMP, which activate transcription factors (Stat3, Tcf3, and Smad1, respectively) that co-occupy Oct4, Sox2, and Nanog-bound promoters and enhancers to add nuance to the gene expression program. Finally, a number of chromatin remodelers are responsible for establishing and maintaining the epigenetic landscape of pluripotent stem cells, which features global hypomethylation, active X chromosomes, and “bivalent” histone modifications that poise differentiation genes for expression upon the proper extracellular cues.

In addition to the regulatory mechanisms described above, post-transcriptional and translational processes also play integral roles in determining the fate of pluripotent cells. These processes cover a whole host of activities that affect the life trajectory of an RNA molecule, including capping, splicing, cleavage, nontemplated nucleotide addition, nucleotide editing, nuclear export, intracellular localization, stability, and translation (Keene, 2007). Remarkably,

post-transcriptional regulation has been reported to account for nearly 75% of the changes in protein levels after Nanog knockdown-induced differentiation (Lu et al., 2009), and control over translational initiation has been shown to dramatically influence the efficiency of somatic cell reprogramming (Tahmasebi et al., 2013). RNA binding proteins (RBPs) and non-coding RNAs, which include long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), are the primary mediators of these post-transcriptional and translational processes. While non-coding RNAs have recently received quite a bit of attention in ESC differentiation and somatic cell reprogramming, (reviewed in Greve et al., 2013; Ghosal et al., 2013; Wright and Ciosk, 2013; Flynn and Chang, 2014), RBPs have been highly under-studied, despite their obvious importance in pluripotency. Gaining a more complete understanding of these mechanisms of gene expression control will allow ESCs and iPSCs to be more effectively utilized at the bench and the bedside.

The role of miRNAs

MicroRNAs (miRNAs) are short 18-24nt noncoding RNAs that bind to complementary sites in mRNAs, typically in the 3'UTR (Lewis et al., 2005), and, in collaboration with the RNA-induced silencing complex (RISC), cause mRNA destabilization or translation inhibition (Bartel 2009; Fabian and Sonenberg, 2012). The outcome of miRNA binding depends on the degree of target site complementarity: perfectly complementary binding results in endonucleolytic cleavage of the mRNA, while partially complementary binding results in mRNA deadenylation or repression of the translational machinery (reviewed in Filipowicz et al., 2008). The seed sequence, a 7-mer sequence that constitutes the 2-8nts of the mature miRNA, is the primary determinant of target recognition (Lewis et al., 2003). MicroRNAs with similar seed sequences constitute a miRNA “family” and are thought to target overlapping sets of transcripts; miRNAs

expressed from the same genomic locus constitute a “cluster” and may contain individual miRNAs from different families (reviewed in Greve et al., 2013).

MiRNAs are critical in pluripotency and early mammalian development, as deletion of *Dgcr8* or *Dicer*, two genes required for miRNA biogenesis, results in major ESC differentiation defects and embryonic lethality (Bernstein et al., 2003; Kanellopoulou et al., 2005; Wang et al., 2007). Moreover, inducible deletion of *Dgcr8* or *Dicer* in MEFs has been found to either severely decrease or completely inhibit reprogramming to iPSCs (Kim et al., 2012; Liu et al., 2015).

While many somatic cells express a variety of different miRNAs, pluripotent stem cells have a distinct miRNA profile dominated by a handful of miRNA clusters: *mir-290~295* (also known as *mir-371~373* in humans [Houbaviy et al., 2005]), *mir-302~367*, *mir-17~92*, and *mir-106b~25* (reviewed in Vidigal and Ventura, 2012; Greve et al., 2013; Table 1). More specifically, *mir-290~295* miRNAs are highly expressed in naïve mouse ESCs and primordial germ cells (Houbaviy, et al., 2003; Marson et al., 2008), while *mir-302~367* miRNAs are highly expressed in primed mouse EpiSCs and human ESCs (Suh et al., 2004; Bar et al., 2008; Jouneau et al., 2012). The *mir-290~295* and *mir-302~367* clusters, in particular, are intimately integrated within the pluripotency network, as their promoters are bound by the core pluripotency transcription factors, Oct4, Sox2, Nanog, and Tcf3 (Card et al., 2008; Marson et al., 2008).

Interestingly, many of the miRNAs in the *mir-290~295*, *mir-302~367*, *mir-17~92*, and *mir-106b~25* clusters have closely related seed sequences. Indeed, these miRNAs are sometimes referred to as the embryonic stem cell-specific cell cycle-regulating (ESCC) family of miRNAs for their role in promoting the structure of the ESC cell cycle (Wang et al., 2008). The ESCC miRNAs also promote ESC self-renewal while buffering against differentiation cues (Melton et

al., 2010) and have been found to be strong promoters of somatic cell reprogramming. Overexpressing ESCC miRNAs can enhance reprogramming by the Yamanaka factors in both mouse and human somatic cells (Judson et al., 2009; Liao et al., 2011; Subramanyam et al., 2011), and, in fact, overexpression of the *mir-302~367* cluster alone, in the absence of other factors, has been shown to induce mouse and human reprogramming (Anokye-Danso et al., 2011; Lin et al., 2011). These remarkably powerful effects of the ESCCs on iPSC generation occur through a number of different mechanisms (Subramanyam et al., 2011), including those that regulate the mesenchymal-to-epithelial transition (Li et al., 2010; Samavarchi-Tehrani et al., 2010), cell cycle, epigenetic modifiers, endoplasmic reticulum trafficking, cellular metabolism, and more.

Table 1. Mature miRNA sequences of the four miRNA clusters most highly expressed in pluripotent stem cells. The seed sequence is underlined. Red indicates the ESCC seed sequence.

miRNA cluster	Mature miRNA	Mature miRNA sequence
miR-290~295	miR-290 miR-291a miR-292 miR-291b miR-293 miR-294 miR-295	A <u>AAGUGC</u> CGCCUAGUUUUAAGCCC A <u>AAGUGC</u> UCCACUUUGUGUGC A <u>AAGUGC</u> CGCCAGGUUUUGAGUGU A <u>AAGUGC</u> AUCCAUUUUGUUUGU AGUGCCGCAGAGUUUGUAGUGU A <u>AAGUGC</u> UCCCCUUUUGUGUGU A <u>AAGUGC</u> CGCCUAGUUUUAAGCCC
miR-17~92	miR-17 miR-18a miR-19a miR-20a miR-19b-1 miR-92a-1	CA <u>AAGUGC</u> UUACAGUGCAGGUAG UAAGGUGCAUCUAGUGCAGAUAG UGUGCAA <u>AUCU</u> AUGCAAACUGA UA <u>AAGUGC</u> UUUAUAGUGCAGGUAG UGUGCAA <u>AUCA</u> UGCAAACUGA UAUUGCACUUGUCCCGGCCUG
miR-302~367	miR-302b miR-302c miR-302a miR-302d miR-367	U <u>AAGUGC</u> UCCAUGUUUUAAGUAG <u>AAGUGC</u> UCCAUGUUUCAGUGG U <u>AAGUGC</u> UCCAUGUUUUGGUGA U <u>AAGUGC</u> UCCAUGUUUGAGUGU AAUUGCACUUUAGCAAUGGUGA
miR-106b~25	miR-106b miR-93 miR-25	UA <u>AAGUGC</u> UGACAGUGCAGAU CA <u>AAGUGC</u> UGUUCGUGCAGGUAG CAUUGCACUUGUCUCGGUCUGA

The role of RNA binding proteins

RNA binding proteins (RBPs) are the chief executors of post-transcriptional and translational regulatory events. Indeed, even miRNAs rely on RBPs for their biogenesis and activity on target mRNAs. However, little is known about RBPs overall: most are classified based on computationally predicted similarities to proteins with known RNA binding domains, and until recently, few of these predictions have been verified in a cellular context *in vivo*. The recent introduction of a technique termed “mRNA interactome capture,” which enables the identification of proteins bound to polyadenylated RNAs *in vivo*, has been a significant development for the field (Baltz et al., 2012; Castello et al., 2012). Using this method, several groups were able to create a comprehensive catalogue of RBPs in different mammalian cells, including 555 RBPs in mouse ESCs (Kwon et al., 2013). However, the mechanism of action of only a small number of RBPs has been examined in any great detail in the context of pluripotency; most studies associate disruption of RBP expression with some aspect of ESC dysfunction while inferring mechanism of action from work done in cell-free biochemical assays or unrelated cell lines (reviewed in Ye and Blelloch, 2014).

In pluripotent cells, RBPs have been found to collaborate with and modulate the activities of ubiquitous core cellular machineries (reviewed in Ye and Blelloch, 2014, Figure 2). For example, RBPs like FOX2, SON, SFRS2, MBNL1 and MBNL2 work with the proteins and small nuclear RNAs of the spliceosome to generate the ESC-specific splicing signature. As another example, RBPs like Nat1, Rbm35a, and Ptbp1 bind to the 5'UTR of RNA transcripts, which can help recruit core translation initiation factors, adjust the accessibility of the RNA to ribosomes, create ribonucleoprotein structures conducive for cap-independent and internal

ribosome entry site-mediated translation, and regulate the movement of ribosomes along the transcript.

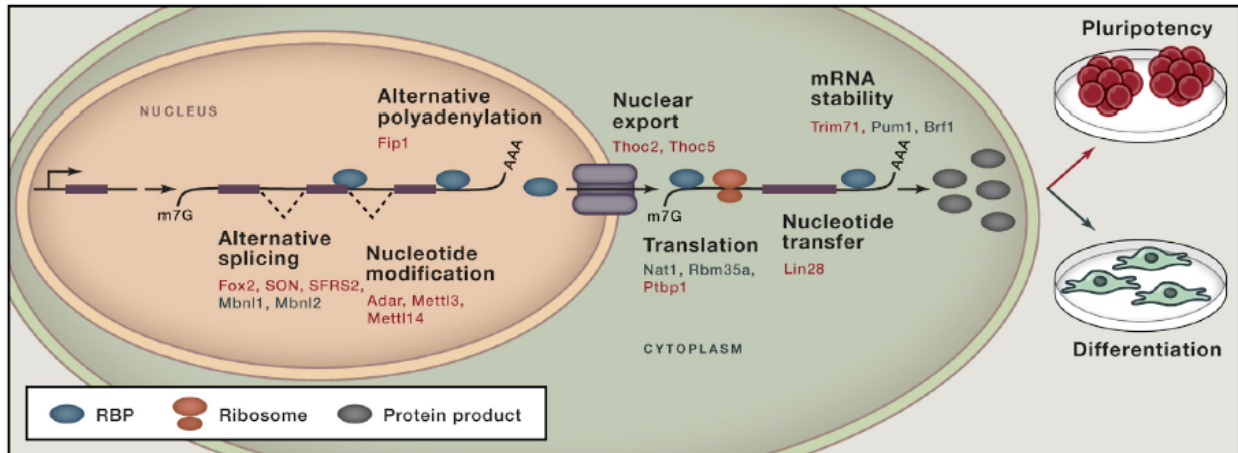


Figure 2. RBPs involved in pluripotency act at many different regulatory steps. Summary of RBPs and the events they regulate in the maintenance of and exit from pluripotency. Starting in the nucleus, RBPs regulate splicing (FOX2, SON, SFRS2, MBNL1, and MBNL2) and alternative polyadenylation (FIP1) simultaneously with transcription. RBPs then regulate export of transcripts (THOC2 and THOC5). RBPs also can induce modifications to RNAs including nucleotide changes (ADAR, METTL3, and METTL14 in the nucleus) and nucleotidyl transfer (LIN28A in association with the TUTases ZCCHC6 and ZCCHC11 in the cytoplasm), which in turn influence mRNA stability and translation. In the cytoplasm, the binding of RBPs to the 3'UTRs of transcripts directly regulates mRNA stability and translation (TRIM71, PUM1, and BRF1). Translation is also influenced by RBPs that bind the 5'UTR of transcripts (NAT1, RBM35A, and PTBP1). Blue circles indicate RBPs. RBP genes in red are positive regulators of pluripotency. RBP genes in green are negative regulators of pluripotency. Gray circles indicate the protein products of the genes whose expression levels are affected by RBPs.

Intriguingly, some RBPs have been found to be involved in multiple regulatory complexes controlling multiple aspects of RNA metabolism. Trim71 is a particularly noteworthy example for the extent of its interconnection with several intracellular circuits. The ability of the ESCC miRNAs to set the ESC cell cycle relies on Trim71 co-binding the 3'UTR of Cdkn1a and inhibiting its translation (Chang et al., 2012). Moreover, Trim71 is a target of the pro-differentiation miRNA let-7, whose biogenesis is inhibited by the RBP Lin28a. Adding to the complexity, Trim71 has also been found to be an E3 ubiquitin ligase for Ago2 and cooperates with Lin28a to inhibit let-7 activity (Rybak et al., 2009). Trim71 can also associate directly with some mRNAs at their UTRs independent of miRNA involvement, and this interaction drives

transcript degradation and inhibits translation (Loedige et al., 2013; Worringer et al., 2014) depending, in part, on the location of the Trim71 binding sites (Aeschmann et al., 2017).

Lin28a is another RBP proposed to play pleiotropic roles. While it is most famous for regulating let-7 biogenesis by directing the uridylation of pre-let-7 by the 3' terminal uridylyl transferases, Zcchc11 and Zcchc6 (Heo et al., 2008; Hagan et al., 2009; Heo et al., 2012), Lin28a also has functions independent of let-7 (reviewed in Shyh-Chang and Daley, 2013). Through binding to an AG-rich motif similar to its pre-let-7 binding motif, Lin28a both positively and negatively modulates the translation of hundreds of mRNAs in human and mouse ESCs (Wilbert et al., 2012; Cho et al., 2012; Peng et al., 2011; Xu et al., 2009).

RBPs like Trim71 and Lin28a may appear to be multifunctional for several reasons: they could have different roles in different cell types, multiple roles in a single cell type, or some mix of the two. Regardless, it is almost certain that the particular combination of targets and cofactors that an RBP encounters influences its functions in context-specific ways—a notion that expands upon the “RNA regulon” model originally proposed by Jack Keene, in which an RBP binds multiple targets to effect changes in various cellular processes (Keene, 2007). In other words, it is possible that instead of regulating multiple targets through just a single mechanism, one RBP could simultaneously participate in several layers of RNA metabolism. In so doing, the RBP becomes part of increasingly complex regulatory modules, with many opportunities for feedback and crosstalk, where one aspect of metabolism such as translation can be linked to another such as splicing. Consequently, perturbation of the RBP could have a cascading effect on the molecular landscape of a cell and precipitate drastic switches in cellular identity. Also, it follows that disrupting subsets of the mRNA targets or cofactors of any one RBP could affect parallel pathways by shifting the RBP's dominant activities to different genes or even different

regulatory modules altogether. For example, a decrease in the levels of a cofactor that enables an RBP to regulate splicing could drive that RBP to shift its main activity to transcription.

Premise of studies

RBP- and miRNA-mediated post-transcriptional regulatory processes are clearly critical to pluripotency. In the following chapters, we discuss our contributions to this field. In chapters 2 and 3, we characterize the roles of *mir-290~295* and *mir-302~367* clusters in somatic cell reprogramming, and in chapter 4, we explore the functions of the RBPs NF45 and NF90/NF110 in ESCs.

Chapter 2—Two miRNA clusters reveal alternative paths in late-stage reprogramming

Summary

Ectopic expression of specific factors such as Oct4, Sox2, and Klf4 (OSK) is sufficient to reprogram somatic cells into induced pluripotent stem cells (iPSCs). In this study, we examine the paths taken by cells during the reprogramming process by following the transcriptional activation of two pluripotent miRNA clusters, *mir-290~295* (*mir-290*) and *mir-302~367* (*mir-302*), in individual cells using knock-in reporters. During embryonic development and ESC differentiation, all cells expressed *mir-290* and *mir-302* sequentially in that order. During OSK-driven reprogramming, however, cells activated the miRNA loci stochastically, in a non-ordered manner. Interestingly, adding Sall4 to the OSK cocktail increased reprogramming efficiency and led to the preferential activation of *mir-302* before *mir-290*. These results show that cells can follow multiple paths during the late stages of reprogramming and that the trajectory of any individual cell is strongly influenced by the combination of factors introduced.

This study was published in *Cell Stem Cell* in 2014 (Parchem et al., 2014).

Introduction

The production of iPSCs by somatic cell reprogramming has become commonplace in the laboratory since its invention in 2006 (Takahashi and Yamanaka, 2006). However, the mechanisms by which the process occurs are still poorly understood. Somatic cell reprogramming is often conceptualized as a backward movement up Waddington's epigenetic landscape theoretically traversing a linear path through developmental stages in reverse (Hochedlinger and Plath, 2009). Indeed, a number of studies have demonstrated the progressive downregulation of somatic markers and upregulation of pluripotency markers over the course of

reprogramming (Brambrink et al., 2008; Stadtfeld et al., 2008; Polo et al., 2012; Chan et al., 2009). Single-cell expression analyses also suggest an early stochastic but late hierarchical pattern of gene activation, which argues that, after a certain point in the reprogramming process, cells undergo an immutable set of cell fate transitions that lead to the iPSC state (Buganim et al., 2012; Golipour et al., 2012). However, these studies focused only on genes that are expressed either in the starting fibroblast population or the final pluripotent population. Because they did not examine genes that are uniquely expressed in an intermediate stage of development, it is unclear whether reprogramming cells de-differentiate along a path that includes transitions through intervening stages of development on their way to naïve pluripotency.

Here, we developed fluorescent reporters for the *mir-290* and *mir-302* loci, which allowed us to follow individual cells over the course of embryonic development, ESC differentiation, and somatic cell reprogramming to iPSCs. Because *mir-290* and *mir-302* are highly expressed in the naïve and primed pluripotent states, respectively, they provide a unique view into the progression of differentiation and directed de-differentiation.

Results

The *mir-290* and *mir-302* clusters define transient and distinct states of pluripotency

To characterize the expression of *mir-290* and *mir-302* at the single cell level, fluorescent markers were knocked into each of the loci, creating a “dual reporter” system in which *mir-290* would be co-expressed with mCherry and *mir-302* with eGFP (Figure 3A). Examining dual reporter mouse embryos, we found that all cells of the embryo proper expressed the reporters in a sequential and stereotyped fashion. The ICM cells of E3.5 embryos uniquely expressed *mir-290-mCherry*, the E5.5 epiblast expressed both reporters, and the E7.5 gastrulating embryo

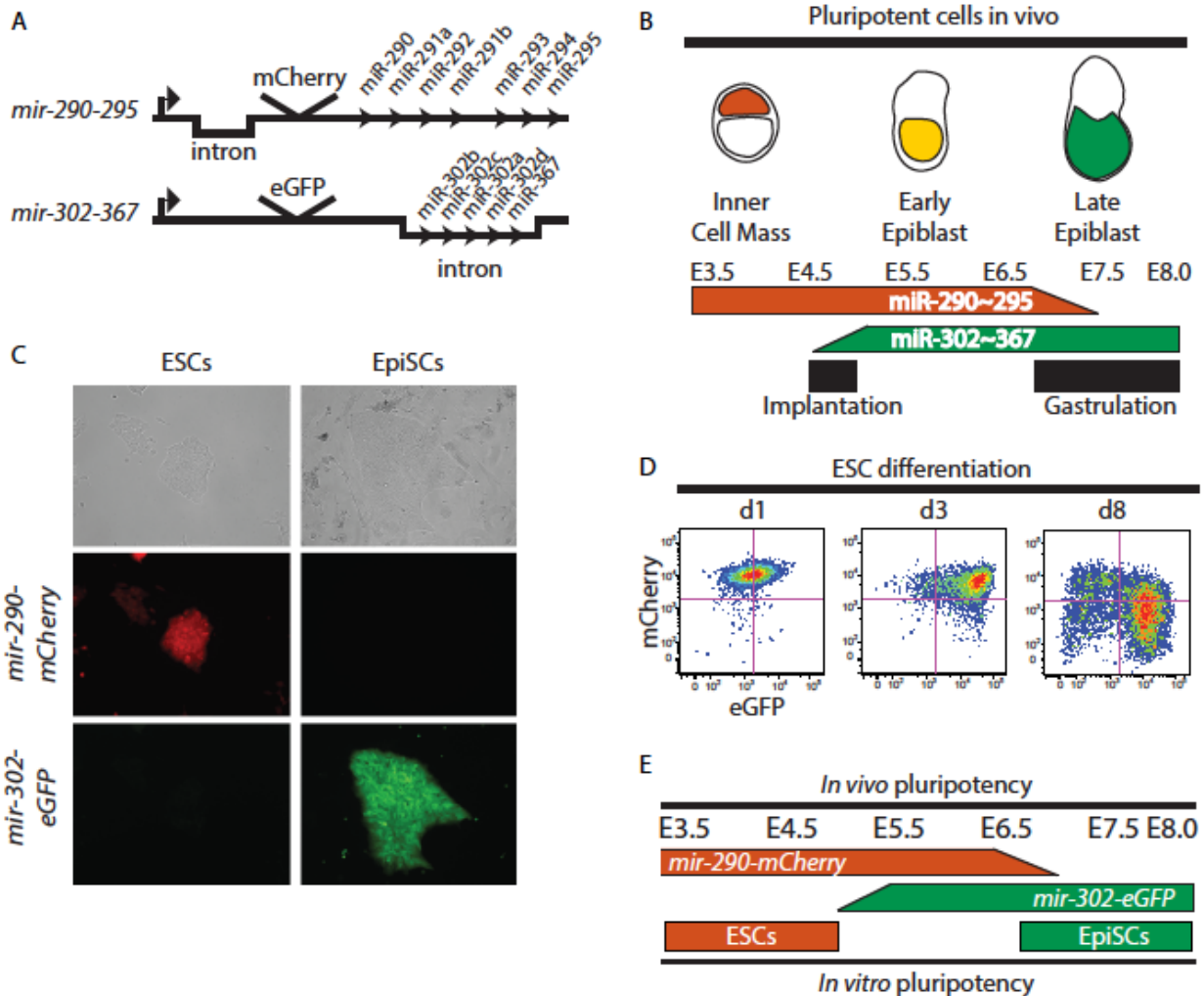


Figure 3. The *mir-290* and *mir-302* clusters define sequential stages of pluripotency *in vivo* and *in vitro*. (A) Schematic of dual reporter design. (B) Schematic summary of *mir-290* and *mir-302* expression during mouse embryonic development. (C) Representative images of dual reporter ESCs and EpiSCs (D) Flow cytometric analysis of mCherry and eGFP expression in dual reporter ESCs cultured without LIF and 2i for 1-8 days. (E) Schematic summary comparing reporter expression *in vitro* and *in vivo*.

specifically expressed *mir-302-eGFP* (Figure 3B). This progression of reporter expression was recapitulated *in vitro*. Naïve pluripotent ESCs derived from E3.5 blastocysts were exclusively *mir-290-mCherry*⁺, while primed pluripotent EpiSCs from the E5.5 epiblast were exclusively *mir-302-eGFP*⁺ (Figure 3C). Additionally, dual reporter ESCs allowed to differentiate upon the removal of LIF and Mek and GSK3 inhibitors (2i) first activated *mir-302-eGFP* and then shut off *mir-290-mCherry* just like their *in vivo* counterparts (Figure 3D). These data show that the

expression of *mir-290* and *mir-302* loci can be used to define progressive stages of differentiation through pluripotent states both *in vivo* and *in vitro* (Figure 3E).

The *mir-290* and *mir-302* clusters are stochastically activated in OSK-driven reprogramming

Given the highly ordered and sequential expression of the *mir-290* and *mir-302* loci during embryonic development and ESC differentiation, we asked whether the sequence of locus activation would occur in reverse during somatic cell reprogramming (Figure 4A). We introduced Oct4, Sox2, and Klf4 (OSK) retroviruses into dual reporter MEFs, treated them with media formulations commonly used for producing iPSCs, and tracked reporter activation using time-lapse fluorescence microscopy. The resulting iPSCs were expanded and shown to silence exogenous pluripotency factors, activate endogenous pluripotency genes, and contribute to germline of chimeric mice (data not shown), thus demonstrating their full acquisition of naïve pluripotent stem cell properties.

While virtually all iPSC colonies formed by day 16 of the experimental protocol were *mir-290-mCherry*⁺/*mir-302-eGFP*⁻ (red) (Figure 4B), examination of the reprogramming intermediates at higher magnification over time revealed multiple patterns of miRNA locus activation before full colony formation. Many patches of cells showed no activation of *mir-302-eGFP* during colony formation, while some colonies arose from *mir-302-eGFP*⁺ patches, and others arose from a heterogeneous mixture of *mir-290-mCherry*⁺/*mir-302-eGFP*⁻, *mir-290-mCherry*⁺/*mir-302-eGFP*⁺, and *mir-290-mCherry*⁻/*mir-302-eGFP*⁺ intermediates (Figure 4C). To more carefully characterize the activation of these loci, we performed flow cytometry throughout the course of reprogramming (Figure 4D and 4E). Fluorescent cells were detected

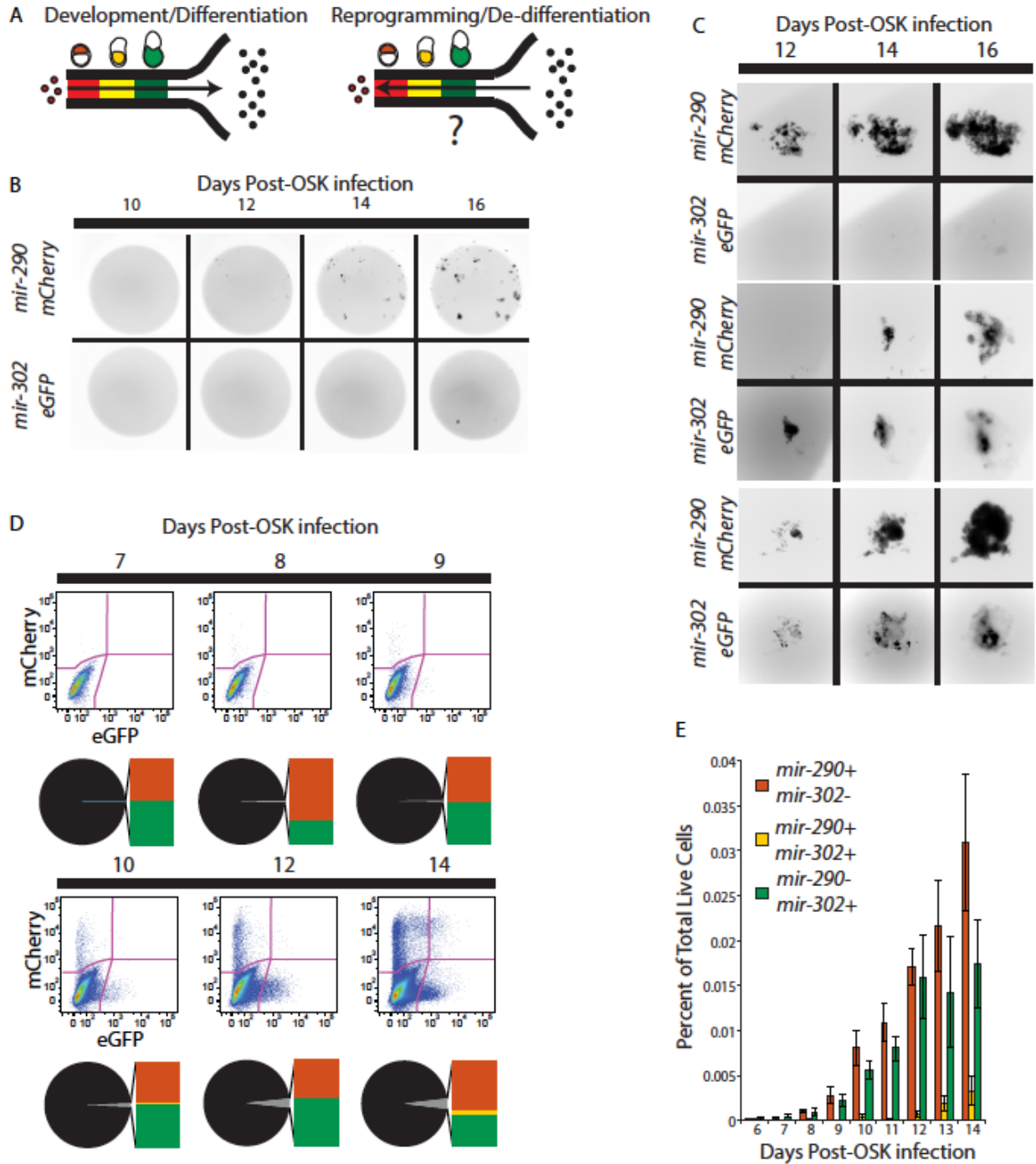


Figure 4. OSK reprogramming leads to late-stage stochastic activation of *mir-290* and *mir-302*. (A) Schematic overview of *mir-290* and *mir-302* expression during normal differentiation toward somatic tissues and hypothetical reversal during reprogramming. (B) Representative (2x) images of *mir-290* and *mir-302* expression in reprogramming wells. (C) Representative high magnification (10x) images of individual patches of reprogramming cells showing multiple patterns of miRNA locus activation before forming *mir-290*-mCherry⁺/*mir-302*-eGFP⁻ iPSC colonies. (D) Representative FACS plots and corresponding bar and pie graphs showing distribution of cells expression *mir-290*-mCherry and/or *mir-302*-eGFP during reprogramming. Black pie slices represent cells that have not activated either miRNA locus. Gray pie slices indicate fluorescent cells, whose fluorescence distributions are represented by the bar graphs. (E) Quantification of (D). Error bars represent the SD of 3 biological replicates.

beginning around day 7. To our surprise, most were either *mir-290-mCherry*⁺ (red) or *mir-302-eGFP*⁺ (green), and they were present in similar numbers, which suggests that these miRNA loci were activated stochastically. Cells that were *mir-290-mCherry*⁺/*mir-302-eGFP*⁺ (yellow) did not appear until a few days later and only small numbers; presumably, these reprogramming intermediates originated from single positive parental cells. Together, our microscopy and flow cytometric results suggest that, in contrast to developmentally programmed differentiation, activation of these miRNA loci during reprogramming is stochastic and unordered.

Next, we asked whether expression of *mir-290*, which is uniquely expressed in the naïve pluripotent state, could be used to prospectively identify cells that are farther along in the de-differentiation process. Reprogramming intermediates were sorted by fluorescence activated cell sorting (FACS) at day 12, because it was the first time point that there were enough cells expressing all possible combinations of the reporters to perform the analysis with and that was early enough to make it unlikely that individual cells had already passed through the expression of both markers. Sorted cells were plated onto irradiated MEFs and evaluated for iPSC colony formation 5-6 days later (Figure 5A). As expected, double negative (black) cells formed colonies with the lowest efficiency (0.03%). Surprisingly, however, *mir-290-mCherry*⁺/*mir-302-eGFP*⁻ (red) and *mir-290-mCherry*⁻/*mir-302-eGFP*⁺ (green) sorted cells displayed similar colony formation efficiencies—1.2% and 0.6%, respectively (Figure 5B and 5C), while the relatively rare (0.015% of all live cells) double positive *mir-290-mCherry*⁺/*mir-302-eGFP*⁺ (yellow) cells demonstrated the highest colony formation efficiency at 12% (Figure 5B and 5C). We also performed mRNA profiling by microarray of black, red, yellow, and green day 12 reprogramming intermediates to compare their overall molecular constitution. Principal component analysis (PCA) showed that red and green cells were more similar to each other than

to yellow cells and that yellow cells were the most similar to iPSCs (Figure 5D). Therefore, the global profiles of these reprogramming populations correlate with their relative potential to form iPSCs. In sum, these data show that, in OSK conditions, it is the number of these miRNA loci activated, not the order of their activation, that predicts the extent of reprogramming, even though one of these loci (*mir-302*) is not expressed in the final iPSC state.

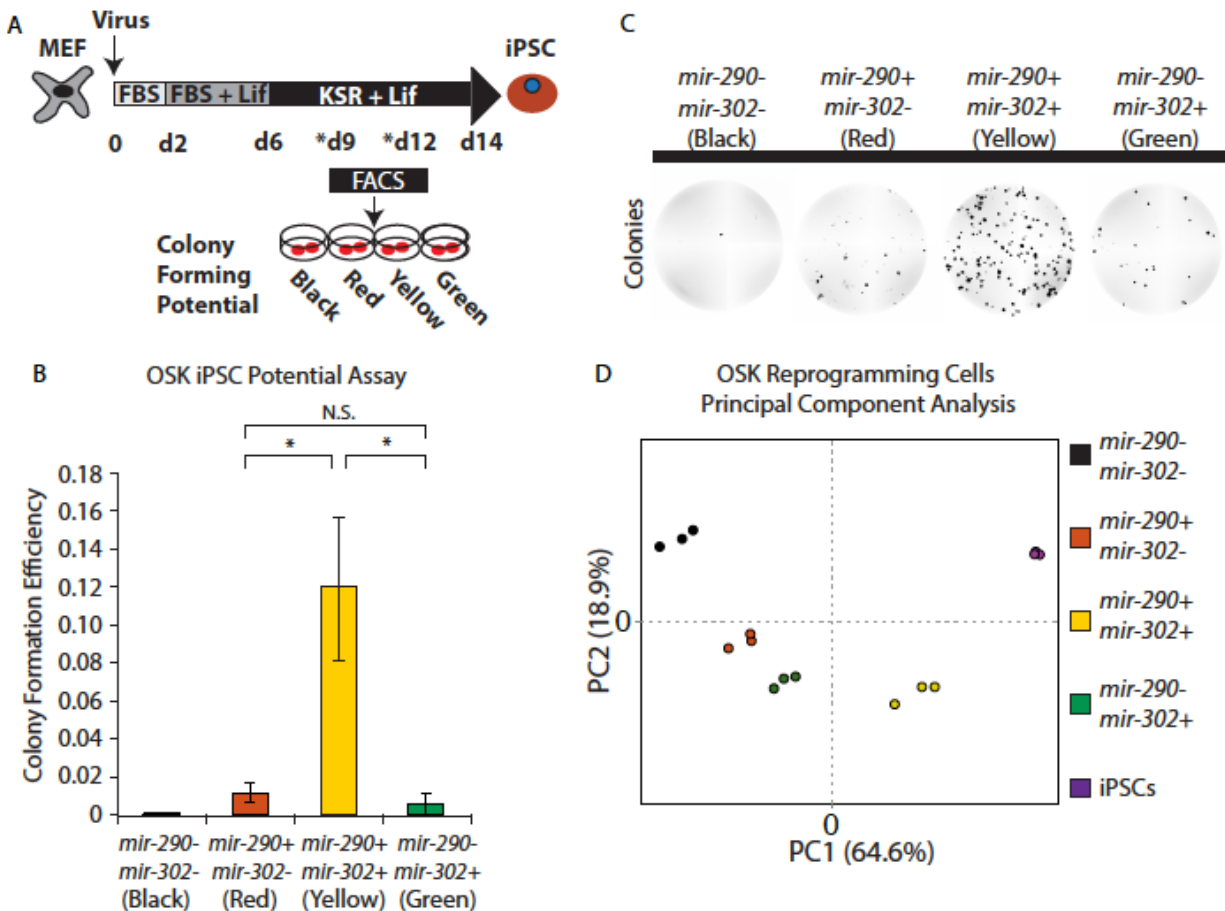


Figure 5. *mir-290-mCherry+*/*mir-302-eGFP+* double positive cells show the greatest potential for generating iPSC colonies. (A) Schematic overview of the iPSC colony forming potential assay. Reprogramming wells were sorted either on day 9 (for OSKS) or day 12 (for OSK) and plated onto irradiated MEFs in ESC medium. Sorted cells were imaged daily and counted 5-6 days after FACS. The colony forming efficiency was calculated as the number of colonies formed per number of cells plated. (B) Colony forming efficiency of day 12 OSK reprogramming intermediates. *p < 0.01, N.S. = not significant as determined by Student's t-test. Error bars represent the SD of 3 biological replicates. (C) Representative images of *mir-290-mCherry+* colony formation wells of OSK reprogramming intermediates sorted at day 12 (1000 cells/well) and imaged 5 days later. (D) PCA of Illumina bead array expression profiles from day 12 OSK reprogramming intermediates and fully reprogrammed iPSCs (n = 3 for each cell type). Percentages reflect the proportion of variance assigned to each principal component.

Sall4 enhances reprogramming and biases the path taken to the iPSC state

We next investigated whether the addition of other reprogramming factors would influence the order of miRNA locus activation. We focused on the transcription factor Sall4, because it is highly expressed in ESCs, regulates the naïve state, and is a powerful enhancer of reprogramming (Tsubooka et al. 2009; Sakaki-Yumoto et al., 2006; Lim et al., 2008; Zhang et al., 2006; Elling et al., 2006). Treating MEFs with OSK and Sall4 (OSKS) dramatically increased the number of iPSC colonies generated compared with OSK (Figure 6A and 6B).

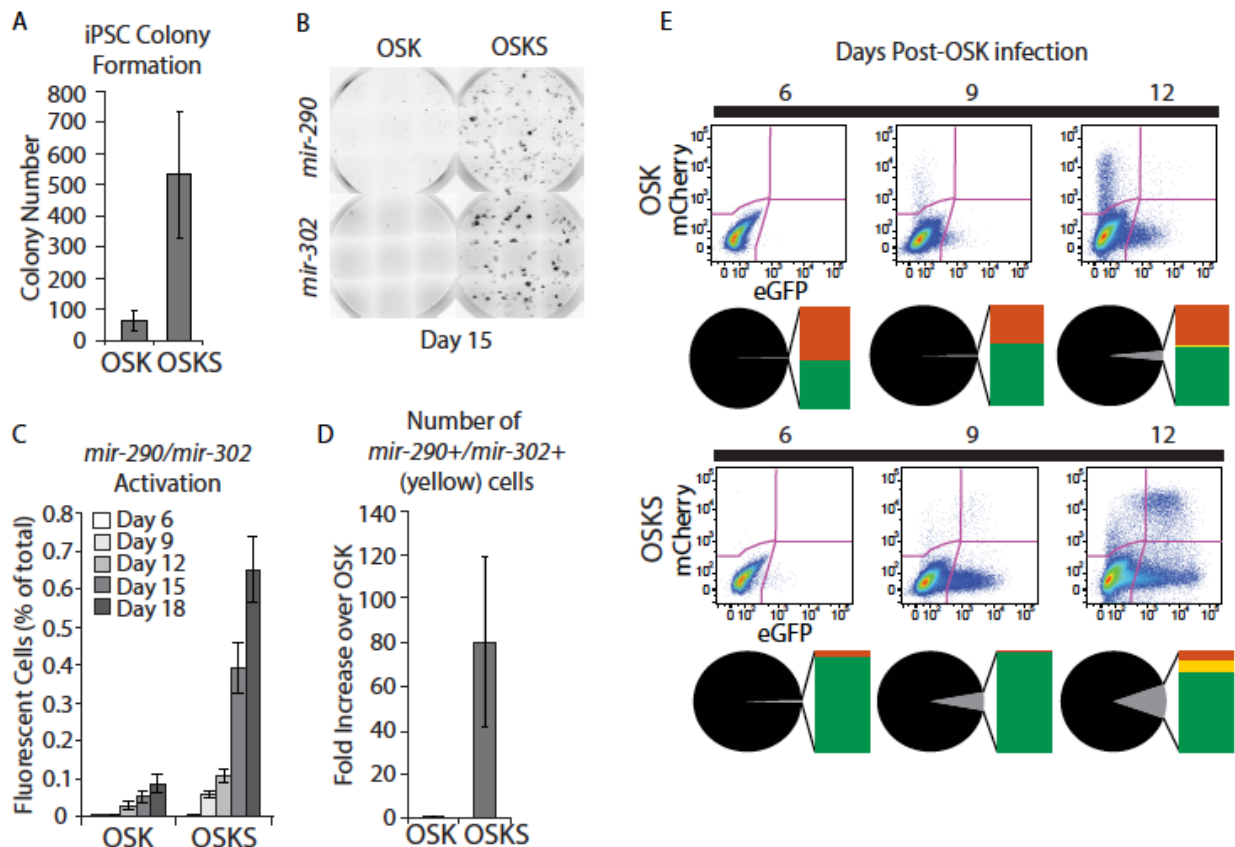


Figure 6. Sall4 enhances reprogramming and biases intermediates to activate *mir-302*. (A) Quantification of iPSC colony formation in OSK or OSKS conditions at day 15 of reprogramming. (B) Representative images of day 15 reprogramming wells. (C) Flow cytometric quantification of the fraction of total fluorescent cells (*mir-290+* and/or *mir-302+*) over multiple days during OSK or OSKS reprogramming. (D) Number of *mir-290-mCherry+/mir-302-eGFP+* double positive cells at day 12 of OSK or OSKS reprogramming. Error bars in A, C, and D represent SD of 3 biological replicates. (E) Distribution of *mir-290* and *mir-302* expressing cells from days 6-12 of reprogramming represented as described in Figure 2D. Pie and bar slices represent the average of 3 biological replicates.

Moreover, flow cytometry over the course of reprogramming revealed that the addition of Sall4 not only increased the frequency and kinetics of reporter activation (Figure 6C) but also increased the number of double-positive (yellow) intermediates detected at day 12 by about 80-fold (Figure 6D). Even more remarkably, OSKS-treated reprogramming cells appeared to preferentially activate the *mir-302* locus before the *mir-290* locus, as suggested by the increase in green reprogramming intermediates (Fig. 6E). High-resolution time-lapse microscopy confirmed this inference (Figure 7A and 7B): tracing individual cells over a period of several days, we found that OSK-treated cells went from expressing neither marker (black) to expressing either *mir-302-eGFP* (green) or *mir-290-mCherry* (red) in comparable numbers, while OSKS-treated cells mostly transitioned from black to green and then to yellow before settling into a final red state. Only a small number of OSKS-treated cells transitioned from black to red and then to yellow. These findings show that cells take multiple routes to the naïve state under OSK conditions but preferentially traverse certain paths under OSKS conditions (Figure 7C).

We tested the colony forming potential of the cells generated at day 9 during OSKS reprogramming, the first time point cells could be detected for all fluorescent populations. As in OSK reprogramming, *mir-290-mCherry*⁺/*mir-302-eGFP* (yellow) cells were the most efficient (Figure 8A). Interestingly, however, we found that OSKS-treated *mir-290-mCherry*⁺/*mir-302-eGFP*⁻ (red) cells formed colonies at a much higher efficiency than red cells generated during OSK reprogramming—indeed, OSKS red and yellow cells formed colonies with comparable efficiency (Figure 8A). To determine whether these differences in iPSC potential in OSK and OSKS conditions are correlated with a global transcriptional signature, we performed microarray profiling on black, red, yellow, and green cells obtained during OSK and OSKS reprogramming. We hypothesized that OSKS red cells, which presumably arose directly from black cells but have

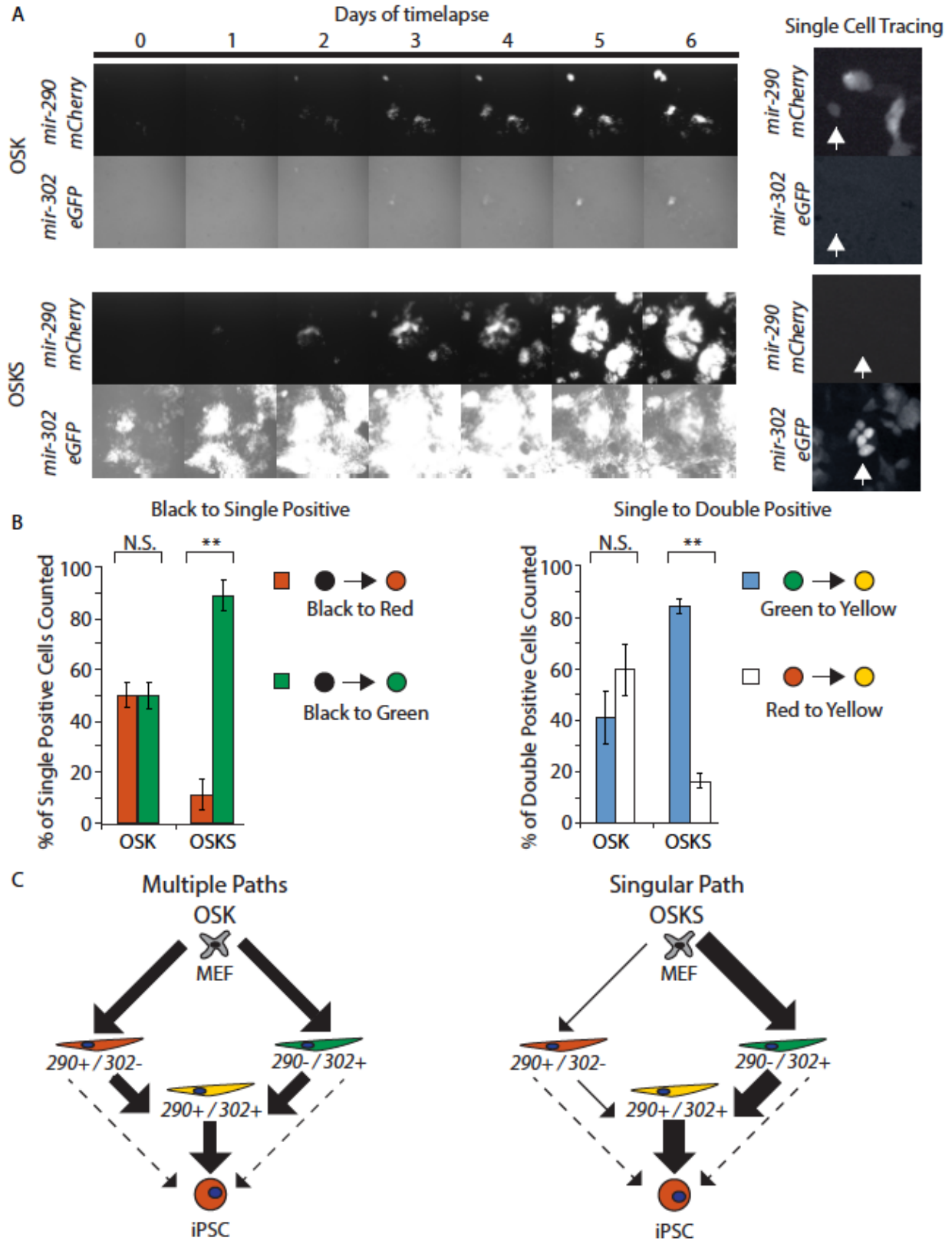


Figure 7. Sall4 biases the path taken during reprogramming. (A) Representative time-lapse images over 6 days during OSK or OSKS reprogramming showing the evolution of miRNA loci expression beginning with a single cell. Right, representative images demonstrating how individual cells were scored. Arrows point to matching cells. (B)

Quantification of transitions in reporter expression observed with time-lapse imaging every 2-4hrs. Activation of *mir-290-mCherry* or *mir-302-eGFP* was quantified in single cells that went on to form colonies. The number of cells that changed reporter expression before forming an iPSC colony is presented as a percentage of the total number of single-positive or double-positive cells traced. In total, 303 single-positive cells were followed in OSK (150 black-to-red and 153 black-to-green), and 467 single-positive cells were followed in OSKS (45 black-to-red and 422 black-to-green). 114 double-positive cells were traced in OSK (46 green-to-yellow and 68 red-to-yellow), and 236 double-positive cells were traced in OSKS (198 green-to-yellow and 38 red-to-yellow). Black-to-double-positive (black-to-yellow) transitions were not observed. Values represent the average of 3 different wells. Error bars represent the SD of cells traced over 3 wells. **p < 0.005, N.S. = not significant as determined by Student's t-test. (C) Schematic summary showing potential paths taken by reprogramming cells with regard to the *mir-290* and *mir-302* reporters. Bold arrows represent the preferred path observed during reprogramming. Dashed arrows represent possible paths that were not quantified in our experiments.

a very high iPSC colony forming potential, would have a transcriptional profile more similar to that of yellow cells than of green cells. However, both PCA and hierarchical clustering of day 9 OSKS reprogramming cell transcriptomes showed similar results to OSK: regardless of the reprogramming factors used, the overall mRNA profiles of red and green cells were the most similar to each other than to that of yellow cells, and yellow cells were the most similar to iPSCs (compare Figure 8B and 8C with Figure 5D). Therefore, we asked if there was a signature common to cells with the greatest iPSC potential (OSK yellow cells and OSKS red and yellow cells) compared with those with lower iPSC potential (OSK red and green cells and OSKS green cells). Using a p value cutoff of 0.01 and log2 fold change greater than 0.4, we identified 312 genes that were differentially expressed across different comparisons of cells with high and low iPSC potential (Figure 8D). Fourteen of these genes overlapped across comparisons (Figure 8D), so we asked if they were enriched in the naïve or primed pluripotent states by performing mRNA profiling of dual reporter ESCs during ESC differentiation as well as derived EpiSCs. We found that the 14-gene signature was downregulated during the naïve-to-primed transition while the non-overlapping genes showed little change (Figure 8E). Consistent with their differential expression in cell populations with the highest iPSC potential, a number of these genes, including *Dnmt3l*, *Dnmt3b*, *Esrrb*, *Dppa5*, *Dppa4*, *Ooep*, and *Rhox5*, have been previously

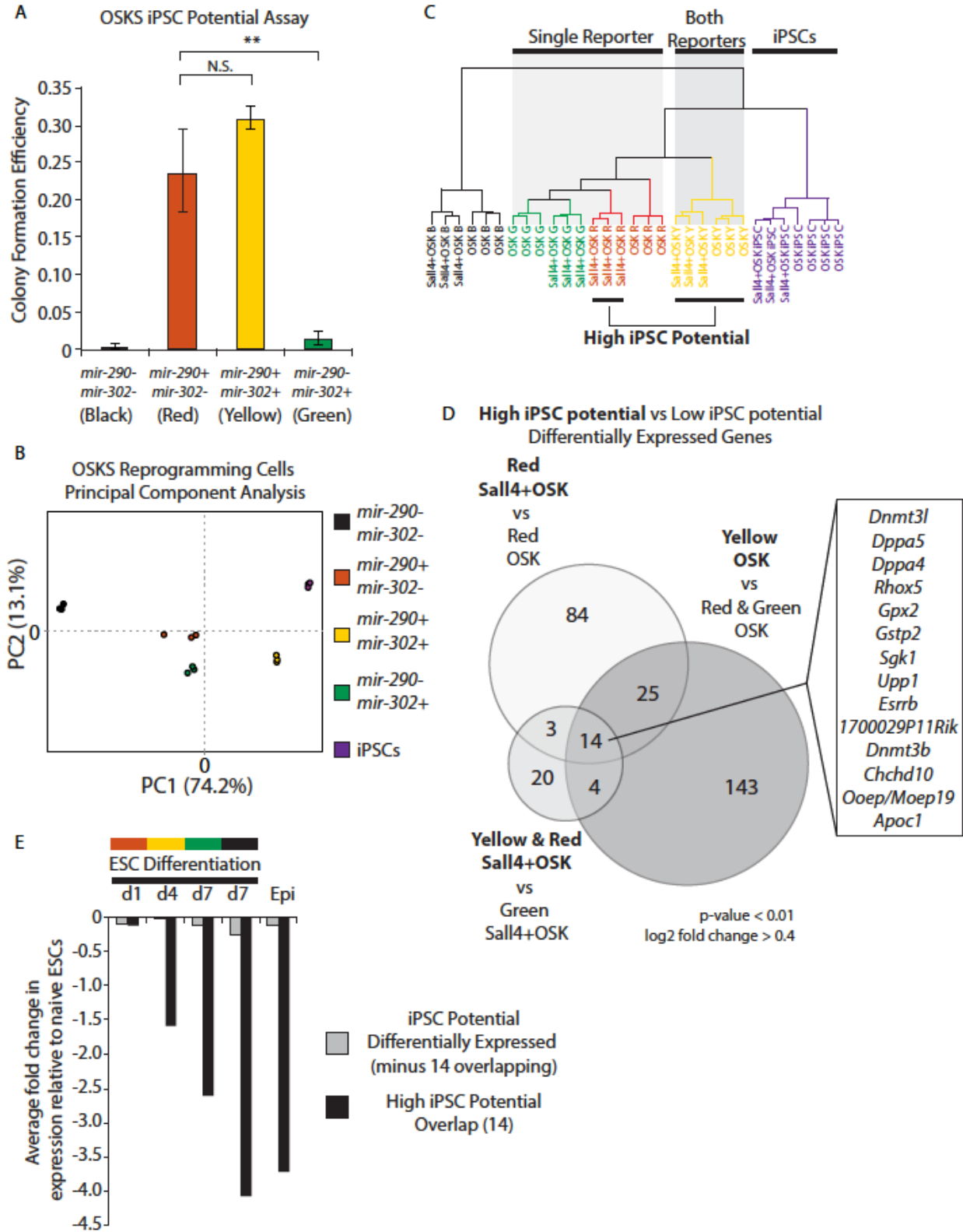


Figure 8. High iPSC potential is associated with a naïve pluripotent gene signature. (A) Colony forming efficiency of day 9 OSKS reprogramming intermediates. Error bars represent SD of 3 biological replicates. ** $p < 0.005$, N.S. = not significant, as determined by Student's t-test. (B) PCA of Illumina bead array expression profiles from day 9 OSKS reprogramming intermediates and fully reprogrammed iPSCs ($n = 3$ for each cell type).

Percentages reflect the proportion of variance assigned to each principal component. (C) Hierarchical clustering of Illumina bead array expression profiles described in (B). (D) Venn diagram showing overlap of genes that are differentially expressed between populations with high (bold font) and low (regular font) iPSC potential. $p < 0.01$, \log_2 fold change > 0.4 . (E) Average expression during ESC differentiation and in primed EpiSC cultures of all differentially expressed iPSC potential genes or the 14 genes associated with high iPSC potential.

identified as regulators of naïve pluripotency (Okano et al., 1999; Amano et al., 2006; Madan, et al., 2009; Tashiro et al., 2010; Festuccia et al., 2012; Martello et al., 2012; Neri et al., 2013).

Thus, these data identify a set of genes that are coregulated with *mir-290* but not *mir-302* locus activation during OSKS reprogramming, are enriched in naïve cells over primed cells, and positively correlate with iPSC colony forming potential.

Discussion

Our dual reporter system showed that changes in *mir-290* and *mir-302* expression delineate the specific cell fate transitions that separate naïve and primed pluripotency during differentiation both *in vivo* and *in vitro*. Although previous work had shown enriched expression of *mir-290* and *mir-302* in ESC and EpiSC populations, respectively (Stadler et al., 2010; Jouneau et al., 2012; Card et al., 2008), it was unclear how these expression patterns were reflected at the single-cell level. Here, we show that pluripotent cells of the embryo transition through three states that are marked by *mir-290* alone, *mir-290* and *mir-302* concurrently, and *mir-302* alone, before silencing both loci in most, if not all, somatic populations. These transitions correlate with important developmental events, such as implantation (activation of *mir-302*) and gastrulation (downregulation of *mir-290*).

Using the activation of *mir-290* and *mir-302* as readouts of developmental status, we found that de-differentiation during OSK reprogramming does not require the activation of genes in a definite order. In fact, reprogramming cells that activated both *mir-290* and *mir-302* showed

greater molecular similarity to naïve ESCs than either single positive population and were the most likely to form iPSC colonies—an observation that contrasts with the situation in normal differentiation, in which double-positive cells have decreased colony forming ability relative to *mir-290-mCherry* single positive cells (data not shown). Together, our colony forming assays and molecular profiling results demonstrate that the activation of *mir-290* alone does not necessarily mark the completion of reprogramming. While some *mir-290-mCherry* single positive cells are likely to form fully reprogrammed iPSCs, others represent intermediates that will not completely reach the naïve pluripotent state.

However, the addition of the reprogramming enhancer *Sall4* biased the process such that a plurality of cells activated *mir-302* before *mir-290*. Moreover, OSKS-treated *mir-290-mCherry* single positive cells showed high colony-forming efficiency comparable to that of double positive cells, which is evidence that a larger fraction of OSKS-treated red cells are functionally closer to the iPSC state than their OSK-treated counterparts. The ability of *Sall4* to lead to early *mir-302* activation cannot be explained simply by the direct activation of the locus because there is a long delay from time of factor introduction and *mir-302* activation, and because *Sall4* is more highly expressed in the naïve (*mir-302*-) state than in the primed (*mir-302*+) state. Interestingly, it was previously shown that Oct4 and Sox2 bind to the promoter of *mir-302* and are required for its expression, but they are clearly not sufficient, since Oct4 and Sox2 are also highly expressed in the naïve (*mir-302*-) state (Card et al., 2008; Marson et al., 2008). Future studies aimed at understanding the transcriptional network of *mir-302* should provide important insight into the drivers of the primed state.

In sum, our results show that the path taken by cells during reprogramming is strongly affected by the combination of factors used. For example, adding *Sall4* promoted a more defined

trajectory, possibly by reducing stochasticity in the later stages of reprogramming. Nevertheless, it is unlikely that any particular iPSC intermediate, either stable or transient, marks a required common route taken by all cells that fully reprogram. Instead, there are most likely multiple paths, and the specific one chosen by any one cell is influenced by the combination of factors introduced and their relative expression.

However, the finding that an enhancer of reprogramming leads to biased activation of *mir-302* before *mir-290* in a manner that would be expected for developmental reversal leaves open the possibility that, although not required, a sequential unfolding of the differentiation program in reverse increases the efficiency of iPSC generation. Such an idea is not unreasonable, considering the fact that directed differentiation of therapeutic cell types from pluripotent stem cells often involves a recapitulation of developmental events. Indeed, the genes associated with high iPSC potential that we identified are differentially expressed in naïve and primed pluripotent cells and are upregulated with increasing iPSC potential in reprogramming intermediates—a phenomenon that is consistent with a backwards recapitulation of normal differentiation.

Materials and Methods

Targeting strategy

Mir-290-mCherry/mir-302-eGFP (dual reporter) ES cells were generated by inserting coding regions for eGFP and mCherry downstream of the transcriptional start sites of *mir-302* and *mir-290*, respectively. Genomic homology arms encompassing the *mir-290* and *mir-302* clusters were subcloned from BACs into pL253 targeting vectors by recombineering. Fluorescent proteins adjacent to SV40pA were cloned into the pL452 recombineering vector next to the

floxed neomycin cassette. The fluorescent protein/SV40 polyA and floxed neomycin cassette were then inserted downstream of the miRNA transcriptional start sites in the pL253 targeting constructs using recombineering. pL253 targeting constructs were linearized and electroporated (20 μ g) into V6.5 ES cells followed by selection with geneticin (G418, 200 μ g/mL) for 7 days. Subclones were screened for proper targeting using long range PCR. Primers specific to the transgene (i.e. fluorescent protein/Neomycin cassette) and genomic regions outside the homology arms of the targeting construct were used to screen for properly targeted colonies. The resulting targeted clones were transfected with a Cre recombinase-expressing plasmid to remove the neomycin cassette. PCR was used to screen subclones for loss of neomycin.

ESC and EpiSC derivation

For ESCs, timed matings of dual reporter mice were used to isolate blastocysts at E3.5. Each blastocyst was plated in one well of a 24-well dish on a MEF feeder layer using ESC medium with 20% KSR and Erk1/2 inhibitor (1 μ M, PD0325901). Blastocyst outgrowths were trypsinized and passaged after 3-4 days until ESC lines were established. For EpiSCs, timed matings of dual reporter mice were used to isolate embryos between E5.5-E7.5. Each epiblast was dissected and plated onto MEF feeder layers in 1 well of a 48-well plate using N2B27 basal EpiSC medium containing bFGF (12ng/mL) and Activin A (20ng/mL). The undifferentiated portions of outgrowths were picked and passaged every 3-4 days until homogeneous cultures were obtained. ESCs were maintained on MEFs or gelatinized plates in 15% FBS in DMEM with L-Glutamine, non essential amino acids, penicillin/streptomycin, 55 μ M beta-mercaptoethanol, and 1000U/mL LIF with or without the addition of 2i (Stemgent, PD0325901 (1 μ M) and CHIR99021 (3 μ M)). EpiSCs could be maintained on MEFs or CellStart CTS (Invitrogen) coated plates in either 20%

KSR/(DMEM:F12) or N2B27 basal medium with L-Glutamine, non essential amino acids, penicillin/streptomycin, 12ng/mL bFGF, and 20ng/mL Activin A.

ESC differentiation

ESCs were plated at 40,000 cells per well in a 12-well plate on day 0 in FBS+LIF+2i conditions (control) or FBS-LIF-2i conditions (differentiation). Culture medium consisted of DMEM, 15% FBS, L-Glutamine, non essential amino acids, penicillin/streptomycin, 55 μ M beta-mercaptoethanol with or without LIF/2i (1000 units/mL LIF and 2i (Stemgent, PD0325901 (1 μ M) and CHIR99021 (3 μ M))). Every day for 8 days, a single well of each condition was trypsinized and analyzed via flow cytometry using the LSR II flow cytometer (BD Biosciences).

MEF Generation

MEFs were generated as previously described (Judson et al., 2009). In brief, dual reporter embryos were harvested at E13.5. Heads and visceral tissue were removed. Remaining tissue was dissociated by trypsinization and physical disruption and plated as Passage 0 (P0) cells in MEF media (high glucose (H-21) DMEM, 10% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, 55 μ M 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-expression plasmids at a ratio of 1:2 following the Fugene 6 manufacturer's protocol. At 24 hours, media was replaced with fresh MEF media. At 48 hours, supernatant was harvested, filtered (0.45 μ m) and frozen at -80°C.

Reprogramming and de-differentiation

MEFs (P4) were plated onto gelatin-coated Greiner uClear black-walled 96-well imaging plates at 900 cells/well or standard 12-well plates at 12,000 cells/well. The next day, 50ul (96-well) or 1.2mL (12-well) of each retrovirus-containing supernatant with 4µg/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, penicillin/streptomycin, 55uM beta-mercaptoethanol and 1000U/ml 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, penicillin/streptomycin, 55 µM beta-mercaptoethanol and LIF] after day 6. Colony formation was assessed on days indicated. High throughput imaging and high content analysis were conducted with the INCell Analyzer 2000 imaging station and INCell Developer software suite (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. P3 colonies were harvested for RT-qPCR and fixed for immunohistochemistry.

iPSC colony formation assay

From days 5 through 14 of reprogramming, individual reprogramming wells (12-well) were trypsinized and analyzed/sorted on a FACS Aria II cell sorter (BD Biosciences). Up to 1000 cells for each population (black, red, yellow, green) were sorted and plated onto irradiated MEF feeder layers (P1) in 96-well plates with reprogramming medium [Knock-out DMEM (Invitrogen), 15% Knock-out Serum Replacement (Invitrogen), non- essential amino acids, L-

glutamine, penicillin/streptomycin, 55 μ M beta-mercaptoethanol and 1000U/ml LIF]. These plates were imaged daily with the INCell Analyzer 2000 (GE) to follow colony formation. The INCell Developer software (GE) was used to count the number of iPSC-like (defined by morphology and reporter expression) colonies 5 or 6 days after sorting and to quantify the area of *mir-290-mCherry* and *mir-302-eGFP* expression in each well. The ratio of colonies formed to the number of cells plated was defined as the iPSC colony forming potential.

mRNA profiling

For analysis of differentiating dual reporter ESCs and reprogramming MEFs, total RNA was isolated from sorted populations using miRNeasy micro columns according to manufacturers protocol (Qiagen). Naive ESCs were maintained in N2B27 medium supplemented with LIF+2i. For ESC differentiation, cells were placed in FGF+Activin to induce differentiation and sorted for reporter expression at day 1 (red), day 4 (yellow) and day 7 (green and black). EpiSCs were derived from embryos and maintained in FGF+Activin. Reprogramming dual reporter cells were cultured as described above and were sorted at either day 9 (OSKS) or day 12 (OSK) as indicated in the main text. Triplicate biological samples were analyzed using Illumina MouseRef-8 v2.0 Expression BeadChips run by the UCLA Neuroscience Genomics Core.

High-resolution time-lapse microscopy

Reprogramming wells were imaged at 10X magnification every 2-4 hours beginning at day 9 following OSK or OSKS retroviral transduction using a BioStation CT (Nikon). CL-Quant software (Nikon) was used to analyze the data and prepare movies. To characterize transitions in reporter expression during reprogramming, activation of *miR-290-mCherry* and *mir-302-eGFP*

in individual cells was counted in areas that went on to form colonies. The activation of *mir-290* or *mir-302* occurred either in black cells (black to single positive) or fluorescent cells (single to double positive). To calculate the frequency of initial activation of fluorescence, the number of instances in which *mir-290* or *mir-302* is activated in black cells was divided by the total number of single positive cells counted. For single positive to double positive transitions, the total number of green-to-yellow or red-to-yellow transitions was divided by the total number of yellow cells counted. At least 10 fields of view, which included at least 15 (OSK) or 20 (OSKS) reprogramming events were analyzed.

Chapter 3—miR-290 and miR-302 are dispensable for somatic cell reprogramming

Summary

The miR-290~295 (miR-290) and miR-302a-d (miR-302) families of miRNAs are important components of the pluripotency and developmental circuitry: they are highly expressed in pluripotent cells, their overexpression increases the efficiency of Yamanaka factor-driven somatic cell reprogramming, and they are critical for embryogenesis. Using genetic knock-out (KO) models of miR-290 and miR-302, we show that, paradoxically, neither miRNA family is required for mouse somatic cell reprogramming to iPSCs. These findings underscore the complexity of miRNA networks and their context-dependent effects on cell state.

This work is being prepared for submission to *Stem Cell Reports*.

Introduction

As discussed in the previous chapters, the *mir-290~295* and *mir-302~367* miRNA clusters play major roles in the regulation of pluripotent stem cells. Along with members of the *mir-17~92* and *mir-106b~25* clusters, many of the miRNAs in these two clusters share a similar seed sequence (reviewed in Vidigal and Ventura, 2012), and are sometimes referred to as the ESCC family of miRNAs. Notably, however, miR-293 and miR-367 have seed sequences that are different. Here, we will refer to *mir-290~295* and *mir-302~367* cluster miRNAs with the ESCC seed sequence as the miR-290 and miR-302 miRNA families.

Given the importance of miR-290 and miR-302 in pluripotency, we asked if they are required for somatic cell reprogramming. We used genetic knock-out (KO) systems of the *mir-290~295* (*mir-290* KO, Medeiros et al., 2011) and *mir-302a-d* (*mir-302* KO, Parchem et al., 2015) loci to address this question. Remarkably, MEFs deficient in either miR-302 or miR-290

were both able to generate iPSCs, and reduction of miR-302 in *mir-290* KO cells did not affect reprogramming efficiency either. These data suggest that while these miRNA families are strong drivers and enhancers of reprogramming, neither is necessary for the process to occur.

Results

Expression of miR-302 is not required for reprogramming

To investigate whether miR-302 miRNAs are required for the de-differentiation of somatic cells to iPSCs, we harvested mouse embryonic fibroblasts (MEFs) from *mir-302a-d*^{-/-} (*mir-302* KO) and *mir-302a-d*^{+/-} (*mir-302* Het) embryos. In this KO model, eGFP was knocked into the *mir-302~367* locus, such that *mir-302a-d* are deleted, while *mir-367* is left intact (Parchem et al., 2015). We transduced the MEFs with retroviruses expressing the pluripotency factors Oct4, Sox2, Klf4 (OSK), treated them with media formulations commonly used for producing iPSCs, and stained them for Nanog expression 15 days later. To our great surprise, *mir-302* KO MEFs were not only able to generate Nanog⁺ colonies, but did so at an efficiency similar to that of their *mir-302* Het counterparts, which we used as controls (Figure 9A and 9E). Initially anticipating that reprogramming would be adversely affected with loss of miR-302, we also transfected OSK-treated MEFs with miR-302b mimics (OSK+302), hoping to demonstrate a “rescue” effect; instead, however, we found that OSK+302 increased the number of Nanog⁺ colonies over OSK similarly in control and *mir-302* KO MEFs (Figure 9B and 9E). These observations suggest that endogenous and exogenous miR-302 have different functional effects on reprogramming. When we added Sall4 to the OSK cocktail (OSKS), control MEFs showed an increase in Nanog⁺ colonies as expected, but, interestingly, this enhancement in reprogramming efficiency was blunted in *mir-302* KO MEFs (Figure 9C and 9E). Because these *mir-302* Het and

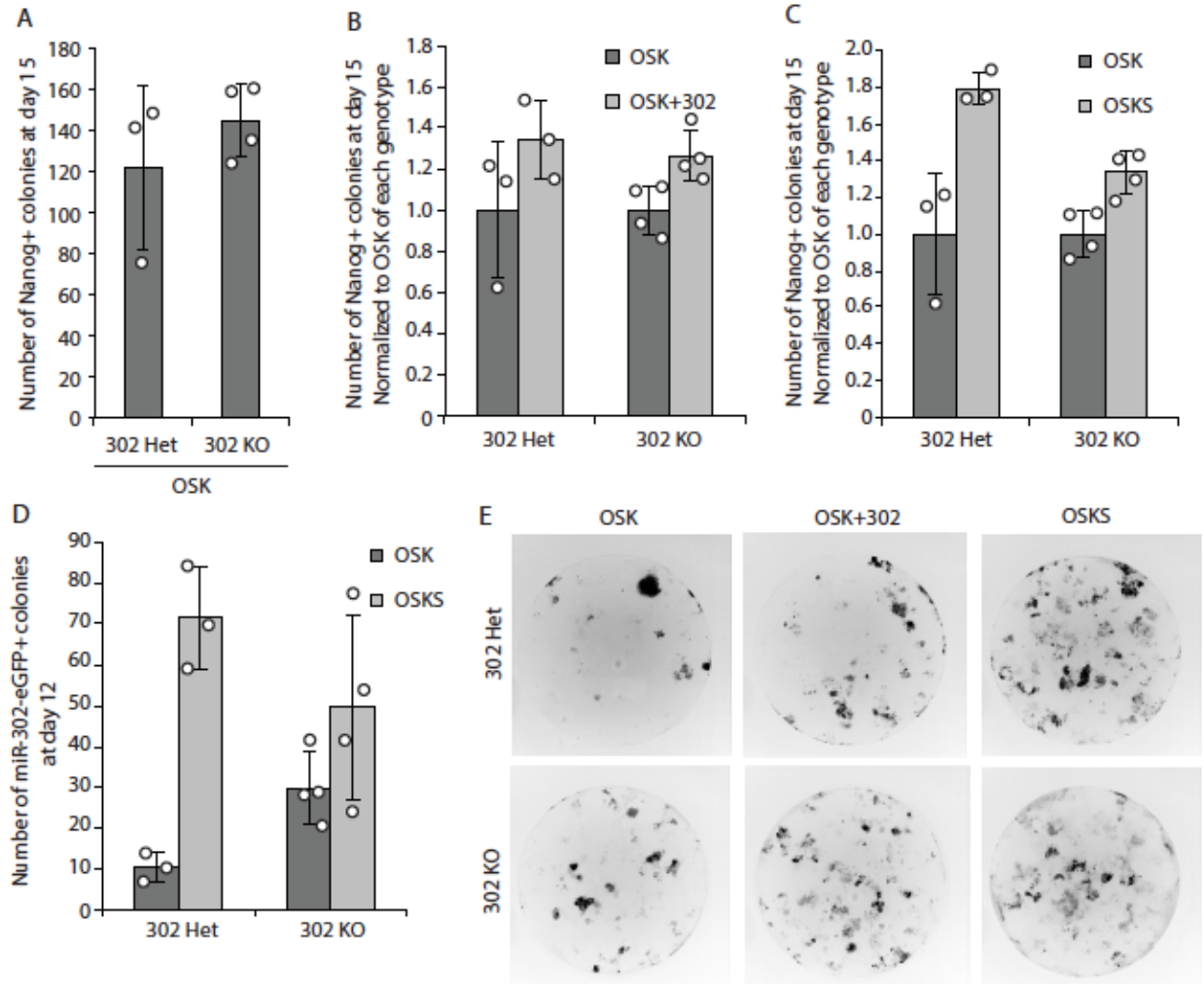


Figure 9. *mir-302* KO MEFs can reprogram to iPSCs. (A) Average number of Nanog+ colonies counted per well 15 days after OSK transduction. (B) Fold change in Nanog+ colonies 15 days after OSK transduction and miR-302b mimic transfection compared with OSK transduction alone. (C) Fold change in Nanog+ colonies 15 days after transduction with OSKS compared with OSK. (D) Number of *miR-302-eGFP*+ colonies counted per well 12 days after OSK or OSKS transduction. (E) Representative images of individual Nanog-stained reprogramming wells at day 15. Error bars represent SD of 3-4 biological replicates. Circles indicate individual data points.

KO cells become eGFP+ when the *mir-302~367* locus is transcribed and because we have previously shown that the addition of Sall4 drives OSK-treated MEFs to preferentially activate this miRNA locus during reprogramming (Chapter 2; Parchem et al., 2014), we counted the number of eGFP+ colonies that formed at an intermediate timepoint during the reprogramming process. We saw that for *mir-302* KO cells, the increase in eGFP+ reprogramming intermediates with OSKS treatment compared to OSK was less dramatic than for control cells, although

baseline eGFP levels in OSK-treated *mir-302* KO cells was higher than control, possibly because *mir-302* KO cells possess 2 eGFP alleles, while *mir-302* Het cells only have 1 eGFP allele

(Figure 9D).

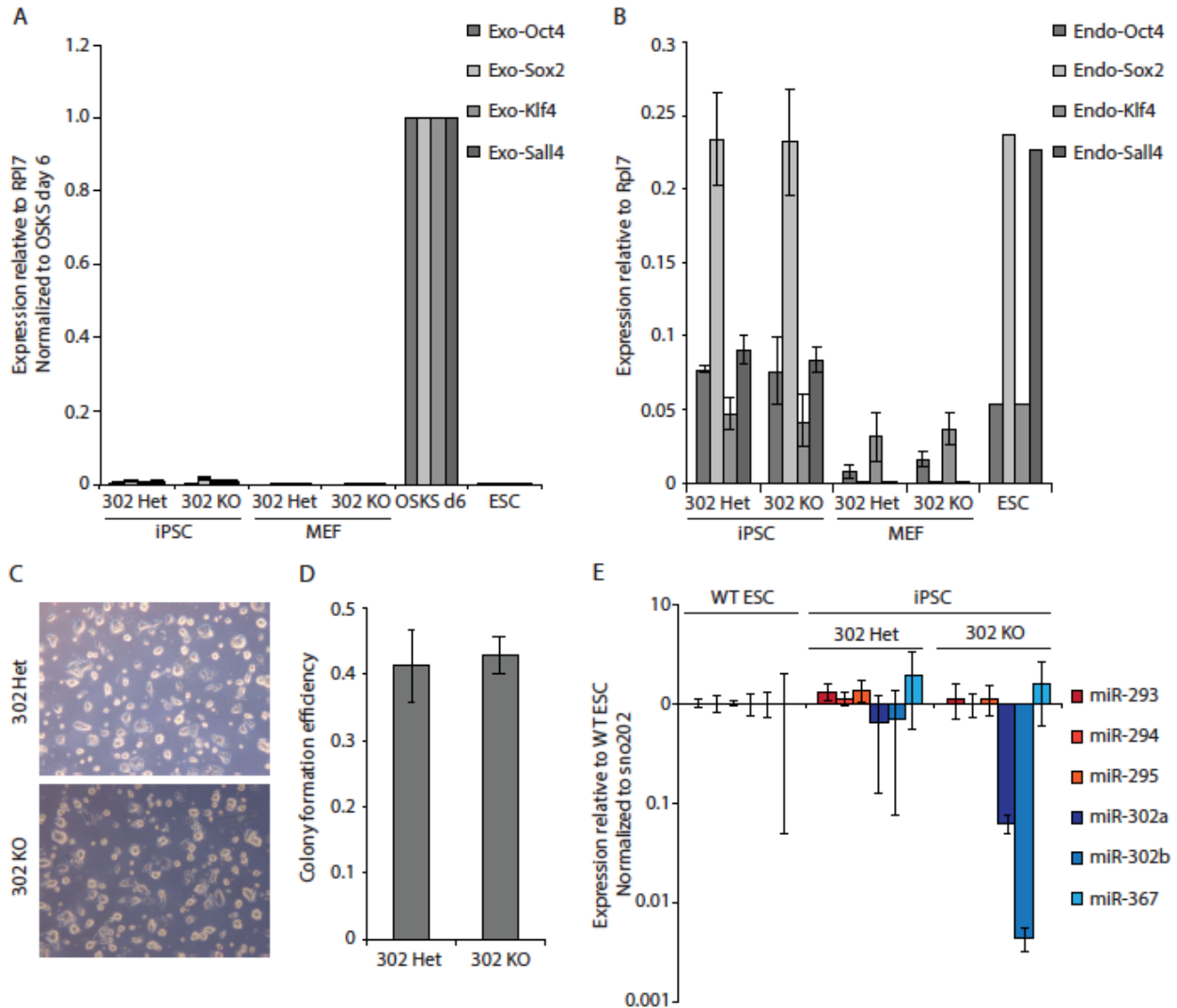


Figure 10. *mir-302* KO iPSCs have no detectable molecular or functional deficits. (A) qRT-PCR analysis of exogenous OSKS expression levels in WT ESCs, and *mir-302* Het and *mir-302* KO MEFs and iPSCs. Data are normalized to MEFs collected 6 days after OSKS transduction. (B) qRT-PCR analysis of endogenous OSKS expression levels in WT ESCs and *mir-302* Het and *mir-302* KO iPSCs. (C) Representative brightfield images of reprogrammed and expanded *mir-302* Het and *mir-302* KO iPSCs. (D) Colony formation efficiency (fraction of cells plated that form Nanog+ colonies) of *mir-302* Het and *mir-302* KO iPSCs. (E) Transcript levels of mature *mir-290*~*295* and *mir-302*~*367* cluster miRNAs in WT ESCs, *mir-302* Het iPSCs, and *mir-302* KO iPSCs. Error bars represent SD of 3-4 biological replicates. Circles indicate individual data points.

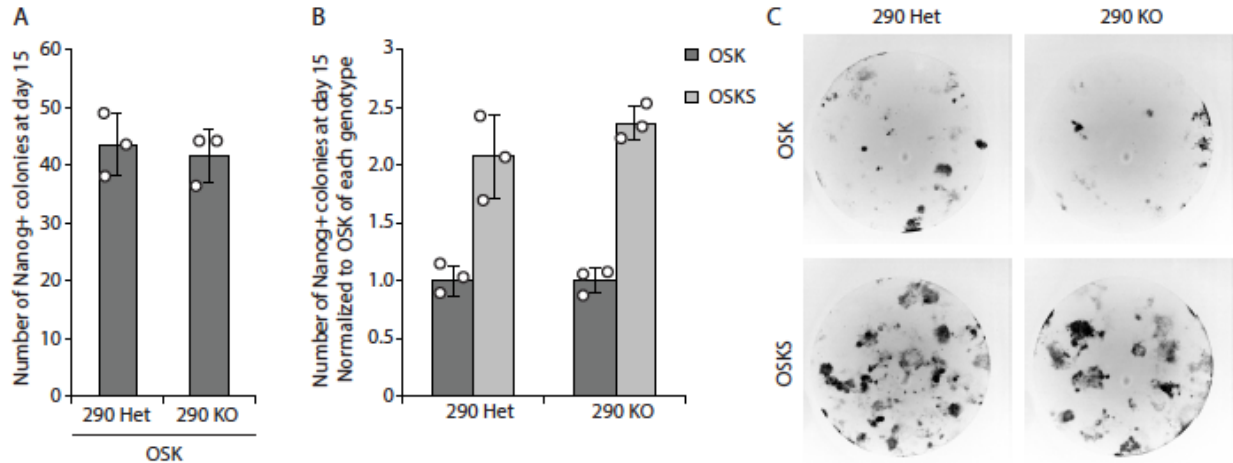


Figure 11. *mir-290* KO MEFs can reprogram to iPSCs. (A) Average number of Nanog+ colonies counted per well 15 days after OSK transduction. (B) Fold change in Nanog+ colonies 15 days after transduction with OSKS compared with OSK. (C) Representative images of individual Nanog-stained reprogramming wells at day 15. Circles indicate individual data points.

Expression of miR-290 is not required for reprogramming

Although the ability of the *mir-302* KO MEFs to reprogram surprised us, it could still be conceptually rationalized, as the *mir-302~367* locus is not expressed in the mouse naïve pluripotent state, which was the endpoint of our de-differentiation experiments. The *mir-290~295* locus, however, is highly expressed in this state. Therefore, we next asked how MEFs from *mir-290~295*^{-/-} (*mir-290* KO, Medeiros et al., 2011) embryos would behave in our reprogramming assay. Again unexpectedly, *mir-290* KO MEFs generated similar numbers of Nanog+ colonies as control (*mir-290* Het) with OSK transduction (Figure 11A and 11C). Interestingly, unlike with *mir-302* KO MEFs, iPSC formation was enhanced with the addition of Sall4 regardless of the *mir-290* genotype (Figure 11B and 11C).

We were able to expand the *mir-290* Het and KO iPSCs, confirm their genotype (data not shown), and demonstrate that they silence exogenous retroviruses (Figure 12A) and activate endogenous pluripotency genes (Figure 12B). Interestingly, *mir-290* KO iPSCs appear to have a slight delay in shutting off exogenous factors, as residual levels are still detectable after

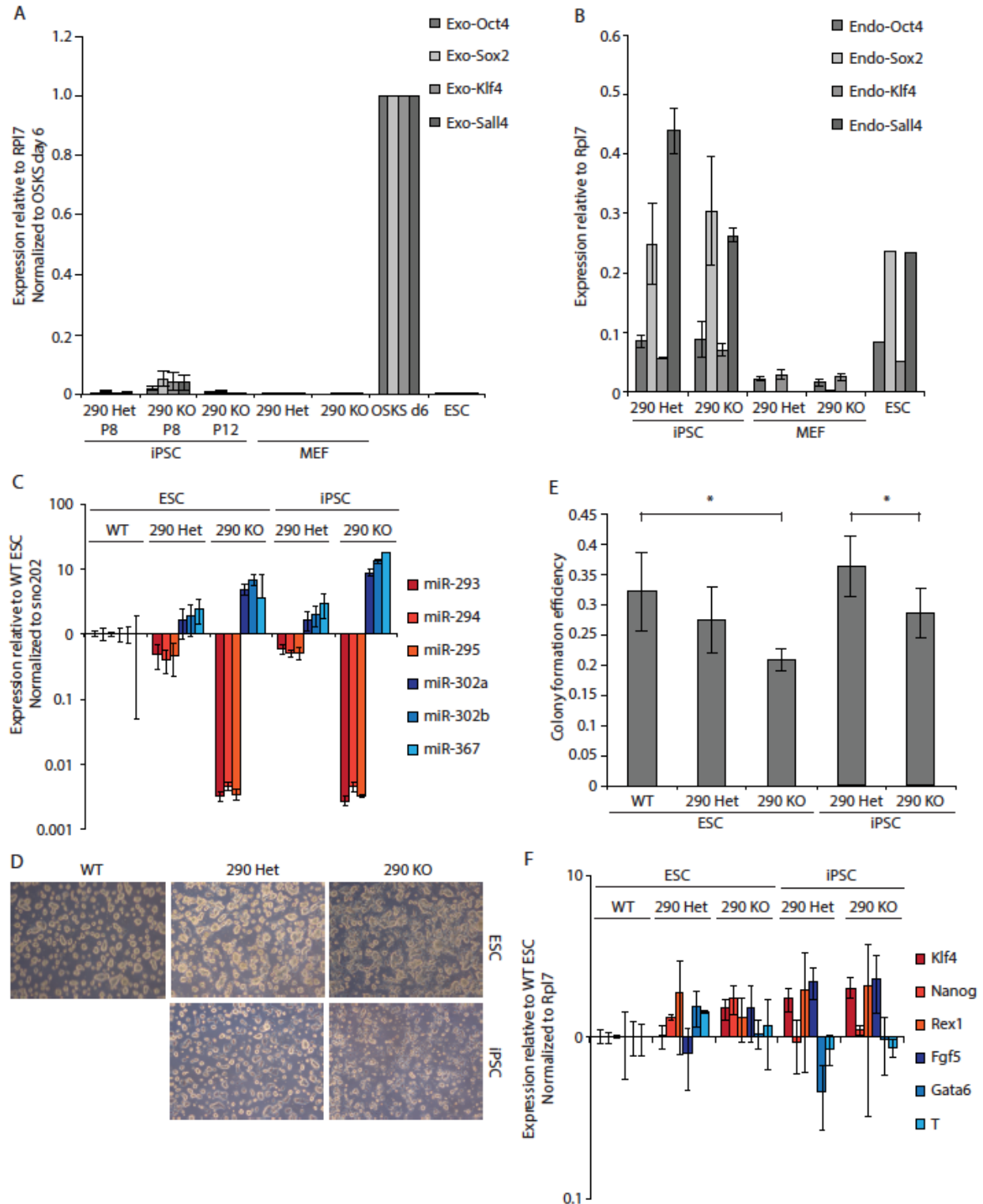


Figure 12. *mir-290* KO iPSCs exhibit minor deficits. (A) qRT-PCR analysis of exogenous OSKS expression levels in WT ESCs, and *mir-290* Het and *mir-290* KO MEFs and iPSCs. Data are normalized to MEFs collected 6 days after OSKS transduction. (B) qRT-PCR analysis of endogenous OSKS expression levels in WT ESCs and *mir-290*Het and *mir-290* KO iPSCs. (C) Transcript levels of mature *mir-290~295* and *mir-302~367* cluster miRNAs in WT ESCs, *mir-290* Het iPSCs, and *mir-290* KO iPSCs. (D) Representative brightfield images of reprogrammed and expanded *mir-302* Het and *mir-302* KO iPSCs. (E) Colony formation efficiency (fraction of cells plated that form

Nanog+ colonies) of *mir-302* Het and *mir-302* KO iPSCs. (F) qRT-PCR analysis of naïve pluripotency markers (Klf4, Nanog, Rex1) and primed pluripotency markers (Fgf5, Gata6, T) in WT ESCs and *mir-290* Het and *mir-290* KO iPSCs. Error bars represent SD of 3 biological replicates (A-C, F) or 4-6 biological replicates (E). Circles indicate individual data points. *P < 0.05

approximately 30 days in culture (“P. 8”) compared to the nearly completely absent expression in control iPSCs; an additional 2 weeks in culture (“P. 12”) was required to shut down retrovirus expression to control levels (Figure 12B), suggesting a lag in the maturation of the iPSC state. Even more striking was our observation that *mir-290* KO iPSCs express increased amounts of *mir-302~367* cluster miRNAs (Figure 12C). This phenotype appears not to be due to a failure to fully achieve naïve pluripotency but is, rather, inherent to the *mir-290* KO state, as *mir-290* Het and KO embryo-derived ESCs also show similarly elevated levels of miR-302 relative to WT ESCs (Figure 12C). Moreover, both *mir-290* KO ESCs and *mir-290* KO iPSCs look morphologically distinct from their *mir-290* Het and WT counterparts, with heterogeneous colony outgrowths that suggest that the cells may be in a metastable state (Figure 12D). Indeed, colony formation efficiency of *mir-290* KO ESCs and *mir-290* KO iPSCs is significantly reduced, demonstrating a reduced capacity for self-renewal (Figure 12E). However, expression of a panel of representative naïve pluripotency and primed pluripotency markers was not appreciably different across the different ESC and iPSC genotypes (Figure 12F), suggesting that *mir-290* KO ESCs and iPSCs can stably maintain at least some aspects of naïve pluripotency. Together, these data suggest that while the *mir-290* cluster is not essential for reprogramming, its loss does delay the attainment of a fully reprogrammed state and negatively affect the self-renewal properties of the end state pluripotent stem cells.

Depleting miR-302 in *mir-290* KO cells does not affect reprogramming

The upregulation of *mir-302~367* cluster miRNAs in *mir-290* KO cells raised the question of whether reprogramming is unaffected in these cells because miR-302 and other miRNAs that share the same seed sequence are able to functionally compensate for the absence of miR-290. This possibility would ideally be addressed genetically—by performing the de-differentiation assay in *mir-290~295*^{-/-}; *mir-302a-d*^{-/-} MEFs; however, these double KO embryos are virtually unrecoverable by E9.5 due to a dramatic arrest in development (Parchem et al., 2015). Therefore, our strategy was to attempt to functionally remove or inactivate miR-302 miRNAs by transducing *mir-290* Het and KO MEFs with a lentiviral construct in which eGFP expression is controlled by a 3'UTR containing binding sites for miR-302b (Judson et al., 2013, Figure 13A). This “miR-302 sponge” would, in theory, sequester miR-302 miRNAs from their endogenous targets, as well as read out miR-302 expression levels, since eGFP expression would be decreased when the 3'UTR is miRNA-bound (Ebert et al., 2007, Ebert and Sharp, 2010, Liu et al., 2008, Kluiver et al., 2012). We found that *mir-290* Het and KO MEFs infected with either an *eGFP* construct lacking a 3'UTR (control) or the *eGFP-miR-302-sponge* construct (*mir-302 sponge*) all expressed eGFP at comparable intensity (Figure 13B), suggesting that the sponge is not suppressed by any miRNAs in this somatic state. We introduced the OSK and OSKS reprogramming factors into these MEFs, and found that, surprisingly, the presence of the miR-302 sponge did not affect OSK reprogramming efficiency compared to eGFP in either *mir-290* Het or *mir-290* KO cells (Figure 14A and 14C). In addition, all MEFs, regardless of genotype or presence of sponge, were able to produce Nanog⁺ colonies more efficiently when Sall4 was added to the reprogramming cocktail (Figure 14B and 14C). The resulting iPSCs had the appropriate genotypes, confirming successful integration of the *miR-302 sponge* construct

(Figure 15A), shut down expression of exogenous pluripotency factors (Figure 15B), and activated expression of endogenous pluripotency factors (Figure 15C). Also, transcript levels of *mir-302~367* miRNAs in *mir-290* KO iPSCs were similar regardless of whether or not they were treated with *miR-302 sponge* construct (Figure 15D), suggesting that cells are not further upregulating *mir-302~367* to counterbalance any miR-302 miRNAs that the miR-302 sponge might have removed from the functional pool.

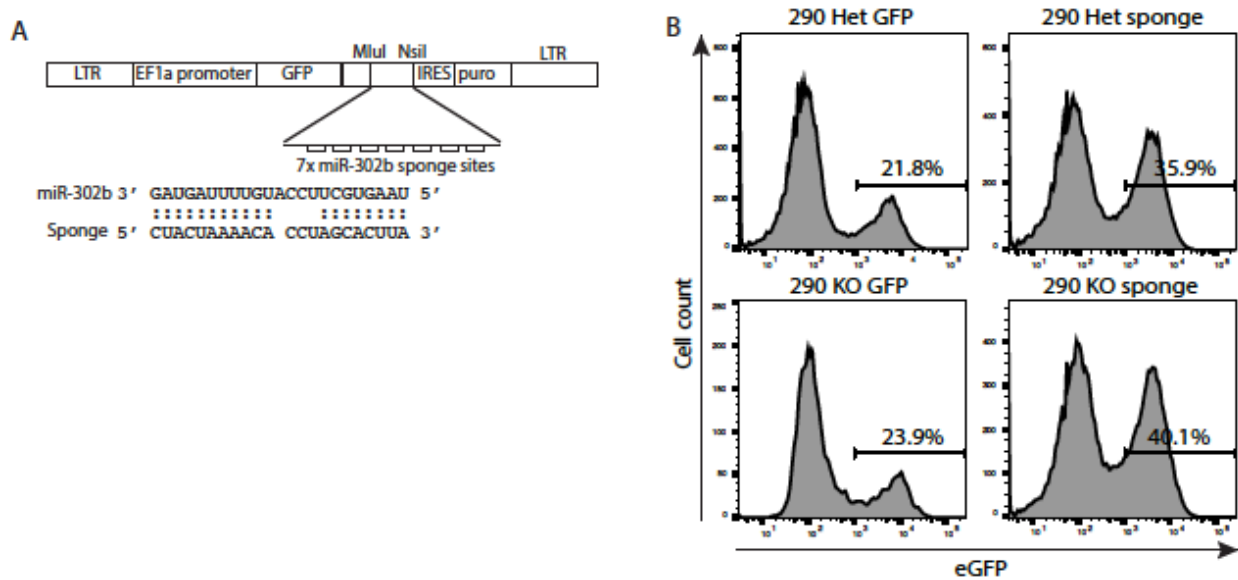


Figure 13. Construction and use of *eGFP-miR-302-sponge* construct. (A) Construct design. Schematic is adapted from Judson et al., 2013. (B) Representative FACS plots of *mir-290* Het and *mir-290* KO MEFs transduced with lentiviral constructs encoding *eGFP* or *eGFP-miR-302-sponge*. Horizontal bar indicates range of *eGFP* fluorescence of MEFs that were sorted and expanded for reprogramming.

Because we did not observe any reprogramming defect in the *mir-290* KO + *miR-302 sponge* MEFs, we wondered if the *mir-302 sponge* construct was truly functional. We attempted to address this issue by examining the expression of miR-302 sponge-controlled *eGFP* protein, transcript levels of the construct itself, and the change in the ability of miR-302b to suppress gene expression in a luciferase assay. *eGFP* expression can provide much information: in theory, high *eGFP* expression in *miR-302 sponge*-treated cells indicates an abundance of unbound miR-302 sponge transcript, while low *eGFP* expression indicates strong miRNA binding to miR-302

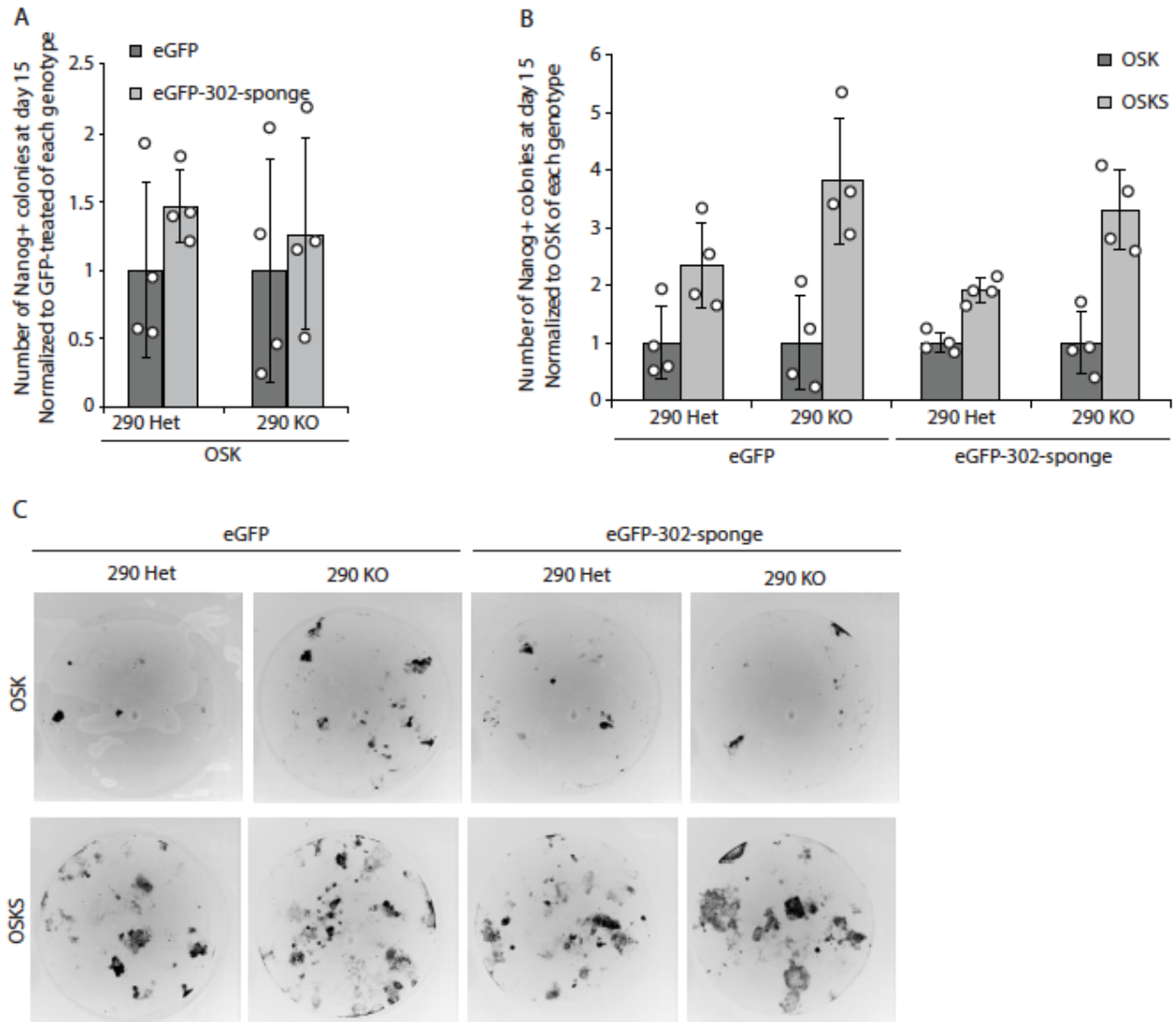


Figure 14. Reducing miR-302 levels in the context of miR-290 loss does not affect somatic cell reprogramming. (A) Fold change in Nanog+ colonies 15 days after OSK transduction in *eGFP* or *eGFP-mir-302-sponge*-treated *mir-290* Het or *mir-290* KO MEFs. (B) Fold change in Nanog+ colonies 15 days after transduction with OSKS compared with OSK. (C) Representative images of individual Nanog-stained reprogramming wells at day 15. Error bars represent SD of 4 biological replicates. Circles indicate individual data points.

sponge, although this scenario is also possible if the starting MEFs were never infected with the construct or if the cells epigenetically silenced the construct at some point during reprogramming. However, the eGFP expression patterns we observed in our cells are highly suggestive not only of miR-302 sponge functionality but also that the miR-302 sponge is also likely binding miR-290 and other miRNAs with the same seed sequence. When scoring day 15

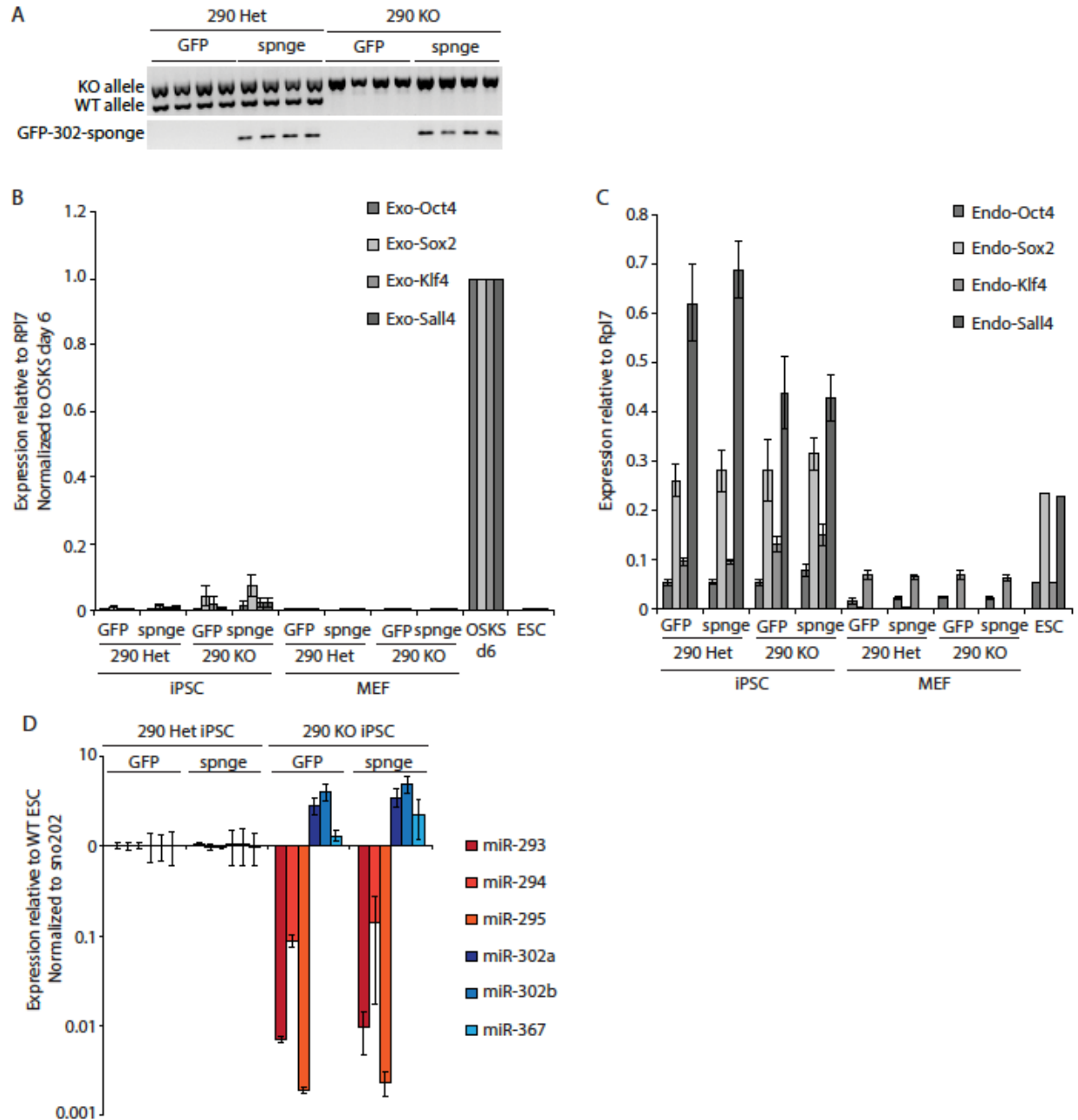


Figure 15. *eGFP-miR-302-sponge*-treated *mir-290* KO iPSCs do not exhibit additional detectable deficits. (A) DNA genotyping results for the *mir-290~295* locus and the *eGFP-miR-302-sponge* construct in the *mir-290* Het and *mir-290* KO iPSC lines. (B) qRT-PCR analysis of exogenous OSKS expression levels in WT ESCs, and *mir-290* Het and *mir-290* KO MEFs and iPSCs treated with *eGFP* or *eGFP-miR-302-sponge*. Data are normalized to MEFs collected 6 days after OSKS transduction. (C) qRT-PCR analysis of endogenous OSKS expression levels in WT ESCs and *mir-290* Het and *mir-290* KO iPSCs treated with *eGFP* or *eGFP-miR-302-sponge*. (D) Transcript levels of mature *mir-290~295* and *mir-302~367* cluster miRNAs in *mir-290* Het and *mir-290* KO iPSCs treated with *eGFP* or *eGFP-miR-302-sponge*. Error bars represent SD of 4 biological replicates.

reprogramming plates for Nanog⁺ colonies, we noticed that the fraction of day 15 Nanog⁺ reprogramming colonies that are also eGFP⁺ was significantly decreased in *mir-290* Het + *mir-302 sponge* iPSCs but not in *mir-290* KO + *mir-302 sponge* iPSCs compared to cells treated with control construct (Figure 16), which suggests that the miR-302 sponge is more strongly suppressed in *mir-290* Het iPSCs than in *mir-290* KO iPSCs. Since *mir-290* Het iPSCs express miR-302 lowly, we inferred that the highly expressed miR-290 is likely binding the miR-302 sponge in these cells. In *mir-290* KO iPSCs, on the other hand, miR-290 is absent and miR-302 is upregulated; however, these Nanog⁺ colonies are largely eGFP⁺, which suggests that the absolute amounts of miR-302 in the cells are insufficient to fully suppress the miR-302 sponge. When we expanded these iPSCs and analyzed their eGFP levels by microscopy and flow cytometry, we saw that *mir-290* Het + *miR-302 sponge* iPSCs had quantitatively lower eGFP intensity than did *mir-290* KO + *miR-302 sponge* iPSCs, while cells expressing control construct had uniformly high eGFP intensity regardless of genotype (Figure 17A-C). Thus, eGFP expression in the iPSC lines supports miR-302 sponge functionality.

We also measured RNA levels of the control and miR-302 sponge constructs (Figure 17D). It was interesting to note that mRNA expression of both the control and *mir-302 sponge* constructs were decreased in iPSCs compared to MEFs, which could be due to differences in promoter efficiency in the two cell types, as well as some degree of epigenetic silencing in the iPSCs. More important, in iPSCs, mRNA levels of miR-302 sponge were lower than that of the control construct, which suggests that the miR-302 sponge mRNA is destabilized upon miRNA binding—likely of miR-290 in the *mir-290* Het iPSCs and miR-302 in the *mir-290* KO iPSCs. However, it is also possible that the *miR-302 sponge* construct is preferentially epigenetically

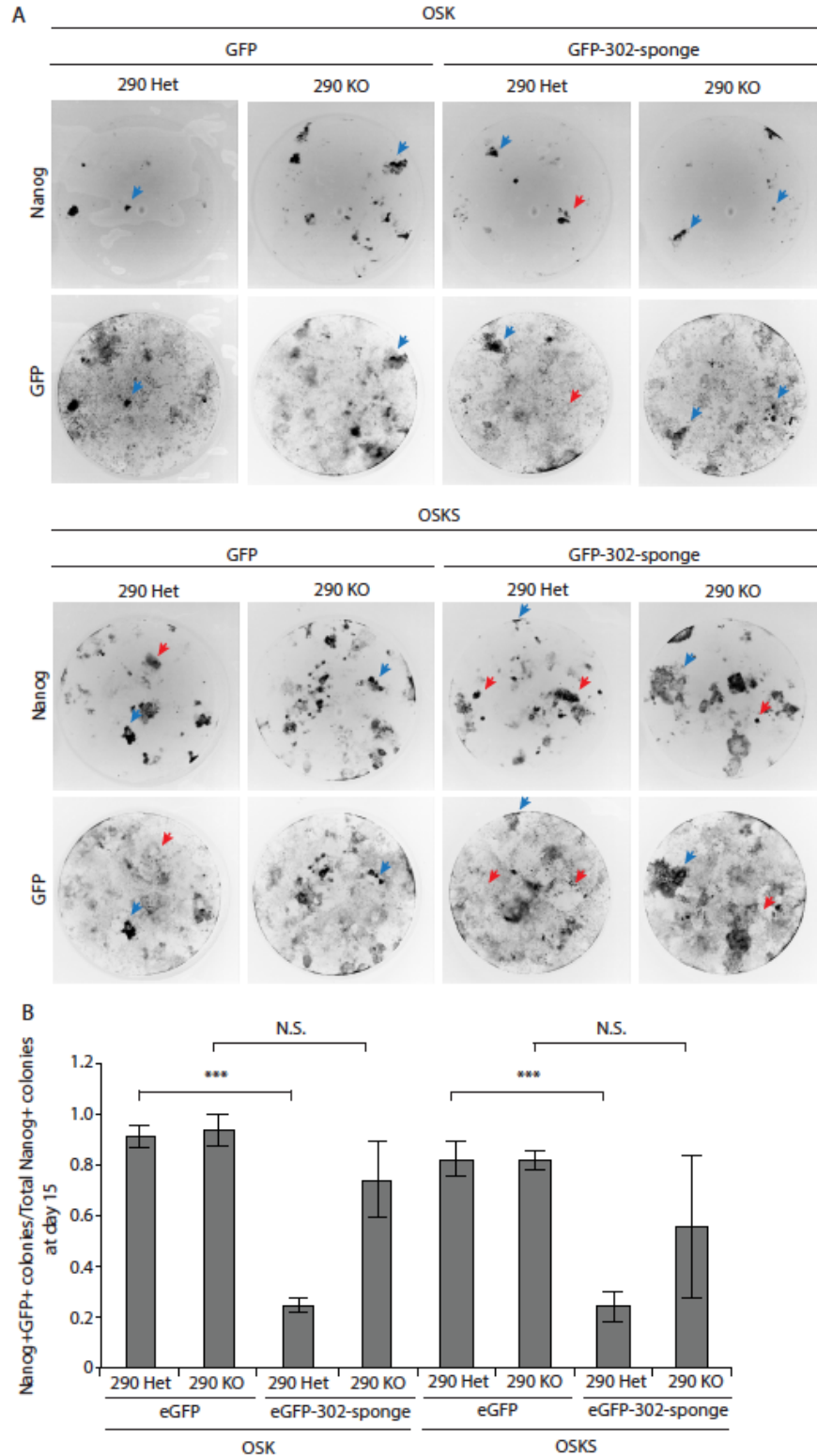


Figure 16. eGFP expression levels in iPSCs is consistent with miR-302 sponge function. (A) Representative images of individual Nanog-stained reprogramming wells at day 15. Blue arrows indicate Nanog+GFP+ colonies; red arrows indicate Nanog+GFP- colonies. (B) Quantification of day 15 Nanog-stained reprogramming plates of the fraction of Nanog+eGFP+ colonies (represents control cells or cells in which the *eGFP-mir-302-sponge* construct is not being fully suppressed). Error bars represent SD of 3-4 biological replicates. *** $P < 0.0005$.

silenced if there is there is a competitive disadvantage to removing both miR-290 and miR-302 family miRNAs from the cell.

Finally, we tested miR-302 sponge function in the iPSCs using a luciferase assay in which Renilla luciferase expression is controlled by a sequence perfectly complementary to miR-302b (Figure 17E). We expected that decreases in Renilla expression would indicate the presence of residual miR-290/miR-302 family miRNAs not bound to the miR-302 sponge, and we hypothesized that the perfectly complementary sequence would bind miR-302 strongly and therefore be a more sensitive readout of miR-302 sponge functionality compared with the imperfect seed sequence matches found in endogenous 3'UTRs. We saw that *mir-290* KO iPSCs had decreased Renilla luciferase activity compared to *mir-290* Het, which is consistent with the upregulation of miR-302 in *mir-290* KO cells. Importantly, *mir-290* Het or KO iPSCs containing *mir-302 sponge* had a small but statistically significant increase in luciferase activity compared to matched counterparts containing control construct, which is evidence that miR-302 sponge is successfully removing miR-290/miR-302 family miRNAs and therefore de-repressing Renilla luciferase. Interestingly, luciferase levels in *mir-290* KO + *mir-302 sponge* iPSCs was not higher than in *mir-290* Het + *mir-302 sponge* iPSCs, which suggests that there may be other miRNAs that are compensating for decreased miR-302 in the *mir-290* KO + *mir-302 sponge* iPSCs. Indeed, Renilla luciferase was partially repressed in all of the iPSCs compared to *Dgcr8* KO ESCs, which express no miRNAs and consequently represent maximal possible Renilla luciferase levels; therefore, even in iPSCs deficient in miR-290 and miR-302, there must still be other miRNAs that are capable of binding the luciferase construct.

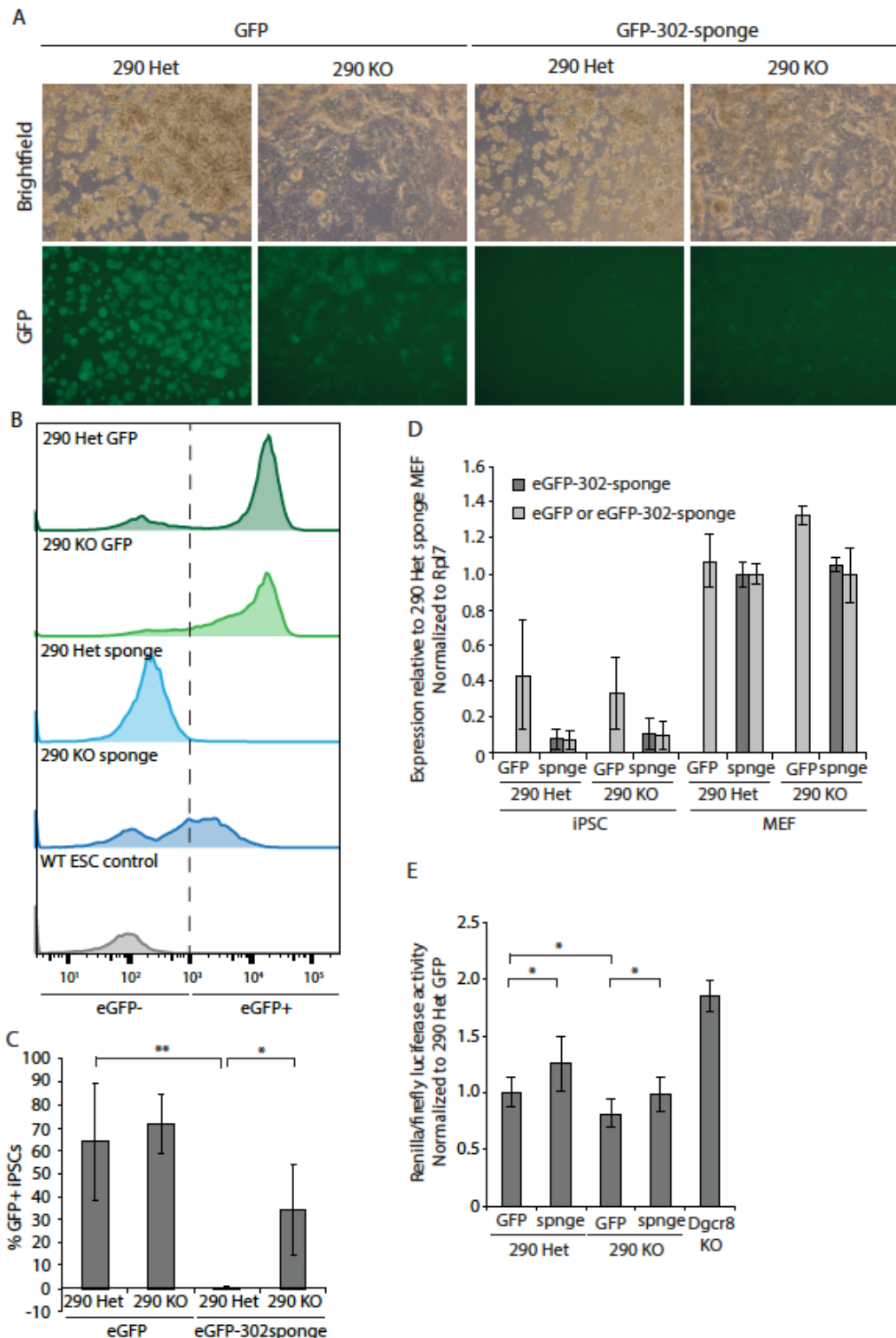


Figure 17. The *eGFP-miR-302-sponge* construct suppresses miR-290 and miR-302 family miRNAs. (A) Representative images of *eGFP* or *eGFP-miR-302-sponge*-treated *mir-290* Het or *mir-290* KO iPSCs. (B) Flow cytometric analysis of GFP expression levels in representative *eGFP* or *eGFP-miR-302-sponge*-treated *mir-290* Het or *mir-290* KO iPSCs. (C) Quantification of flow cytometry analysis in (B), with eGFP+ and eGFP- cells being arbitrarily separated by the dotted line in (B). (D) qRT-PCR analysis of *eGFP* or *eGFP-miR-302-sponge* transcript

expression levels in *mir-290* Het and *mir-290* KO MEF and iPSC lines using primers that can detect both transcripts (eGFP or eGFP-302-sponge) or the sponge construct specifically (eGFP-302-sponge). (E) Dual luciferase assay using a construct in which a sequence perfectly complementary to miR-302b is inserted 3' of Renilla luciferase. *Dgcr8* KO ESCs were used as a control for maximal luciferase expression. Error bars represent SD of 3-4 biological replicates for A, C, D; error bars represent SD of 7 biological replicates for E. *P < 0.05, **P < 0.005.

Discussion

In this study, we used genetic KO models of the miR-290 and miR-302 miRNA families to demonstrate that, despite their high expression in pluripotent cells and remarkable ability to enhance somatic cell reprogramming, they are not required for the reprogramming process in mouse cells. We were able to successfully generate Nanog⁺ iPSCs from *mir-290* KO MEFs as well as *mir-302* KO MEFs, and the absence of these miRNAs produced only fairly subtle phenotypes on the reprogramming process and the resulting iPSCs.

For instance, *mir-302* KO MEFs were less responsive to the addition of Sall4 to the OSK reprogramming mix than were control MEFs (Figure 9C and 9E). An established enhancer of reprogramming (Tsubooka et al., 2009, Parchem et al., 2014), Sall4 drives reprogramming intermediates to activate the *mir-302~367* locus (Parchem et al., 2014, Chapter 2). That *mir-302* KO MEFs do not generate Nanog⁺ colonies more efficiently when treated with OSKS compared to OSK suggests that the ability of Sall4 to increase reprogramming frequency and kinetics depends on the presence of miR-302. More generally, our data intriguingly implies that although somatic cells can traverse different paths consisting of different combinations of molecular changes occurring in different orders to arrive at the same final pluripotent state (Parchem et al., 2014), the particular series of molecular events any one cell takes can be limited by that cell's genetic makeup. Epigenetic status of the starting somatic cell, too, could affect the availability of paths to pluripotency; for example, the order of events possible for a de-differentiating B cell could be distinct from that of MEFs.

Although we did not detect any defect in the ability of *mir-290* KO MEFs to form Nanog+ iPSC colonies in the presence of either OSK or OSKS, we found that *mir-290* KO iPSCs exhibited a slight delay in the silencing of exogenous retroviruses (Figure 12A). This finding suggests that the *mir-290~295* cluster plays a role in the kinetics of the maturation phase of the reprogramming process (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007; Stadfeld et al., 2008). It may do so by directly influencing retroviral silencing, as miR-290 is known to target multiple epigenetic regulating factors (Benetti et al., 2008, Sinkkonen et al., 2008, Gruber et al., 2014). Another interesting finding was that both *mir-290* KO iPSCs and ESCs expressed higher levels of *mir-302~367* cluster miRNAs (Figure 12C). As miR-302 is activated early in the differentiation of ESCs, this data would be consistent with these cells being in a metastable state with some cells moving toward an EpiC state even under culture conditions containing 2i, which stabilizes mouse ESCs in the naïve ground state (Ying et al., 2008; Nichols et al., 2009; Marks et al., 2012). Indeed, miR-302 expression is seen in ESC cells grown in serum and LIF conditions (Houbaviv et al., 2003, Chen et al., 2007; Ciaudo et al., 2009, Parchem et al., 2014), which is also believed to result in a metastable ESC culture (Marks et al., 2012; Graf and Stadfeld, 2008; Hayashi et al., 2008). Therefore, miR-290 may be acting downstream of 2i in promoting a homogeneous ESC population. Indeed, recent single-cell sequencing experiments have shown that miR-290 promotes a transcriptionally homogeneous population under 2i conditions (Gambardella et al., 2017). Alternatively, miR-302 may be more directly part of a feedback mechanism with miR-290 and thus able to compensate for miR-290 in naïve cells. More consistent with the former possibility, both *mir-290* KO iPSCs and ESCs showed a defect in self-renewal and clonogenicity (Figure 12E), again a potential reflection of some cells in the population moving toward an EpiC state. Prior to our study, the only phenotype reported for *mir-*

290 KO ESCs was that they are sensitive to genotoxic stress (Zheng et al., 2011); therefore, our data reveal novel phenotypes for *mir-290* KO pluripotent cells.

We attempted to study the effect of functionally removing both miR-290 and miR-302 on reprogramming by putting a miR-302 sponge into *mir-290* KO cells. We expected that the sponge would unveil defects that might be masked in the *mir-290* KO condition due to the possible compensatory effects of miR-302. The absence of any impact on reprogramming efficiency in *mir-302* sponge-treated *mir-290* KO MEFs has a couple of possible interpretations: neither miR-290 nor miR-302 miRNAs are required for de-differentiation, or the sponge is insufficient for removing all functional miRNAs with that seed sequence. We have several pieces of evidence supporting sponge functionality (Figure 16 and 17), including a luciferase assay conducted using a sequence perfectly complementary to miR-302b to control Renilla luciferase expression (Figure 17E). The effect size is small, however, which suggests that either our assay is not sensitive enough to detect changes in miR-302 levels or that the absolute amount of miR-302 in *mir-290* KO cells is small and therefore changes in miR-302 are not functionally relevant. On the other hand, *mir-302* sponge-treated *mir-290* KO MEFs, unlike *mir-302* KO MEFs, did not show any defects in reprogramming enhancement in OSKS, which would suggest that the sponge did not create a functional *mir-302* KO; however, differences in the genetic backgrounds of *mir-290* KO and *mir-302* KO lines could also account for the variability that we observe when reprogramming with biologically distinct MEF lines. In the end, we were unable to conclusively determine whether our *mir-302* sponge-treated *mir-290* KO cells were truly completely deficient for both miR-290 and miR-302; it is likely that they are hypomorphic for miR-302. Whether simultaneous loss of both of these two miRNA families has any impact on somatic cell

reprogramming would ultimately be probably best addressed with an inducible double KO model, which is beyond the scope of this study.

Nevertheless, it is certain from our data that miR-290 and miR-302 are individually dispensable for generating mouse iPSCs. This result in itself is important and exceptional for two reasons. First is simply that miR-290 and miR-302 seem to play critical roles in development and induced pluripotency. Indeed, *mir-290* KO mice have a partially penetrant embryonic lethality (Medeiros et al., 2011), *mir-302* KO mice have a neural tube defect and are embryonic lethal (Parchem et al., 2015), and *mir-290/mir-302* double KO mice die of gross malformations by E9.5 (Parchem et al., 2015). It is important to note, however, that all of these KO embryos survive past the early epiblast stages that ESCs represent, suggesting that miR-290 and miR-302 are not required for the pluripotent state, even though they are highly expressed in wild-type situations. Although a recent study suggested that *mir-290~295* and *mir-302~367* are required for differentiation to the EpiC state, the defect appears slight and could a phenomenon limited to *in vitro* culture (Gu et al., 2016). In reconciling the different effects of these two miRNA families in development compared with reprogramming, it is also important to note that the enhancement effects of miR-290 and miR-302 overexpression on de-differentiation appear to occur early, during the initiation phase of reprogramming (Judson et al., 2013), not late in reprogramming when endogenous miR-290 and miR-302 are actually activated (Judson et al., 2009, Polo et al, 2012, Parchem et al., 2014). This observation suggests that, when ectopically expressed, these miRNAs likely target a set of mRNAs that is different from those that they bind under endogenous developmental contexts.

The second reason why our results are surprising is that there have been several studies showing a negative impact on reprogramming upon knockdown of miR-302~367 miRNAs in

mouse and human fibroblasts (Liao et al., 2011, Zhang et al., 2013) and upon knockout of the *mir-302~367* locus in human fibroblasts. Several possibilities could explain the discrepancy between those studies and ours. First, the antagomirs, TALE-KRAB repressors, and TALENs used to knock down or knock out the expression of *mir-302~367* miRNAs may have off-target effects on unrelated molecular and cellular processes. Our use of genetic KO models circumvents these technical issues. Second, our *mir-302* KO model does not remove *mir-367*, which has a seed sequence distinct from that of miR-302a-d and miR-290~295. The role of miR-367 in somatic cell reprogramming has not been fully investigated so far: knockdown of miR-367 with antagomiRs did not affect mouse reprogramming (Liao et al., 2011), yet the ability of miR-302~367 to induce mouse and human reprogramming independent of other factors requires the presence of miR-367 (Anokye-Danso et al., 2011). A genetic KO model for *mir-367* would be best for addressing this issue. Finally, because mouse ESCs are very different from human ESCs, the endpoints of the human and mouse reprogramming assays are difficult to compare. Human ESCs are molecularly more similar to mouse EpiSCs (Tesar et al., 2007, Brons et al., 2007), and miR-302 is the predominant miRNA in these two cell types (Suh et al., 2004, Bar et al., 2008, Jouneau et al., 2012). We attempted to reprogram *mir-302* Het and *mir-302* KO MEFs to into induced EpiSCs (Han et al., 2011) and were able to produce miR-302+ cells that upregulated EpiSC markers like *Fgf5*, *Gata6*, and *T* (data not shown); however, we were unable to reliably derive self-renewing EpiSCs for further characterization. Therefore, it is conceivable that miR-302 could be required for reprogramming to the mouse EpiSC/human iPSC state while being dispensable for generating mouse iPSCs.

In sum, our work on miR-290 and miR-302 in reprogramming clearly illustrates the complexity of miRNA function: the same miRNAs that have been implicated in a host of critical

activities in a cell type turn out to actually be largely dispensable for attaining and maintaining that same cell state. There is undoubtedly much still to be learned about how miRNAs physically and functionally interact with their mRNA targets—as well as with each other—in specific cellular contexts under endogenous conditions. Being cognizant and mindful of these intricacies will help us better understand how miRNAs help determine cell states and navigate transitions between them.

Materials and Methods

MEF generation

MEFs were generated as previously described (Judson et al., 2009). In brief, *mir-302*^{+/-}, *mir-302*^{-/-}, *mir-290*^{+/-}, and *mir-290*^{-/-} embryos were harvested at E13.5. After removal of the head and visceral tissue, the remaining tissue was dissociated by physical disruption and trypsinization and plated as Passage 0 (P0) cells in MEF medium (high glucose (H-21) DMEM, 10% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, and 55uM beta-mercaptoethanol). MEFs were expanded to P3 and frozen. For the *eGFP-mir-302-sponge* experiments, P0 *mir-290*^{+/-} and *mir-290*^{-/-} MEFs were infected with *pSin-eGFP* or *pSin-eGFP-mir-302-sponge* lentiviral constructs (Judson et al., 2013), sorted by FACS 2 days later (BD FACSAria3u), expanded and frozen 2-3 days after FACS.

Virus production

For retrovirus production, HEK293T cells grown in MEF medium were seeded at 2 million cells/10cm plate and transfected the following day with 5ug pCL-Eco and 5ug pMXs-expression plasmid with 30ul Fugene 6. Medium was replaced 24hr after transfection, and at 48hrs post-

transfection, supernatant was harvested, filtered (0.45um), aliquoted, and frozen at -80C. For lentivirus production, HEK293T cells were seeded at 4 million cells/10cm plate and transfected the following day with 4ug pSin-expression plasmid and 2ug each of VSVG, RSV, and MDL packaging vectors with 30ul Fugene 6. Supernatant was harvested, filtered, aliquoted, and frozen at -80C 72hrs post-transfection. Fresh aliquots were used for each experiment.

Reprogramming/de-differentiation

P4 MEFs were plated onto 0.2% gelatin-coated Greiner uClear black-walled 96-well imaging plates at 2000 cells/well or standard 12-well plates at 20,000 cells/well. The next day (day 0), cells were treated with 40ul (96-well) or 400ul (12-well) of each retrovirus aliquot and 4ug/ml polybrene. At day 1, the medium was replaced with fresh MEF medium and, if indicated, transfected with 100nM miR-302b mimic and Dharmafect1 transcription reagent per the manufacturer's protocol (GE/Dharmacon). Medium was replenished every other day with FBS+LIF medium (DMEM, 15% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, 55uM beta-mercaptoethanol, 1000U/ml LIF) between days 2 and 6. Thereafter, medium was replenished with KSR+LIF medium (Knock-out DMEM (Invitrogen), 15% Knock-out serum replacement (Invitrogen), non-essential amino acids, L-glutamine, penicillin/streptomycin, 55uM beta-mercaptoethanol, 1000U/ml LIF). Expression of the eGFP reporter in the *mir-302* KO allele was used to track reprogramming progression and kinetics, as the plates were imaged throughout reprogramming using the INCell Analyzer 2000 (GE). Reprogramming plates were analyzed for successful iPSC generation at day 15 by immunostaining and high-content, high-throughput imaging using the INCell Analyzer 2000 (GE).

ESC and iPSC culture and analyses

ESC (*mir-290*^{+/-} and *mir-290*^{-/-}) lines were derived from individual E3.5 blastocysts plated in one well of a 24-well dish on a MEF feeder layer using ES medium with 20% KSR and Erk1/2 inhibitor (1 μ M, PD0325901). Blastocyst outgrowths were trypsinized and passaged after 3-4 days until ESC lines were established. iPSC lines were generated from reprogramming experiments described above. ESC and iPSC lines were expanded and passaged under feeder-free conditions on 0.2% gelatin-coated plates in FBS+LIF+2i medium (DMEM, 15% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, 55 μ M beta-mercaptoethanol, 1000U/ml LIF, 1 μ M PD0325901, 3 μ M CHIR99021) and depleted of un- or partially-reprogrammed MEFs by transferring un-adhered cells to new plates 30min-1hr after passage. Gene expression in the iPSCs was examined by qRT-PCR after ~25-40 days in culture. For the *eGFP-mir-302-sponge* experiments, iPSCs were analyzed for eGFP expression by flow cytometry (BD LSR II). For ESC and iPSC colony formation/clonogenicity assays, cells were plated in 96-well plates at 500 cells/well, stained with Nanog antibody (CST8785), and Nanog⁺ colonies were counted. The ratio of Nanog⁺ colonies formed to the number of cells plated was defined as the colony formation potential.

qRT-PCR

Total RNA was collected using TRIzol according to the manufacturer's protocol. For cDNA synthesis, RNA was treated with DNaseI (Invitrogen) and reverse-transcribed with oligo-dT primers using the SuperScript III kit (Invitrogen). Total cDNA was diluted 1:5, and qPCR was performed using gene-specific primer sets (listed below) and SensiFast SYBR Hi-ROX master mix (Bioline). MiRNA qRT-PCR was performed with the polyA and SYBR Green method as

previously described using miRNA-specific forward primers and a 3' RACE adaptor reverse primer (Shi and Chiang, 2005). Primer specificity was verified through analysis of dissociation curves in experimental, no RT, and water-only samples.

Immunohistochemistry

Cells were fixed for 15min in 4% PFA, washed in PBST (PBS + 0.1% Triton X-100), incubated for 1hr at room temperature with blocking buffer (PBST, 1% goat serum, 2% BSA), then incubated overnight at 4C with 1:500 Nanog antibody (CST8785) in blocking buffer. Cells were then washed in PBST, incubated for 1hr at room temperature in 1:500 secondary antibody in blocking buffer (AlexaFluor 594 goat anti rabbit IgG), washed in PBST with 1:10000 Hoechst 33342 (Invitrogen), and stored in PBS before imaging.

Luciferase assay

A sequence perfectly complementary to miR-302b was cloned into the 3'MCS of the Renilla cassette in the psiCHECK2 luciferase vector (Promega). Cells were plated at 8000 cells/well in FBS+LIF+2i medium on gelatinized 96-well plates and transfected the following day with 100ng of the construct and 1ul Fugene 6 (Roche) in a final volume of 100ul/well. Cells were lysed and the luciferase assays were performed 24hr later using the Dual-Luciferase Reporter Assay System (Promega) on a dual-injecting SpectraMax L luminometer (Molecular Devices) according to the manufacturer's protocol. Luciferase activity was defined as the ratio of Renilla luciferase readings to firefly luciferase readings.

Primers

The following primers were used for qRT-PCR or DNA genotyping PCR.

Primer	Sequence
302 wt/mut genotyping F	CTCTTTGGGAGGCGGTCACG
302 wt genotyping R	GAGACAGAAAGCATTCCCATG
302 mut genotyping R	CTTGCCGTAGGTGGCATCGC
290 wt/mut genotyping F	TCCAGGTTTCCTTCAGGTTG
290 wt genotyping R	GATGGCCGCTACATAGGTGT
290 mut genotyping R	CGTGCAATCCATCTTGTTC
pSin-eGFP-mir-302-sponge genotyping F	GAGCAAAGACCCCAACGAGA
pSin-eGFP-mir-302-sponge genotyping R	GTAGCGGCCTTCTAAGTGCT
Endo-Oct4-qPCR F	TCTTTCCACCAGGCCCGGCTC
Endo-Oct4-qPCR R	TGCGGGCGGACATGGGGAGATCC
Endo-Sox2 F	TAGAGATAGACTCCGGGCGATGA
Endo-Sox2 R	TTGCCTTAAACAAGACCACGAAA
Endo-Klf4 F	GAATTGTGTTTCGATGATGC
Endo-Klf4 R	TCGCTTCCTCTCCTCCGACACA
Endo-Sall4 F	CAGCCTTATGCCCTTGGATA
Endo-Sall4 R	AGGGGTTGGAGGCATACTCT
Exo-Oct4 qPCR F	TCTCCCATGCATTCAAAGT
Exo-Sox2 qPCR F	CTGCCCTGTGCGCACATGTG
Exo-Klf4 qPCR F	CCTTACACATGAAGAGGCAC
Exo-OSK qPCR R	CTTTTATTTTATCGTCGACC
Exo-Sall4 qPCR F	GAGGAAAATAAGATTGCTGTCAGC
Exo-Sall4 qPCR R	CCTGACCTTGATCTGAACTTCT
Fgf5-qPCR F	CCTTGCGACCCAGGAGCTTA
Fgf5-qPCR R	CCGTCTGTGGTTTCTGTTGAGG
Gata6-qPCR F	TGACTCCTACTTCTTCTTCTC
Gata6-qPCR R	TACTTGAGGTCAGTCTCTCG
Klf4-qPCR F	TGTGGCAAACCTATAACCAAGAG
Klf4-qPCR R	CACAGCCGTCCCAGTCAC
Nanog-qPCR F	AACCAAAGGATGAAGTGCAAGCGG
Nanog-qPCR R	TCCAAGTTGGGTTGGTCCAAGTCT
Rex1-qPCR F	GAAAGTGAGATTAGCCCCGAG
Rex1-qPCR R	GTCCCCTTTGTCATGTACTCC
T/Brachyury-qPCR F	CTGGGAGCTCAGTTCTTTCGA
T/Brachyury-qPCR R	GAGGACGTGGCAGCTGAGA
Cdkn1a-qPCR F	GTCTGAGCGGCTGAAGATT
Cdkn1a-qPCR R	AAGACCAATCTGCGCTTGG
Lats2-qPCR F	TGACTGGATTTCAGGTGGAC
Lats2-qPCR R	CTGTCTCCACAGCGACAGTT
Tgfbr2-qPCR F	TGTGTGGAGAGCATGAAAGAC
Tgfbr2-qPCR R	CAGCACTCGGTCAAAGTCTC

pSin GFP or sponge-qPCR F	CACATGGTCCTGCTGGAGTT
pSin GFP or sponge-qPCR R	CCGAGGAGAGGGTTAGGGAT
pSin-eGFP-mir-302 sponge qPCR F	TTACCGGCGCGCTACTAAAA
pSin-3GFP-mir-302 sponge qPCR R	GTAGCGGCCTTCTAAGTGCT
Rpl7-qPCR F	GAACCAAAGCTGGCCTTTGTCATC
Rpl7-qPCR R	CAATGTATGGCTCCACAATCCGCA
Sno202 qPCR F	GTACTTTTGAACCCTTTTCCATCTGATG
miR-293 qPCR F	AGTGCCGCAGAGTTTGTAGTGT
miR-294 qPCR F	AAAGTGCTTCCCTTTTGTGTGT
miR-295 qPCR F	AAAGTGCTACTACTTTTGTAGTCT
miR-302a qPCR F	TAAGTGCTTCCATGTTTTGGTGA
miR-302b qPCR F	TAAGTGCTTCCATGTTTTAGTAG
miR-367 qPCR F	AATTGCACTTTAGCAATGGTGA
3' RACE adaptor outer for miRNA qPCR	GCGAGCACAGAATTAATACGACT

Chapter 4—NF45 and NF90/NF110 coordinately regulate ESC pluripotency and differentiation

Summary

While years of investigation have elucidated many aspects of embryonic stem cell (ESC) regulation, the contributions of post-transcriptional and translational mechanisms to the pluripotency network remain largely unexplored. In particular, little is known in ESCs about the function of RNA binding proteins (RBPs), the protein agents of post-transcriptional regulation. We performed an unbiased RNAi screen of RBPs in an ESC differentiation assay and identified two related genes, NF45 (Ilf2) and NF90/NF110 (Ilf3), whose knockdown promoted differentiation to an epiblast-like state. Characterization of NF45 KO, NF90+NF110 KO, and NF110 KO ESCs showed that loss of NF45 or NF90+NF110 impaired ESC proliferation and led to dysregulated differentiation down embryonic lineages. Additionally, we found that NF45 and NF90/NF110 physically interact and influence the expression of each other at different levels of regulation. Globally across the transcriptome, NF45 KO ESCs and NF90+NF110 KO ESCs show similar expression changes. Moreover, NF90+NF110 RNA immunoprecipitation (RIP)-seq in ESCs suggested that NF90/NF110 directly regulate proliferation, differentiation, and RNA-processing genes. Our data support a model in which NF45, NF90, and NF110 operate in feedback loops that enable them, through both overlapping and independent targets, to help balance the push and pull of pluripotency and differentiation cues.

This work has been submitted to *Nucleic Acids Research*.

Introduction

As discussed in Chapter 2, early mammalian development is a complex phenomenon driven by a myriad of dynamic and precise molecular changes that enable cells to establish,

maintain, and exit from the pluripotent state. Embryonic stem cells (ESCs) provide a valuable *in vitro* platform for dissecting these processes in detail, as they can both self-renew in culture indefinitely and differentiate into all of the tissue types of the body. ESC maintenance and differentiation are strongly affected by post-transcriptional mechanisms, which control alternative splicing, nuclear export, transcript stability, and translational efficiency. These functional activities are directed by non-coding RNAs and RNA binding proteins (RBPs), and the functions of RBPs in particular are poorly understood (Keene, 2007; Van Nostrand et al., 2016a). Building a more complete picture of the RBP-mediated molecular changes that regulate pluripotency is important, as it will enable the more informed use of stem cells for disease modeling, drug development, and regenerative medicine.

We chose to identify and characterize pluripotency-associated RBPs in a well-defined, developmentally relevant *in vitro* system—specifically, the transition from an ESC to an epiblast cell (EpiC) (Krishnakumar et al., 2016), a progression that parallels the earliest cell fate decision that occurs in the mammalian embryo proper. Through an unbiased siRNA screen for putative RBPs that affect the ESC-to-EpiC transition, we came to focus on two related genes, *Ilf2* (NF45) and *Ilf3* (NF90/NF110), the latter of which has a 90kDa isoform (NF90) and a 110kDa isoform (NF110). In addition to the EpiC-promoting phenotype their knockdown produced in our screen, these genes were interesting to us for several reasons. First, they have been implicated in pluripotency and development (Lu et al, 2009, Shi et al., 2005; Horie et al., 2011; Wang et al., 2006). Second, although they were first identified in a complex together as a transcriptional activator of IL2 in Jurkat T cells (Shaw et al., 1988; Corthesy et al., 1994; Kao et al., 1994), they were uncovered in all three mammalian “mRNA interactome capture” studies (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013), which provides strong evidence that they have RNA

binding functions *in vivo*. Indeed, NF45 and NF90/NF110 have been found to participate in a diverse assortment of post-transcriptional processes (Shi et al., 2005; Singh et al., 2012; Sakamoto et al., 2009; Chu et al., 2015; Wandrey et al., 2015; Merrill and Gromeier, 2006; Graber et al., 2010; Lee et al., 2011; Shim et al., 2002; Kuwano et al., 2010; Neplioueva et al., 2010; Guo et al., 2016), and their frequent co-occurrence therein suggests that they might interact with each other both physically and functionally. Importantly, none of the previous studies examining NF45 and NF90/NF110 molecular mechanisms were performed in ESCs, so we sought to investigate how NF45 and NF90/NF110 behave in the specific context of pluripotency and whether their activities are interconnected. Using genetic knockout models of these genes, we found that absence of NF45 or NF90+NF110 impaired ESC proliferation and differentiation down embryonic lineages. Additionally, we found that NF45 and NF90/NF110 interact with one another both physically and functionally, controlling a number of genes involved in proliferation, differentiation, and post-transcriptional processes either alone or in concert. Our data suggest that, through feedback loops, NF45, NF90, and NF110 modulate the expression levels of one another, thereby coordinately influencing the regulatory programs that govern pluripotency and differentiation.

Results

RNAi screen identifies NF45 and NF90/NF110 as promoters of pluripotency

To discover RBPs important for regulating pluripotency and differentiation, we performed a siRNA knockdown screen of 356 putative RBPs in an *in vitro* mouse ESC differentiation system that monitors the ESC-to-EpiC transition. The RBPs used in the screen either have been annotated as such in the literature (McKee et al., 2005) or are associated with RNA-related Gene Ontology (GO) terms. Additionally, they are differentially expressed in a

microarray study comparing ESCs with *in vitro* differentiated EpiCs (Parchem et al., 2014). The ESC differentiation system uses the fluorescent reporters described in Chapter 3 to track the expression of two miRNA clusters associated with different stages of pluripotency. Specifically, mCherry was knocked into the *mir-290~295* (*mir-290*) locus and eGFP into the *mir-302~367* (*mir-302*) locus (Parchem et al., 2014). When maintained in the naïve pluripotent conditions, these dual reporter ESCs express miR-290 exclusively and are mCherry+/eGFP-; when the ESCs are allowed to differentiate through the removal of Leukemia Inhibitory Factor (LIF) and GSK and MEK inhibitors (2i), they turn on *mir-302* and later shut off *mir-290*, thus becoming first mCherry+/eGFP+ and then mCherry-/eGFP+. The initial activation of *mir-302* is highly homogeneous and represents the acquisition of the primed pluripotent EpiC state (Krishnakumar et al., 2016). Therefore, we focused on this phase of the transition to identify genes whose knockdown might disrupt the kinetics, direction, or extent of the differentiation process.

For the screen, dual reporter ESCs transfected with a pool of 4 siRNAs against each gene of interest were induced to differentiate and assessed for their miR-290-mCherry and miR-302-eGFP expression levels after 3 days using high-content, high-throughput microscopy (Figure 18A). The screen yielded a number of hits. Focusing only on the 146 genes whose knockdown did not negatively affect cell proliferation and viability (defined in the Materials and Methods), we found 49 genes whose knockdown significantly changed the expression levels of miR-290-mCherry or miR-302-eGFP (Figure 18B). Many of these genes (35 genes) led to a reduction in miR-302-eGFP expression, which represents an inhibition of the ESC-to-EpiC transition. These genes were candidate promoters of differentiation, but we reasoned that the reduced miR-302-eGFP levels could also represent a general nonspecific manifestation of cellular toxicity that might occur when any important cellular gene is disrupted. More interestingly, there was only

one gene, NF45 (Ilf2), whose knockdown resulted in increased miR-302-eGFP expression (Figure 18B). This result therefore suggests that NF45 may normally function to help maintain naïve pluripotency.

There is much literature on NF45 acting as a post-transcriptional regulator in addition to evidence of its involvement in the pluripotency network, as it was found to be complexed with Nanog (Wang et al., 2006) and is downregulated in a Nanog-knockdown model of ESC differentiation (Lu et al., 2009). These pieces of data highlighted the potential significance of

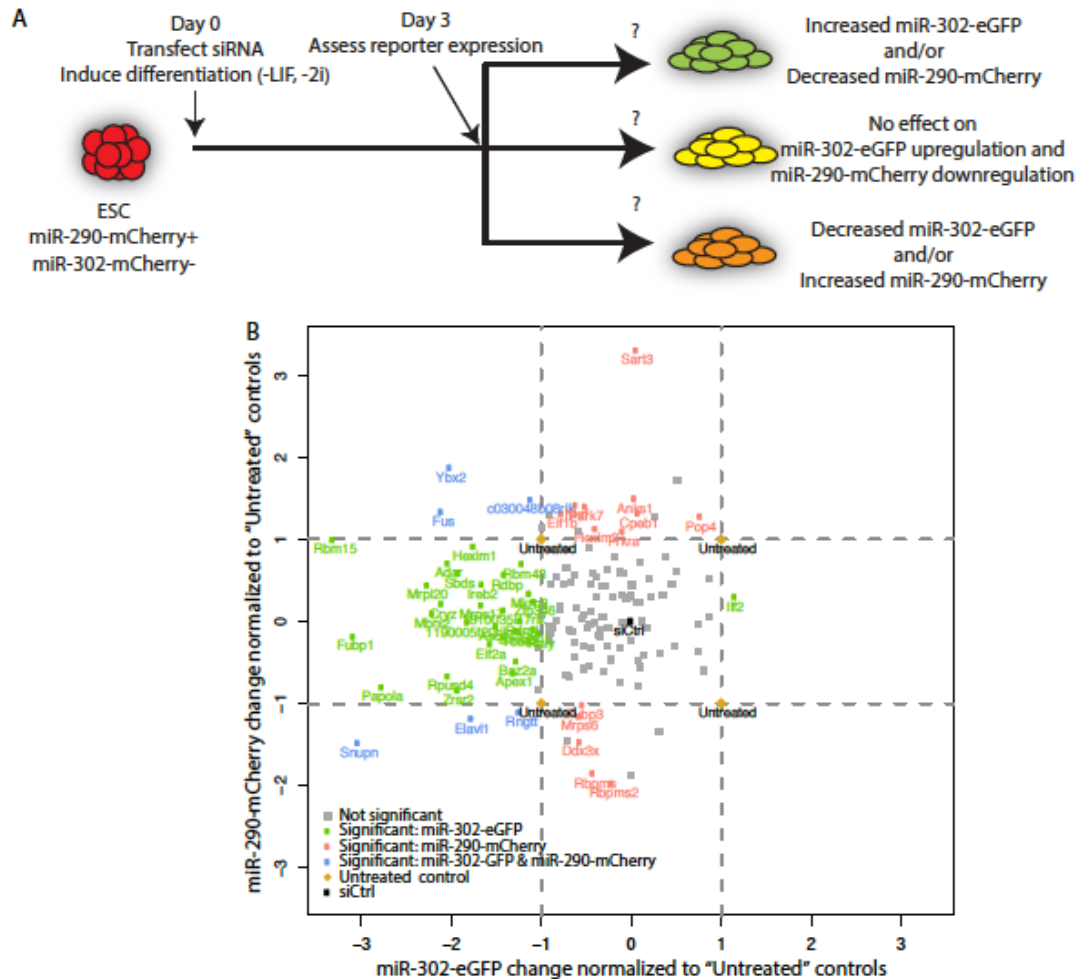


Figure 18. An siRNA screen identifies NF45, NF90, and NF110 as promoters of pluripotency. (A) Schematic of siRNA screen design. (B) Scatterplot of siRNA screen hits with normalized t-statistics for miR-302-eGFP and miR-290-mCherry on the x and y axes, respectively. Coloring denotes genes whose knockdown significantly affected levels of miR-302-eGFP, miR-290-mCherry, or both ($q < 0.001$) after first z-score normalizing from a center of zero as defined by siCtrl-treated samples and then scaling by the t-statistic of the untreated samples. Genes whose knockdown negatively affected cell proliferation and viability ($n = 210$) were removed from this analysis. Note that some genes with large t-statistics are not called significant due to cutoffs described in the Materials and Methods.

NF45; therefore, we chose to focus our subsequent investigations on this gene. To validate the NF45 knockdown phenotype, we performed flow cytometry on cells treated with siRNAs against NF45 and confirmed the increase in miR-302-eGFP expression compared with the cells treated with siCtrl non-targeting siRNA at 72 hours of differentiation (Figure 19A and 19B). Using qRT-PCR, we also confirmed the increase in the levels of processed miR-302 miRNAs (Figure 19C). As NF45 is known to physically interact with NF90/NF110 (Guan et al., 2008; Wolkowicz et al., 2012), which were not included in the original siRNA screen, we asked how depletion of NF90/NF110 would affect miR-302 expression. siRNAs against NF45, NF90+NF110, and NF110 downregulated their targets to similar levels (~20-50% of wt) (Figure 19D and 19E). Knockdown of either NF90+NF110 or NF110 alone promoted miR-302 expression (Figure 19A-C), although the increase in miRNA levels compared to siCtrl was less dramatic than seen with NF45 knockdown (Figure 19C). Therefore, these data suggest that decreased levels of NF45, NF90, and/or NF110 are associated with a possible accelerated shift toward the EpiC fate during ESC differentiation.

NF45 and NF90/NF110 regulate ESC proliferation and differentiation

We next employed genetic knockout models to further dissect the functions of NF45 and NF90/NF110 in pluripotency and differentiation. NF45^{-/-} (NF45 KO) ESCs created through a gene-trap system were obtained from Horie et al. (2011). As the wild-type (NF45 WT) control, we used their “revertant” counterpart, an ESC line in which the gene-trap construct was removed through FLP-FRT recombination. ESCs deficient for both NF90 and NF110 (NF90+NF110 KO) or for NF110 alone (NF110 KO) were generated by CRISPR-mediated mutagenesis (Figure

20A); the parental line (NF90+NF110 WT) served as their control. We confirmed the loss of the appropriate proteins in all KO ESC lines (Figure 20B and 20C).

To determine whether NF45 and NF90/NF110 are required for ESC self-renewal and proliferation, we performed colony formation and proliferation assays. We found that NF45 KO,

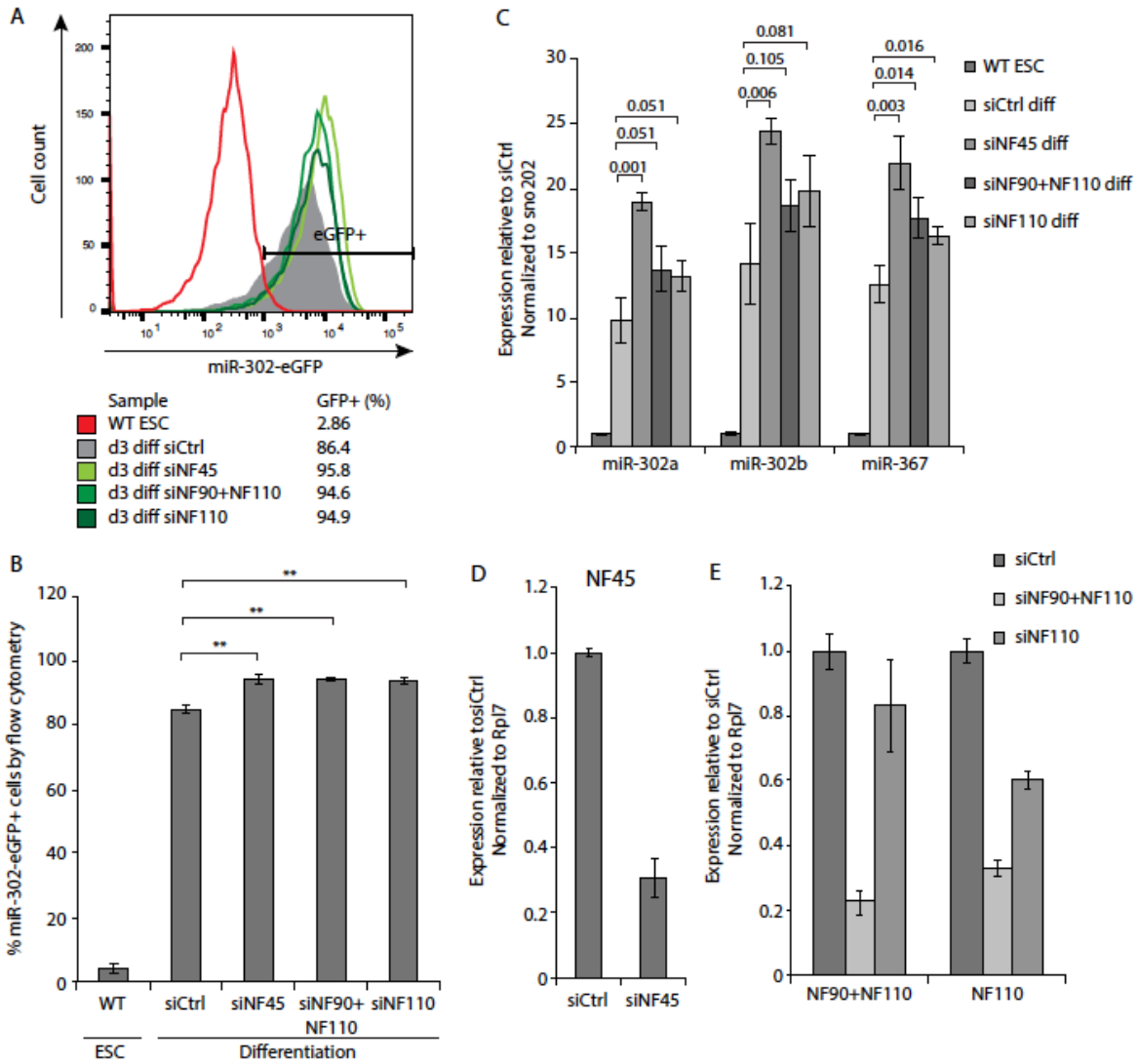


Figure 19. Validation of siRNA screen. (A) Flow cytometry of miR-302-eGFP expression of ESCs and ESCs differentiated for 3 days in the presence of control siRNA (siCtrl) or siRNAs against *Ilf2*/NF45 (siNF45), NF90+NF110 (siNF90+NF110), or NF110 only (siNF110). (B) Quantification of miR-302-eGFP+ cells from (A). (C) Transcript levels by qRT-PCR of mature miR-302 miRNAs in ESCs treated with the indicated siRNAs and differentiated for 3 days. Expression analysis by qRT-PCR of (D) NF45 and (E) NF90 and NF110 in ESCs or ESCs treated with the indicated siRNAs and differentiated for 3 days. Error bars represent SD of 3 biological replicates. Numbers above the graph in (C) indicate the p-values of the comparisons marked. **p < 0.005.

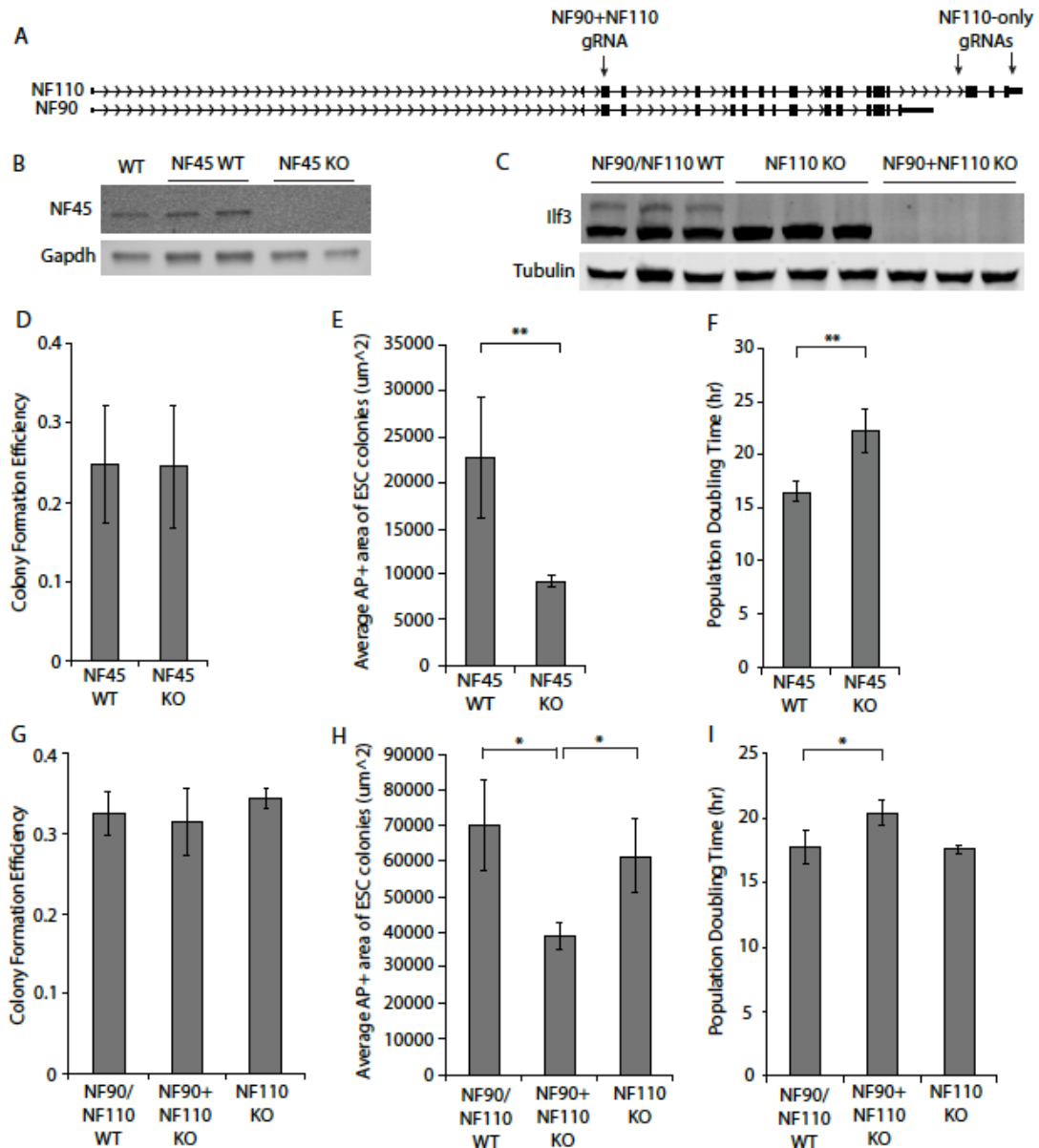


Figure 20. NF45 and NF90 promote ESC proliferation. (A) Schematic showing location of gRNAs for generating NF90+NF110 and NF110 only KO ESCs. (B) Western blot showing protein expression of WT (dual reporter ESCs), NF45 WT, and NF45 KO ESCs. (C) Western blot showing protein expression in NF90+NF110 WT, NF110 KO, and NF90+NF110 KO ESCs. (D) Colony formation efficiency, (E) area of alkaline phosphatase staining per colony, and (F) population doubling time of NF45 WT and NF45 KO ESCs. (G) Colony formation efficiency, (H) area of alkaline phosphatase staining per colony, and (I) population doubling time of NF90/NF110 WT, NF90+NF110 KO, and NF110 KO ESCs. Error bars represent SD of 3-4 biological replicates. * $p < 0.05$, ** $p < 0.005$.

NF90+NF110 KO, and NF110 KO ESCs have colony forming potentials similar to their WT counterparts (Figure 20D and 20G), indicating that their loss does not affect ESC self-renewal. However, NF45 KO ESCs form smaller colonies (Figure 20E) and have a longer doubling time

(Figure 20F), suggesting a decreased proliferative capacity consistent with a previous report (Horie et al., 2011). Notably, NF90+NF110 KO ESCs also exhibit a proliferation defect of comparable magnitude, but NF110 KO ESCs are unaffected (Figure 20H and 20I). These results show that both NF45 and NF90 promote ESC proliferation. It is possible that NF110 also promotes ESC proliferation but that NF90 can function redundantly to compensate for its loss; however, we were unable to test this idea with our genetic knockout models.

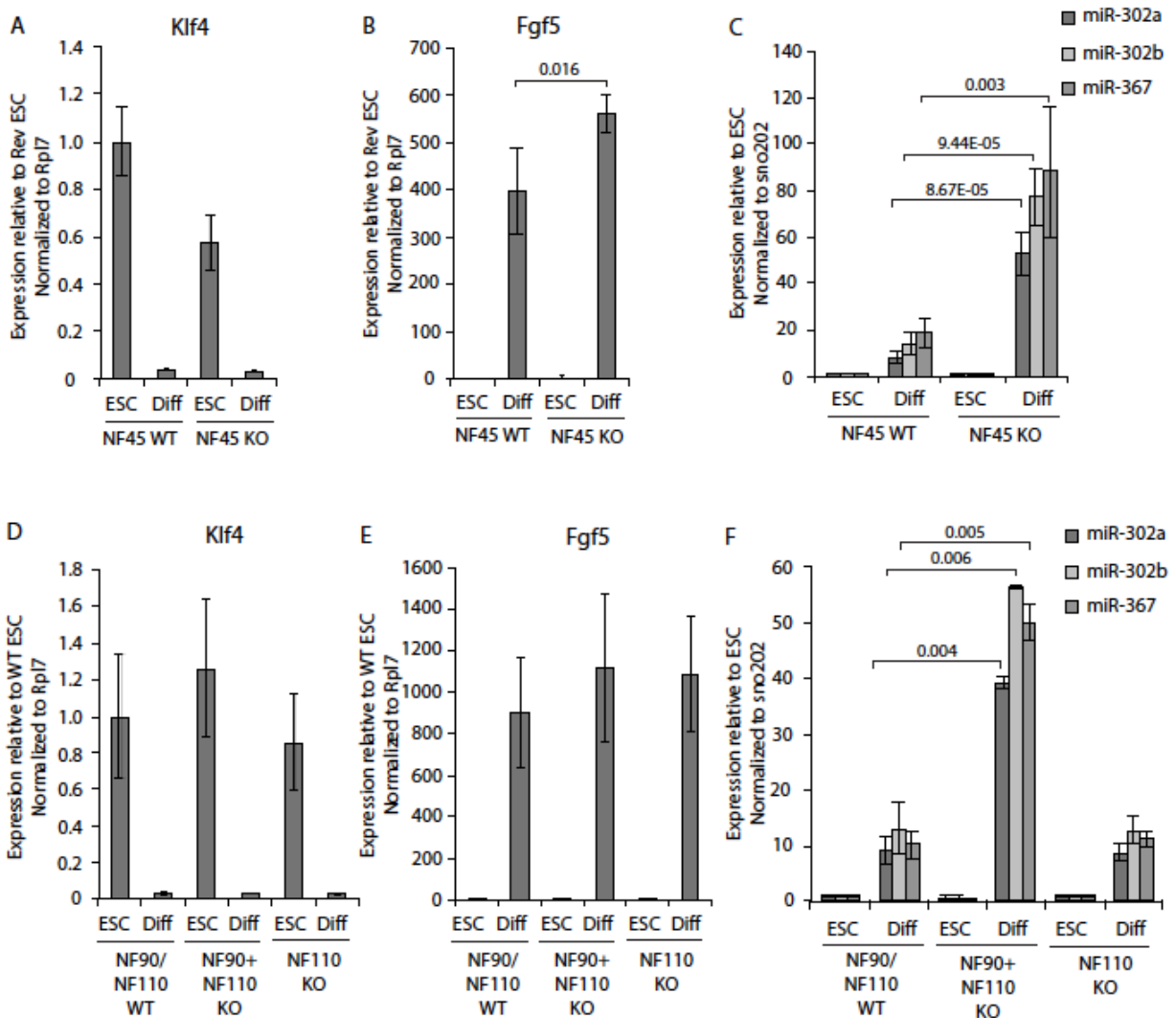


Figure 21. Loss of NF45 and NF90/NF110 dysregulate differentiation to EpiCs. Transcript levels by qRT-PCR of (A, D) *Klf4*, (B, E) *Fgf5*, and (C, F) miR-302 family miRNAs in NF45 WT, NF45 KO, NF90+NF110 WT, NF90+NF110 KO, and NF110 KO ESCs and 2D differentiated epiblast cells. Error bars represent SD of 3-4 biological replicates. Numbers above graphs indicate the p-values of the comparisons marked.

We then assayed the mRNA levels of a selection of pluripotency and differentiation markers to see if to ask whether the propensity of NF45- and NF90/NF110-deficient ESCs to differentiate to the EpiC state begins with an inherent defect in the self-renewing state. We found that NF45 KO ESCs and NF90/NF110 KO ESCs did not display obvious dysregulation of these markers; although NF45 KO ESCs expressed slightly elevated EpiC markers, these levels were still far lower than those seen during differentiation into EpiCs (data not shown).

Having examined the roles of NF45, NF90, and NF110 in naïve pluripotency, we next sought to better characterize how loss of these genes affects the molecular network when ESCs are induced to differentiate. Using the 2D ESC-to-EpiC differentiation assay employed for the siRNA screen, we found that NF45 KO ESCs appropriately downregulated ESC markers like *Klf4* (Figure 21A) but overactivated EpiC markers like *Fgf5* and *miR-302* (Figure 21B and 21C). Although NF90+NF110 KO and NF110 KO ESCs showed a similar trend in gene dysregulation (Figure 21D-F), the degree to which EpiC markers were upregulated was milder, and only NF90+NF110 KO but not NF110 KO ESCs exhibited an overactivation of *miR-302*.

In order to more carefully dissect the impact of these genes upon exit from pluripotency, we studied the effects of NF45, NF90, and NF110 loss in a 3D embryoid body differentiation assay, which allows for undirected differentiation to the three embryonic germ layers. As in the 2D system, NF45 KO ESCs downregulated ESC markers (*Klf4*) appropriately (Figure 22A) while overactivating EpiC markers (*Fgf5*, Figure 22B). However, they were unable to activate markers of ectoderm (*Pax6*), mesoderm (*T/Brachyury*), or endoderm (*Gata6*) (Figure 22C-F), suggesting that while NF45 KO ESCs are predisposed to achieving the primed EpiC state, they cannot properly differentiate further to form the embryonic lineages. While lineage

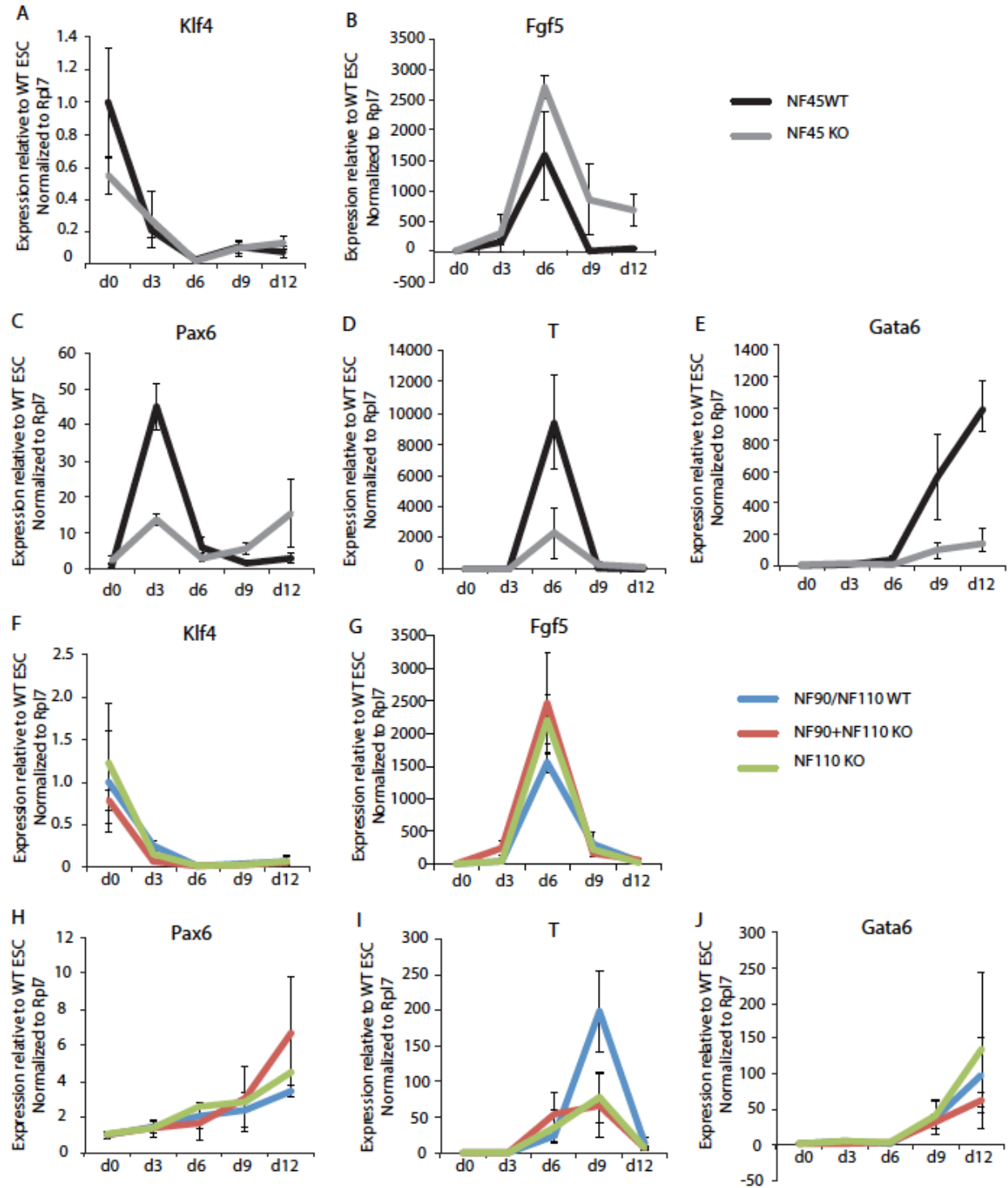


Figure 22. Loss of NF45 and NF90/NF110 dysregulate differentiation down embryonic lineages. Transcript levels by qRT-PCR of (A, F) Klf4, (B, G) Fgf5, (C, H) Pax6, (D, I) T/Brachyury, and (E, J) Gata6 in NF45 WT, NF45 KO, NF90/NF110 WT, NF90+NF110 KO, and NF110 KO ESCs (d0) and EBs (d3-d12). Error bars represent SD of 3-4 biological replicates.

differentiation also showed signs of dysregulation in NF90+NF110 KO and NF110 KO ESCs (Figure 22F-J), the defects, especially in the NF110 KO ESCs, were different and less pronounced than those in the NF45 KO ESCs. Specifically, NF90+NF110 KO ESCs appear to have a particular defect in activating the mesoderm program, which is indicated by their failure to activate T/brachyury expression (Figure 22I). In summary, our focused molecular and cellular analyses revealed that NF45 and NF90/NF110 have dissimilar effects on differentiation, where loss of NF45 resulted in more dramatic impairments in differentiation than loss of NF90+NF110 or NF110 alone. These results are consistent with the observation that NF90/NF110 KO mice survive to birth (Shi et al., 2005), while NF45 KO mice exhibit embryonic lethality (<http://www.informatics.jax.org/marker/MGI:1915031>).

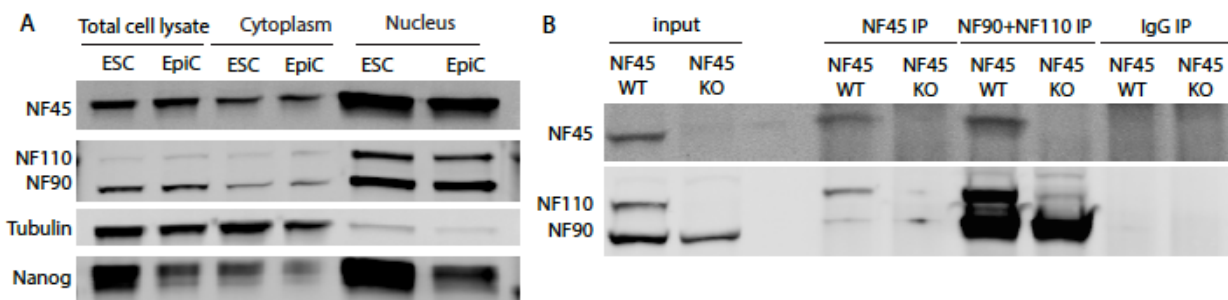


Figure 23. NF45 and NF90/NF110 are largely localized to the nucleus and physically interact. (A) Subcellular localization of NF45, NF90, and NF110 proteins in WT ESCs and EpiCs (3 days of -LIF, -2i diff). Nanog was used as a nuclear marker as well as an indicator for differentiation progression. Equal amounts of protein were loaded per well. (B) Western blots showing co-IP interactions between NF45 and NF90/NF110. Input lanes show 2% input.

NF45 and NF90/NF110 form complexes and influence expression levels of each other

Given the partially overlapping molecular and cellular phenotypes of their knockout ESC lines, we tested if and how NF45 and NF90/NF110 are physically or functionally interconnected. First, we examined the subcellular localization of NF45, NF90, and NF110 by fractionation. We found that all proteins are present in both the cytoplasm and the nucleus, although they are largely concentrated in the nucleus (Figure 23A). Expression and localization of these proteins

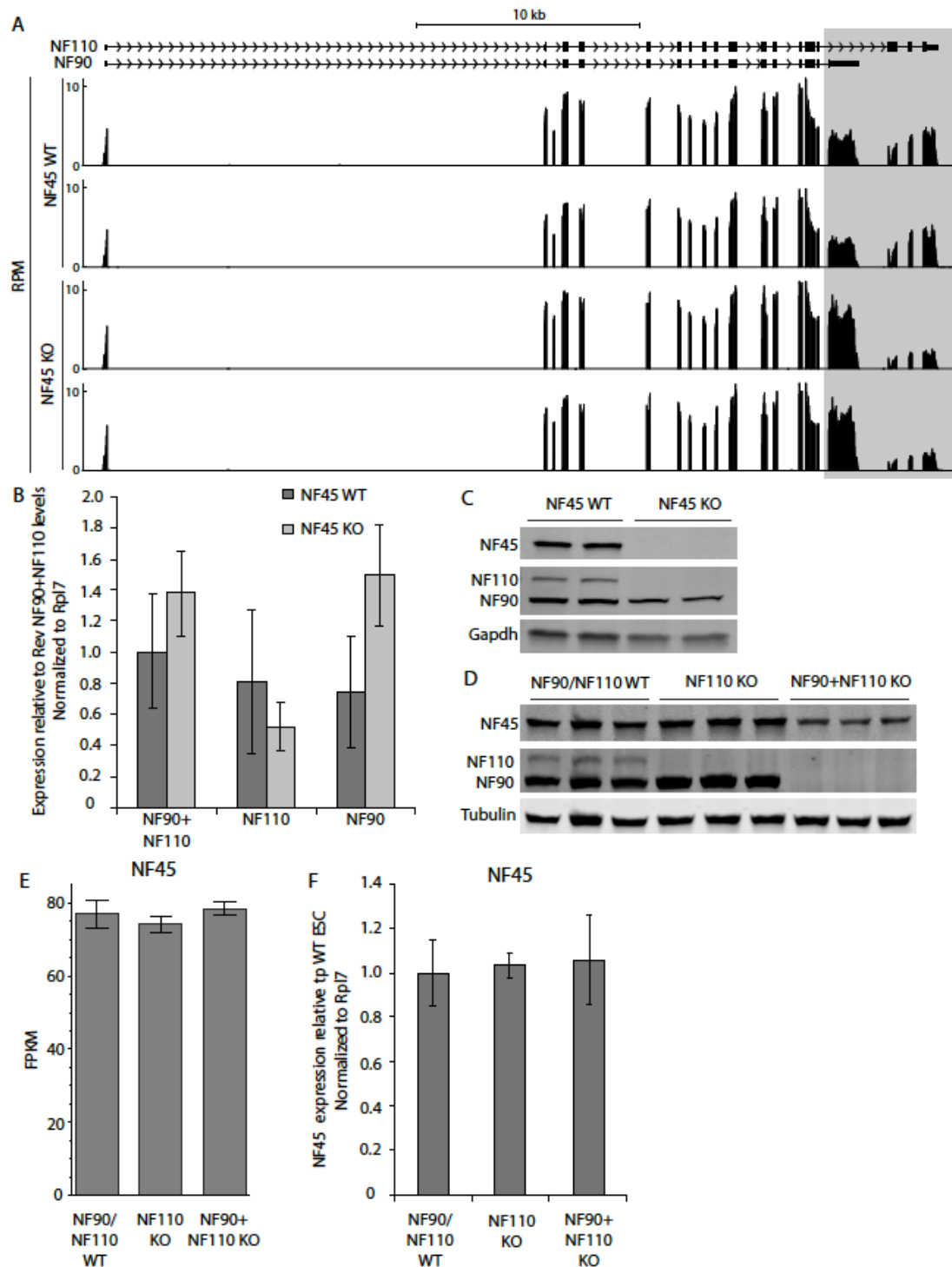


Figure 24. NF45 and NF90/NF110 influence the expression of each other at different levels of regulation. (A) Read coverage in RPM (reads per million mapped reads) of NF90/NF110 transcript in NF45 WT and NF45 KO ESCs. Differences between the NF90 and NF110 isoforms are highlighted. (B) Transcript levels by qRT-PCR of NF90 and NF110 in NF45 WT and NF45 KO ESCs. (C, D) Western blots showing NF45, NF90, and NF110 protein expression levels in the KO ESC lines indicated. (E) Mean FPKM (fragments per kilobase of transcript per million mapped reads) of NF45 transcript in NF90/NF110 WT, NF110 KO, and NF90+NF110 KO ESCs. (F) Transcript levels by qRT-PCR of NF45 in NF90/NF110 WT, NF110 KO, and NF90+NF110 KO ESCs. Error bars in (B) and (F) indicate SD of 3-4 biological replicates; error bars in (E) indicate standard error of the mean.

did not change with differentiation to the EpiC state (Figure 23A). Their intracellular localization suggests that both NF45 and NF90/NF110 likely participate predominantly in nuclear processes.

We then asked whether NF45 and NF90/NF110 interact with and influence the expression of each other in ESCs. Indeed, work in other cellular systems frequently discover both NF45 and NF90/NF110 in the processes being studied, and X-ray crystallography has shown that these proteins can heterodimerize with each other through the dsRNA-binding domain associated with zinc fingers (DZF) motif that both proteins possess (Wolkowicz et al., 2012). Co-immunoprecipitations showed that NF45 and NF90/NF110 physically interact with each other in our ESCs (Figure 23B). Intriguingly, we found that loss of either NF45 or NF90/NF110 impacts the expression of the other at different stages of molecular regulation. NF45 deletion leads to a splicing isoform switch of NF90/NF110, with transcripts of the NF90 isoform increasing in abundance at the expense of the NF110 isoform (Figure 24A and 24B). The effects of this isoform switch are further amplified at the protein level, where NF110 protein is entirely absent in NF45 KO cells (Figure 24C), which suggests that in addition to controlling the balance of NF90 and NF110 mRNA levels, NF45 also positively regulates NF110 protein stability. NF90 protein levels are also reduced in the NF45 KO but to a much smaller degree than the NF110 isoform. Conversely, simultaneous loss of NF90 and NF110 but not loss of NF110 alone leads to a decrease in NF45 protein levels (Figure 24D), while NF45 transcript levels are unaffected (Figure 24E and 24F). This observation suggests that NF90 promotes NF45 protein stability, although NF110 may also play a redundant role. These interactions suggest not only that NF45 and NF90/NF110 have overlapping phenotypes but also that being physically complexed allows them to positively regulate each other's protein levels.

Genome-wide expression analysis supports functional interactions between NF45 and NF90/NF110

To further investigate the functional interactions of NF45, NF90, and NF110, as well as their roles in ESC proliferation and differentiation, we analyzed global gene expression in NF45 KO, NF90+NF110 KO, and NF110 KO ESCs by RNA-seq. Examining the transcriptomes of these KO ESC lines in relation to their corresponding WT controls, we identified 1670 genes that are differentially expressed in at least one comparison (Figure 25A). By requiring the genes to be robustly expressed (mean FPKM ≥ 10) in at least one condition, we retained 971 genes that represent candidate regulatory targets, both direct and indirect, of each protein.

Grouping these 971 genes by unsupervised hierarchical clustering, we identified 9 expression modules (Figure 25B). We focused on genes that were consistently upregulated (Module III) or downregulated (Module IX) in all three KO ESC lines, as they represent genes that NF45 and NF90/NF110 coordinately regulate either negatively or positively. Gene Ontology analysis revealed that genes in Module III (upregulated in all lines) show an overrepresentation of biological processes related to development and differentiation, such as the determination of left/right symmetry, placenta development, and cell migration, while genes in Module IX (downregulated in all lines) are highly biased towards cell cycle and cell division related genes (Figure 25C). These results are broadly consistent with our phenotypic data showing that deletion of NF45 or NF90/NF110 generally resulted in decreases in ESC proliferation and defects in differentiation.

Interestingly, the overall global transcriptomic changes associated with loss of NF45 resemble those associated with the loss of NF90+NF110 KO as seen with correlation analyses performed at the individual gene level (Figure 26A, top) and when aggregated into KEGG

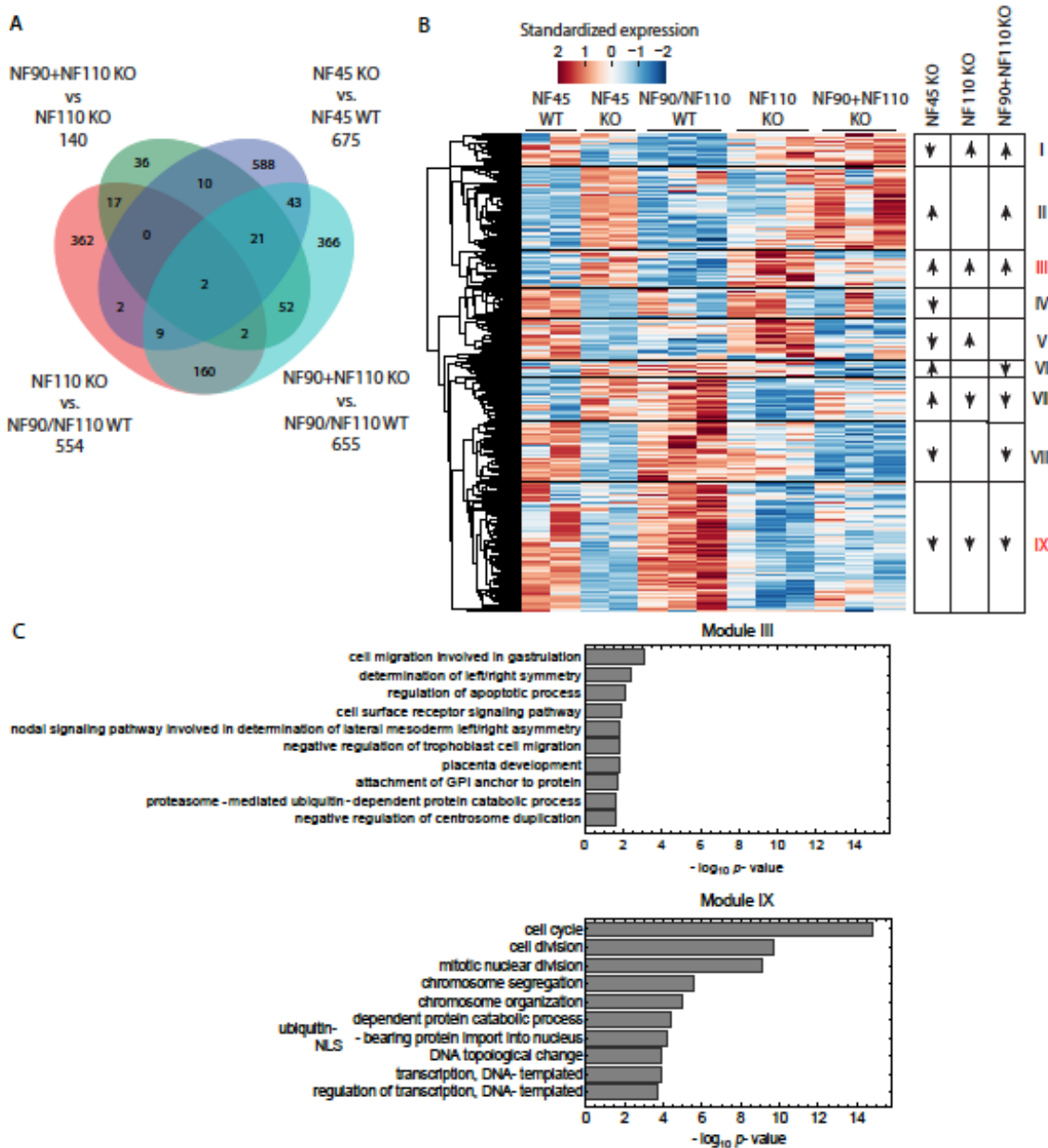


Figure 25. Global expression analysis is consistent with functional interactions among NF45, NF90, and NF110. (A) Venn diagram of differentially expressed genes in RNA-seq analysis of NF45 WT, NF45 KO, NF90+NF110 WT, NF90+NF110 KO, and NF110 KO ESCs. (B) Heat map of robustly expressed genes that are differentially expressed in at least one comparison considered in (A). Genes are categorized through unsupervised clustering into 9 modules based on their pattern of expression in the different lines. (C) GO analysis of category III and IX genes as defined in (B).

pathways (Figure 26B, top) or GO terms (Figure 26C, top). This similarity is absent when comparing NF45 KO with NF110 KO (Figure 26A-C, bottom), which suggests either that NF90 is the primary functional isoform or that NF90 and NF110 act redundantly. Combined with our

phenotypic data, our results indicate that the molecular and cellular impairments of NF45 KO ESCs are similar yet not identical to that of NF90+NF110 KO ESCs, suggesting both collaborative and independent functions among the three proteins.

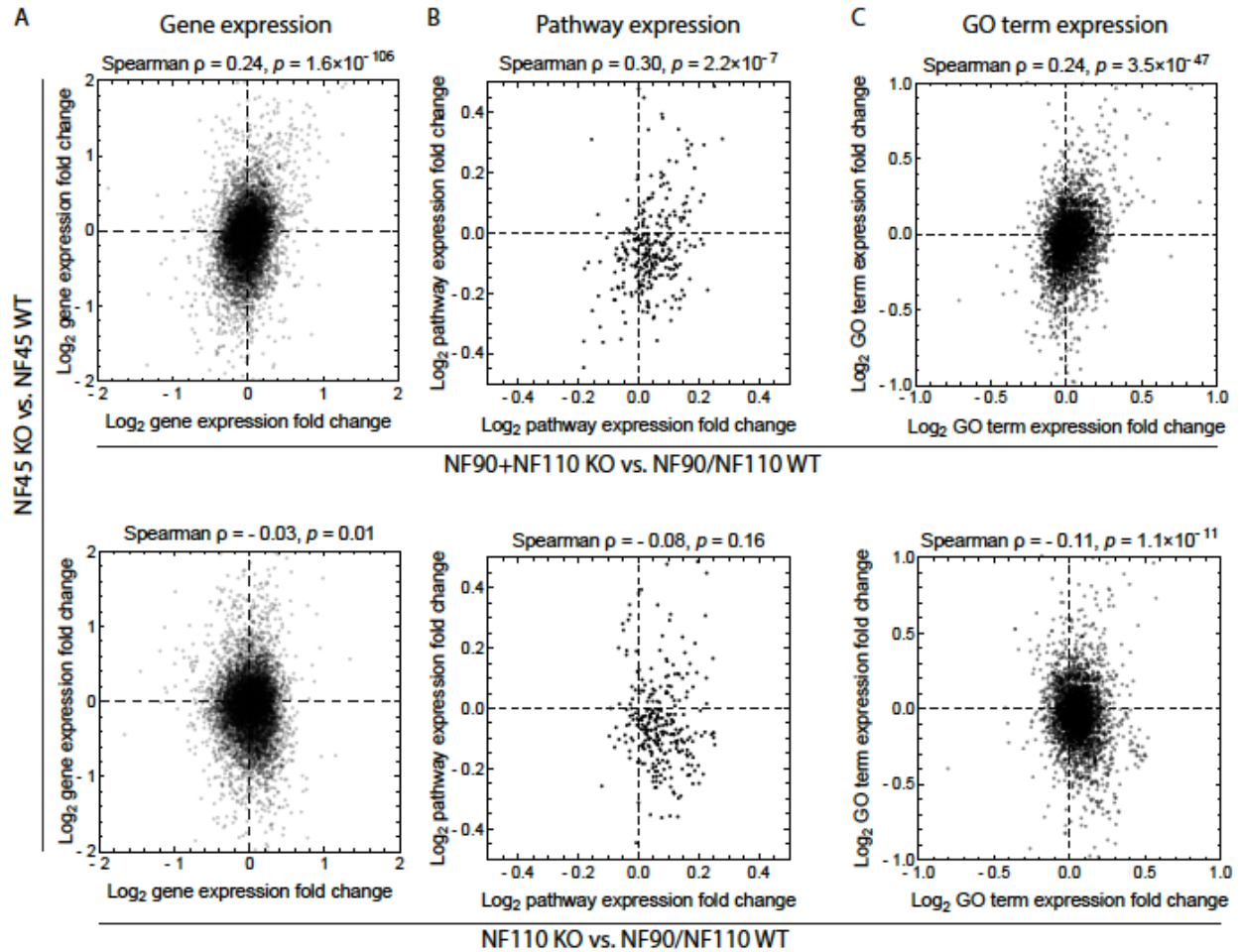


Figure 26. Transcriptomic changes in NF45 KO correlate with NF90+NF110 KO but not NF110 KO. Correlation analyses of NF45 KO/NF45 WT vs. NF110 KO/NF90+NF110 WT or NF90+NF110 KO/NF90+NF110 WT transcriptomes with respect to (A) individual genes, (B) molecular pathways, and (C) GO terms.

Identification of NF90/NF110 RNA targets

Although our transcriptomic analyses suggested overlapping roles for NF45 and NF90/NF110 in mRNA regulation, they could not distinguish between the primary and secondary effects of the proteins. Therefore, we sought to identify the combined direct RNA targets of NF45 and NF90/NF110. To do so, we performed NF90/NF110 RNA

immunoprecipitation and sequencing (RIP-seq) in NF110 KO, NF90+NF110 KO, and NF90+NF110 WT ESCs, expecting that the immunoprecipitated NF90/NF110-bound RNAs would also include transcripts bound by NF45-NF90/NF110 complexes but not alternative NF45 complexes. Western blot analysis of input and RIP samples confirmed that immunoprecipitation of NF90/NF110 pulled down the expected proteins in each sample and that targets of NF45-NF90/NF110 complexes would also be included in our datasets (Figure 27A). To distinguish between RNA targets of NF90 and NF110, we compared the RIP-seq datasets of NF110 KO, NF90+NF110 KO, and WT ESCs (Figure 27B). In principle, the WT dataset contains both NF90 and NF110 targets; the NF110 KO dataset contains only NF90 targets, including those that could be redundantly bound by NF110 or NF90 (Figure 27B, top panel); and the NF90+NF110 KO dataset contains non-specific targets of the NF90/NF110 antibody and experimental noise, both of which are also present in the WT and NF110 KO data. Thus, removing genes found in the NF110 KO RIP-seq data from the list of genes found in the WT RIP-seq data yields targets bound only by NF110 (“NF110 only targets”); a similar comparison of WT and NF90+NF110 KO data yields targets bound by NF90 and/or NF110 (“NF90+NF110 targets”), and a comparison of NF110 KO and NF90+NF110 KO data yields targets bound by NF90 that could also be redundantly bound by NF110 (“NF90 targets,” Figure 27B, bottom panel).

We developed two complementary computational approaches to identify NF90 and NF110 targets based on these comparisons. One utilizes the stochastic nature of count data but treats the IP and input data as independent samples (Poisson ratio test), while the other takes into account the paired nature of the IP and input samples but does not make full use of the distributive properties of count data (log ratio test). We used a multi-objective optimization algorithm based on the concept of Pareto dominance (Diaz et al., 2015) to combine results from

the two methods, obtaining 208, 81, and 167 genes in NF110 only targets, NF90 targets, and NF90+NF110 targets, respectively (Figure 27C). Notably, NF90 targets previously identified in HeLa cells (Kuwano et al., 2010) also showed IP to input enrichment in our NF90 target identification method (data not shown), which demonstrates the robustness of our analysis and suggests that NF45, NF90, and NF110 likely regulate targets that are common across many cell types, as well as targets that are specific to certain cellular contexts.

Gene Ontology analysis showed that the identified targets are broadly implicated in biological processes related to cell cycle, development, protein folding, and multiple steps in RNA metabolism, including transcription, splicing, and translation (Figure 27D), which is consistent with our molecular and cellular evidence that NF45 and NF90/NF110 are crucial for the regulation of ESC proliferation and differentiation, as well as published work associating NF45 and NF90/NF110 with different RNA-related activities. Indeed, NF45 and NF90/NF110 have been shown to participate in miRNA processing (Gregory et al, 2004; Sakamoto et al., 2009), mRNA splicing through its interactions with the components of the exon junction complex (Singh et al., 2012) and other splicing regulators (Damianov et al., 2016), protein translation through its association with precursors to the 60S ribosomal subunit (Wandrey et al., 2015), and lncRNA function through its interactions with Xist (Chu et al., 2015). Additionally, NF45 has been shown to regulate translation at cellular and viral internal ribosomal entry sites (Merrill et al., 2006; Graber et al., 2010, Lee et al., 2011), and NF90 functions as an mRNA stabilizer and translational repressor, which, indeed, is perhaps its most established role (Shi et al., 2005; Shim et al., 2002; Kuwano et al., 2010; Guo et al., 2016).

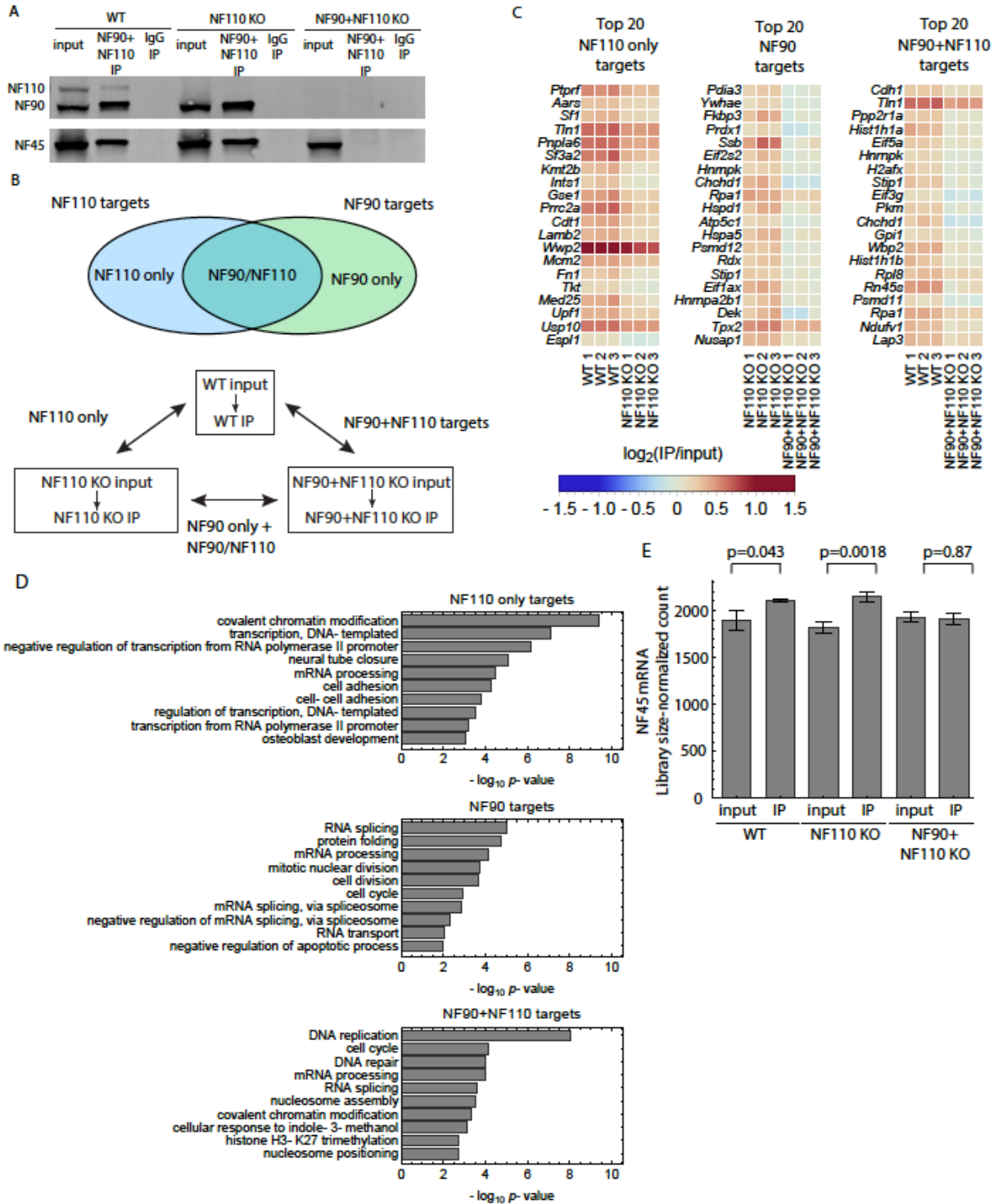


Figure 27. Identification of NF90/NF110 RNA targets. (A) Representative RIP-Western blot showing immunoprecipitation of NF90, NF110, and NF45 in WT, NF110 KO, and NF90+NF110 KO ESCs. (B) Identification of NF90, NF110, and NF90+NF110 RNA binding proteins in ESCs by RIP-seq. Top panel defines NF110 and NF90 targets. Bottom panel defines the comparisons used to call RNA targets. (C) Heatmap of $\log_2(\text{IP}/\text{input})$ RIP-seq expression values of top 20 NF90, NF110 only, and NF90+NF110 RNA targets as identified by a combination of the Poisson ratio test and the log ratio test. (D) GO analysis of identified NF90, NF110 only, and NF90+NF110 targets. (E) Mean library-size normalized read count (as calculated by DESeq2) of NF45 in the RIP-seq libraries. Error bar indicates standard error of the mean.

Interestingly, our data showed enriched NF45 mRNA in the IP compared to the input of both WT and NF110 KO but not NF90+NF110 KO ESCs. The enrichment is small but statistically significant (Figure 27E) and suggests that NF45 mRNA is bound by NF90/NF110. NF90/NF110 could potentially regulate NF45 post-transcriptionally, which could explain the reduced NF45 protein levels in the NF90+NF110 KO (Figure 24D). Together, these data uncover hundreds of potential targets of NF90 and NF110, many shared but others distinct, thus demonstrating that NF110 is not simply redundant with NF90.

Discussion

Post transcriptional and translational processes and the RBPs that drive them have been historically understudied components of ESC regulatory circuitry. In this study, we use an RNAi screen for RBPs that affect the ESC-to-EpiC transition to identify NF45 and NF90/NF110 as factors important for maintaining the ESC state. Using genetic knockout models, we found evidence for both physical and functional interactions of NF45, NF90, and NF110 with one another. Our data suggest that these proteins can function both independently and in complexes in ESCs to regulate RNA processing, cell proliferation, and cell differentiation. Interestingly, a similar screen of RBPs in ESCs recently identified components of the small subunit processome of rRNA biogenesis as important regulators of pluripotency (You et al., 2015), which highlights the significance of post-transcriptional and translational regulation of the ESC state. Due to differences in experimental design, however, there was only partial overlap with the RBPs tested here and no overlap in the hits identified.

Our study design, which combined directed molecular experiments with genome-wide analyses, yielded rich datasets that allow us to propose theoretical models of NF45, NF90, and

NF110 expression and function (Figure 28). First, our results indicate that NF45, NF90, and NF110 influence the expression of one another through regulatory mechanisms diagrammed in Figure 28A. NF90 and NF110 are two alternative isoforms originating from the same gene and are thus in natural competition with each other. Loss of NF45 resulted in an isoform switch between NF90 and NF110, with NF90 transcripts increasing at the expense of NF110 transcripts (Figure 24A), which suggests that NF45 normally promotes splicing of NF110 over NF90. Further, NF90+NF110 KO but not NF110 KO ESCs showed decreased NF45 protein level (Figure 24D), while NF45 transcript levels were unchanged (Figure 24E), which indicates that NF90 promotes NF45 protein stability (Figure 28A), although NF110 might be able to function redundantly. Taken together, our data support a model (Figure 28A) in which NF45 and NF90 regulate each other through a negative feedback loop, with NF90 promoting NF45 protein levels, while NF45 indirectly decreases NF90 expression by promoting NF110 splicing. This negative feedback loop may help reduce stochastic fluctuations, ensuring proper downstream function of the complexes and individual components.

Second, our genomic analyses allowed us to dissect the functional interactions among NF45, NF90, and NF110 in unprecedented detail. In principle, these three proteins can function either on their own or cooperatively by forming complexes, which allows for a variety of possible regulatory modes. We were able to assess the actuality of each mode by using a combinatorial gene set expression analysis framework based on the following observations: (1) there exists a set of genes that is regulated only by NF110, independent of both NF45 and NF90; (2) genes regulated by the NF45-NF110 complex could also be redundantly regulated by the NF45-NF90 complex and vice versa; and (3) there exists a set of genes that is regulated by either NF90 or NF110, independent of NF45. Applying this analysis framework to our transcriptomic

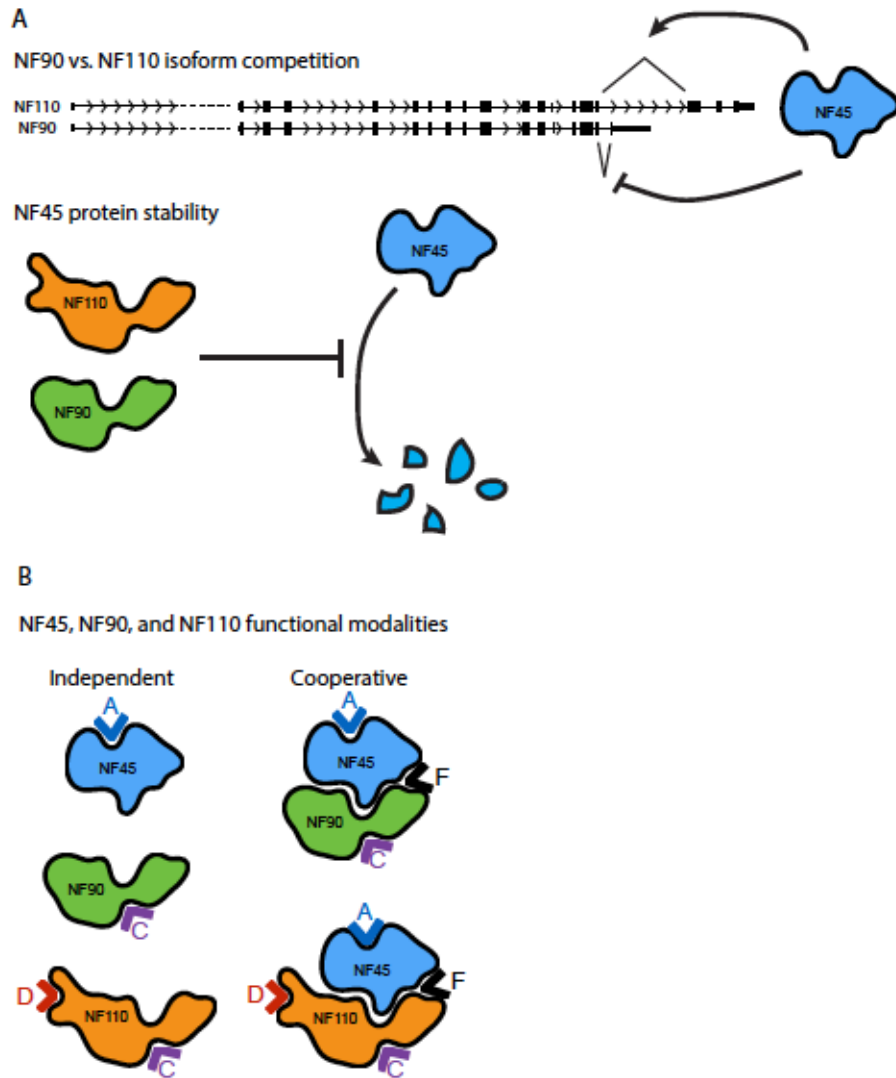


Figure 28. Models of NF45, NF90, and NF110 regulation and function. (A) NF45, NF90, and NF110 expression levels are controlled through a homeostatic regulatory loop. NF45 causes an alternative splicing-driven isoform competition between NF90 and NF110, and NF90 (and possibly NF110) promote NF45 protein stability. The dotted line indicates a truncated representation of NF90/NF110 genomic locus. (B) NF45, NF90, and NF110 can operate either independently or cooperatively. Carets indicate different functional modes: Modes A, C, and D represent regulatory groups of NF45, NF90 or NF110, and NF110, respectively; Mode F represent regulatory groups of NF45-NF90 or NF45-NF110 complexes.

data allowed us to identify four distinct regulatory modes that are supported by our experiments (Figure 28B): Mode A, which consists of genes regulated by NF45 only; Mode C, which consists of genes regulated by NF90 or NF110 redundantly, but independent of NF45; Mode D, which consists of genes regulated by NF110 only, independent of both NF90 and NF45; and Mode F, which consists of genes regulated by either NF45-NF90 and/or NF45-NF110 complexes. NF45,

NF90, and NF110 regulate ESC proliferation, pluripotency, and differentiation through these four distinct modes, and the genes in each of these modes should serve as a useful reference for future studies.

While some of the functions of NF45 and NF90/NF110 are probably shared across mammalian cells, others are likely specific to the unique needs of ESCs. For instance, the changes in NF45 and NF90/NF110 protein levels in the absence of each other have also been observed in HeLa cells (Guan et al., 2008), which suggests that establishing a proper balance of these factors could be important for maintaining cellular homeostasis generally. On the other hand, there was only partial concordance of our NF90/NF110 RIP-seq data with that of previously published NF90/NF110 targets, and indeed, NF90/NF110 RIP-chip performed in HeLa and 293T cells yielded completely non-overlapping targets (Kuwano et al., 2010; Neplioueva et al., 2010), all of which suggests that NF90/NF110 RNA targets can be cell type-dependent. These differences could also be explained by technical variability under different experimental conditions. Indeed, capture of RBP-bound RNAs is inefficient, despite the development of a number of different UV crosslinking-based techniques that strive to increase the sensitivity and specificity of target identification (reviewed in Van Nostrand et al., 2016a; Wang et al., 2015). These challenges are compounded by the very nature of RNA binding as well, since many RNA binding domains have limited sequence specificity (reviewed in Singh et al., 2005; Lunde et al., 2007; Chen and Varani, 2013; Janowsky and Harris, 2015). Additionally, double-stranded RNA (dsRNA) binding proteins like NF45 and NF90/NF110 (the DZF motif is dsRNA-binding, and NF90/NF110 also possesses dsRNA binding motifs (Guan et al., 2008)) rely at least in part on recognition of the 3D RNA structure (Chen and Varani, 2013). While these attributes make identifying RBP binding motifs difficult, it could explain why a number of

RBPs, including NF45 and NF90/NF110, have been found to participate in multiple layers of RNA metabolism (Ye and Blelloch, 2014). With their relative binding sequence degeneracy, interactions with other protein partners is likely critical for establishing binding specificity. This could result in one RBP working with several other proteins to regulate different groups of transcripts in different ways in a single cell type.

In sum, our study uses novel experimental designs and analyses to establish a critical role for NF45 and NF90/NF110 in the ESC circuitry. Further work will be necessary to fully elucidate the mechanisms by which these proteins influence the cell fate decisions underlying early embryonic development.

Materials and Methods

Cell culture

All ESCs were maintained on 0.2% gelatin-coated plates in FBS+LIF+2i medium (DMEM, 15% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, 55uM beta-mercaptoethanol, 1000U LIF, 1uM PD0325901, 3uM CHIR99021). Dual reporter (*mir-290-mCherry/mir-302-eGFP*) ESCs (Parchem et al., 2014) and NF45 WT and NF45 KO ESCs (Horie et al., 2011) have been previously described. NF90+NF110 and NF110 KO ESCs were generated using CRISPR (described below).

RNAi screen development

For the siRNA screen, dual reporter ESCs were plated in gelatin-coated 96-well plates at 1000 cells/well in FBS+LIF+2i medium (day -1). The next day (day 0), LIF and 2i were removed to induce differentiation, and ESCs in each well were transfected with 25nM SMARTpool

siGENOME siRNAs (pool of 4 distinct siRNAs) to 356 separate genes using Dharmafect 1 transfection reagent (GE/Dharmacon). The siRNA to NF110 was custom designed (sense: GCAGAAAGGCUAUGGCCAUUU, antisense: AUGGCCAUAGCCUUUCUGCUU [GE/Dharmacon]). On day 3, plates were fixed and stained with Hoechst 33258 (ThermoFisher Scientific) and imaged on the INCell Analyzer 2000 (GE). The area and intensity of miR-302-eGFP (GFP) and miR-290-mCherry (RFP) fluorescence were measured using the INCell Developer Toolbox software (GE). ESCs maintained in FBS+LIF+2i (“12i”), ESCs allowed to differentiate in the absence of siRNA and transfection reagent (“untreated”), ESCs treated with transfection reagent only (“dfect”), and ESCs treated with control non-targeting siRNA (“siCtrl”) served as differentiation and transfection toxicity controls. ESCs treated with siRNAs against GFP (“siGFP”) and Oct4 (“siOct4”) were used as positive controls for siRNA efficiency (knockdown of Oct4 resulted in cell death). Specifically, each plate contained sets of siRNAs against 32 test genes, as well as 4 replicates of every control condition described above, resulting in a total of 68 conditions evaluated on each plate. Each plate design was repeated in triplicate to ensure consistency in the response of any test gene.

For screen hits, cells were analyzed by flow cytometry on an LSRII flow cytometer (BD Biosciences) to evaluate levels of eGFP and mCherry fluorescence. Cells were also analyzed by qRT-PCR to confirm knockdown of the targeted gene as well as changes in miR-290 and miR-302 expression.

siRNA screen analyses

Two levels of normalization were used to control for sample size during testing and to ensure that values across replicates and plate designs were comparable. (1) A variable of interest (DAPI [Hoechst] area, GFP area, or RFP area) was first normalized across all the replicates of each plate design. This was done by shifting the data by a location estimator (Maronna and Zamar, 2002) and then scaling the data by the estimator from the same source (the implementation uses the `scaleTau2` function defined in R's `robustbase` package). This procedure assumes that the distribution of the variable is the same across replicates. However, this is not guaranteed to be a valid assumption across different plates because we cannot ensure that all plates follow an identical distribution for a given variable. (2) To normalize across plates, we made the assumption that the normalized distribution of all the controls within each plate (not including the test genes) should be identical across plates. Therefore, the distribution of the control siRNAs for each plate was used to calculate a robust measurement of location and scale by which to shift and divide that plate's data. This approach re-scales the data to describe each observation in terms of robust control deviations away from the control center.

This 2-step normalization was first used to transform the DAPI area values to be comparable across all replicates and plate designs. The mean and standard deviation of DAPI area were then calculated across siOct4 wells to estimate the parameters of the background distribution of cell death. We then calculated the z-score for all other genes and denote any gene that is above the 95th FDR-adjusted quantile as live cells. In this way, we identified 49 genes whose knockdown induced cell death. To ensure that we test for changes in differentiation upon knockdown, we removed any target genes whose effect might be poorly estimated. We used stricter exclusion

criteria and removed any genes that had a large percentage of cell death in at least one of its wells, which we defined as having a DAPI area value below the maximum value in the siOct4 samples after normalizing the DAPI area of siOct4 across all replicates and plate designs. For the analysis, we only tested for changes in the RFP and GFP area in the 146 genes whose 3 replicates have DAPI areas that are all above maximum siOct4 DAPI area levels, indicating that the knockdown of that gene does not induce cell death.

To identify the genes whose knockdown significantly changed GFP or RFP signal, we first took the difference of the log fluorescence area and log DAPI area to account for the total amount of cells present. Next, we normalized the log fluorescence data across all replicates and applied the t-test with equal variance for each gene siRNA against the combined siCtrl samples within each plate design. We then controlled for multiple hypothesis testing by calculating the q-value of each test using the qvalue package in bioconductor. To identify the siRNAs that significantly alter fluorescence signal, it is not enough to only use a cutoff on the q-value, since that ignores other controls that could be useful in assessing an RBP's importance. For example, the “untreated” controls were often marked as being “significantly” different from the siCtrl samples. Therefore, we considered RBPs to be “significant” only if they are below a q-value cutoff ($q < 0.001$) and have a q-value less than the “untreated” control.

Generation of NF90+NF110 and NF110 KO ESCs by CRISPR

CRISPR was performed as previously described (Ran et al., 2013). In brief, guide RNAs (sequences obtained from GeCKO [<http://genome-engineering.org/gecko/>] or designed using crispr.mit.edu) were inserted into the PX458 construct or a modified PX458 in which eGFP was

replaced with BFP. The oligos used to create the NF90+NF110 KO allele target exon 3 of NF90/NF110 (CACC-GAATGATGATCGCCACGTGA, AAAC-TCACGTGGCGATCATCATTC). The oligos used to create the NF110 KO allele target the intron preceding the exons private to NF110 (CACC-G-TGCTCTGAGCAGTCGGCTGC, AAAC-GCAGCCGACTGCTCAGAGCA-C) and the 3'UTR of NF110 (CACC-G-TGCCTGGTGTGAGTTCCATC, AAAC-GATGGAACTCACACCAGGCA-C); the intron-targeting guide was cloned into PX458 and the 3'UTR-targeting guide was cloned into PX458-BFP. The constructs (3-10ug) were introduced into V6.5 mouse ESCs (3 million) using the Amaxa Mouse ES Cell Nucleofector Kit (Lonza). GFP+ (for NF90+NF110 KO) or GFP+/BFP+ (for NF110 KO) ESCs were purified by fluorescence-activated cell sorting (FACS) on a FACSAria II UV machine (BD Biosciences) 2 days after nucleofection, and individual colonies were picked for analysis 6 days after sorting. ESCs that had successfully undergone genetic modifications of interest were identified by PCR. For NF90+NF110 KO ESCs, a PCR product different from 387bp when using the primer pair CGTAGAGTGCCCCATTCCTG and CAGCCCCTTCTTTGCTCTCA indicates a mutation in NF90/NF110; for NF110 KO ESCs, the absence of a product when using the primer pair TTGGCTCAGGCCCTAATCAC and CCTGTGCACTCTAAAGCCCT, as well as the presence of a product smaller than 2447bp when using the primer pair CATCTGCAAAGCTGCTGTCG and AGCGGAATTCAAATGTACTGTCT indicate successful removal of the exons private to NF110. PCR products were TOPO-cloned (ThermoFisher Scientific) and sequenced to determine the nature of the mutations. NF90/NF110 protein expression status was confirmed by Western blot.

Colony formation/clonogenicity assay

Cells were plated in 96-well plates at 500 cells/well and stained with Vector Red alkaline phosphatase (AP) 4 days later (Vector Laboratories). The ratio of AP+ colonies formed to the number of cells plated was defined as the colony formation potential.

Population doubling time assay

Cells were plated in 6-well plates at 100,000 cells/well and counted 24, 48, and 72hrs later. Population doubling time was calculated as previously described (Wang et al., 2007) using the formula, $Y_{\text{end}} = Y_{\text{start}} \times 2^{(t/T)}$, where T is the cell population doubling time, Y_{start} is the starting number of ESCs (e.g., number of ESCs counted at 24hrs), and Y_{end} is the ending number of ESCs (e.g., number of ESCs counted at 72hrs) after growth for a period of time t (e.g., 48hrs).

Western blot

For whole-cell extract, cells were lysed in protein lysis buffer (25mM Tris pH 7.9, 150mM NaCl, 0.1mM EDTA, 0.1% Triton X-100, 10% glycerol) with protease inhibitors and resolved on 10% Mini-PROTEAN TGX gels (BioRad). Cytoplasmic and nuclear fractions were collected using ThermoFisher Scientific's Subcellular Fractionation Kit for Cultured Cells according to the manufacturer's protocol. Equal amounts of protein were loaded per sample. The following antibodies were used for blotting: 1:10000 NF45 (Everest Biotech EB07784), 1:1000 NF90/NF110 (BD Biosciences clone 21/DRBP76), 1:10000 Nanog (Calbiochem SC1000), 1:2000 Gapdh (Santa Cruz Biotechnology sc-25778), 1:10000 Tubulin (Abcam ab6160).

Co-immunoprecipitation

ESCs were harvested, pelleted, and lysed in 10 pellet volumes of low NP40 lysis buffer (12.5mM Tris pH 7.9, 150mM NaCl, 0.1mM EDTA, 0.1% Triton X-100, 10% glycerol) with cOmplete protease inhibitors (Roche). Lysates were freeze-thawed 3 times and then cleared by centrifugation at maximum speed for 15min at 4C. Antibodies (Ab) for immunoprecipitation (5ug NF45 Ab/1ml lysate (Everest Biotech EB07784), 5ug NF90/NF110 Ab/1ml lysate (BD Biosciences clone 21/DRBP76), 6ug normal rabbit IgG/1ml lysate (ThermoFisher Scientific) were conjugated to 40ul Protein G Dynabeads/1ml lysate (ThermoFisher Scientific) in low NP40 lysis buffer for 2 hours at room temperature. Lysates were incubated with Protein G Dynabead-conjugated antibodies overnight at 4C. Immunoprecipitates were washed 2x with low NP40 lysis buffer and resolved and visualized by Western blot as described above.

Embryoid body (EB) generation

Embryoid bodies were generated using Aggrewell 800 plates (STEMCELL Technologies) according to the manufacturer's protocol. In brief, 1.5 million ESCs were seeded per Aggrewell (~5000 cells per EB) in differentiation medium (DMEM, 15% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, 55uM beta-mercaptoethanol). After 3 days, EBs were transferred to ultra-low attachment plates (Corning) and harvested every 3 days for qRT-PCR analysis.

qRT-PCR

Total RNA was collected using TRIzol (ThermoFisher Scientific) according to the manufacturer's protocol. For cDNA synthesis, RNA was treated with DNaseI (Invitrogen) and

reverse-transcribed with oligo-dT primers (or for RIP samples, a 1:1 mix of oligo-dT and random hexamer primers) using the SuperScript III kit (Invitrogen). Total cDNA was diluted 1:5, and qPCR was performed using gene-specific primer sets (listed below) and SensiFast SYBR Hi-ROX master mix (Bioline). MiRNA qRT-PCR was performed with the polyA and SYBR Green method as previously described using miRNA-specific forward primers and a 3' RACE adaptor reverse primer (Shi and Chiang 2005). Primer specificity was verified through analysis of dissociation curves in experimental, no RT, and water-only samples.

Primer	Sequence
NF45 qPCR F	AAGCCAGCACCTGATGAGAC
NF45 qPCR R	TTCCTGGGGCCACAATCAAG
NF90+NF110 qPCR F	TGTGGAGGTAGACGGCAGTA
NF90+NF110 qPCR R	CTGGAGTCTCTGCCTTCAGC
NF110 qPCR F	GGGCTCTGACTACAGCTACG
NF110 qPCR R	CCTCCATGTGAGCCTGTGTT
NF90 qPCR F	ACATGAATGCTGGTGCTGGA
NF90 qPCR R	CGCTCTAGGAAGCCCCAAA
Rpl7-qPCR F	GAACCAAAGCTGGCCTTTGTCATC
Rpl7-qPCR R	CAATGTATGGCTCCACAATCCGCA
T/Brachyury qPCR F	CTGGGAGCTCAGTTCTTTCGA
T/Brachyury qPCR R	GAGGACGTGGCAGCTGAGA
Gata6 qPCR F	TGACTCCTACTTCTTCTTCTC
Gata6 qPCR R	TACTTGAGGTCACTGTTCTCG
Fgf5 qPCR F	CCTTGCGACCCAGGAGCTTA
Fgf5 qPCR R	CCGTCTGTGGTTTCTGTTGAGG
Klf4 qPCR F	TGTGGCAAACCTATACCAAGAG
Klf4 qPCR R	CACAGCCGTCCCAGTCAC
Rex1 qPCR F	GAAAGTGAGATTAGCCCCGAG
Rex1 qPCR R	GTCCCCTTTGTCATGTACTCC
Pax6 qPCR F	CGGGAAAGACTAGCAGCCAA
Pax6 qPCR R	TTGCTGGCCTGTCTTCTCTG
Nanog qPCR F	AACCAAAGGATGAAGTGCAAGCGG
Nanog qPCR R	TCCAAGTTGGGTTGGTCCAAGTCT
Pax6 qPCR F	CGGGAAAGACTAGCAGCCAA
Pax6 qPCR R	TTGCTGGCCTGTCTTCTCTG
Sno202 qPCR F	GTACTTTTGAACCCTTTTCCATCTGATG
miR-293 qPCR F	AGTGCCGCAGAGTTTGTAGTGT
miR-294 qPCR F	AAAGTGCTTCCCTTTTGTGTGT
miR-295 qPCR F	AAAGTGCTACTACTTTTGTAGTCT

miR-302a qPCR F	TAAGTGCTTCCATGTTTTGGTGA
miR-302b qPCR F	TAAGTGCTTCCATGTTTTAGTAG
miR-367 qPCR F	AATTGCACTTTAGCAATGGTGA
3' RACE adaptor outer for miRNA qPCR	GCGAGCACAGAATTAATACGACT

RNA-seq library generation

For NF45 WT and NF45 KO ESCs, 1ug of RNA per sample was processed using the Illumina TruSeq Stranded mRNA Sample Prep kit according to the manufacturer's protocol. For NF90/NF110 WT, NF90+NF110 KO, and NF110 KO ESCs, 1ug of RNA per sample was processed using the KAPA Stranded RNA-Seq with RiboErase kit according to the manufacturer's protocol.

RNA immunoprecipitation (RIP)

RIP was performed essentially as previously described (Keene et al., 2006; Peritz et al., 2006). ESCs were harvested, pelleted, and lysed in 10 pellet volumes of polysome lysis buffer (100mM KCl, 5mM MgCl₂, 10mM HEPES pH 7.5, 0.5% NP40 substitute, 1mM DTT) with cOmplete protease inhibitors (Roche), 100U/ml Superase-In (ThermoFisher Scientific), and 400uM vanadyl ribonucleoside complex (NEB). Lysates were freeze-thawed once, incubated with 100U/ml Turbo DNase (ThermoFisher Scientific) for 15min at 37C, and then cleared by centrifugation at maximum speed for 15min at 4C. Aliquots of the cleared lysate were taken at this step to serve as input for Western blot and qRT-PCR analysis. Antibodies for immunoprecipitation (5ug NF90/NF110 Ab/1ml lysate [BD Biosciences clone 21/DRBP76], 6ul normal rabbit IgG/1ml lysate [ThermoFisher Scientific]) were conjugated to 40ul/1ml lysate Protein G Dynabeads (ThermoFisher Scientific) in polysome lysis buffer for 2 hours at room temperature. Lysates were then incubated with Protein G Dynabead-conjugated antibodies for

4hr at 4C. Immunoprecipitates were washed 3x with polysome lysis buffer and then 3x with polysome lysis buffer containing 1M urea. An aliquot of the immunoprecipitate was taken for Western blot, while the rest of the immunoprecipitate, along with the qRT-PCR input lysate, was incubated with 1.2mg/ml proteinase K (Roche) and 2x proteinase K buffer (100mM Tris-HCl pH7.5, 12.5mM EDTA, 150mM NaCl, 2% SDS) for 30min at 55C. Samples were then lysed in TRIzol LS (ThermoFisher Scientific), and RNA was extracted and used for qRT-PCR or sequencing.

Mapping reads

Both RNA-seq and RIP-seq reads were mapped to the mm10 reference genome using Tophat2 (Kim et al., 2013) with options -g 1 --prefilter-multihits. Cuffnorm (Trapnell et al., 2012) was used with the gtf file from UCSC mm10 (Illumina iGenomes July 17, 2015 version) as transcript annotation to evaluate relative expression level of genes (fragments per kilobase of transcript per million mapped reads (FPKM)). Genes belonging to the following categories were removed from subsequent analysis: tRNA, Mt-tRNA, rRNA, scRNA, snRNA, snoRNA, miRNA, and misc-RNA.

Differential expression

HTSeq-count (Anders et al., 2014) was used to summarize read counts for each gene. The resulting read count matrices were used in DESeq2 (Love et al., 2014) analysis of differential expression. Because paired-end 50 bp reads were used in NF45 RNA-seq experiments, while single-end 50 bp reads were used in NF90/NF110 RIP-seq experiments with much lower sequencing depth, NF45 RNA-seq data generally had higher sensitivity in calling differentially

expressed genes compared to NF90/NF110 RIP-seq data. We thus used different thresholds for statistical significance in the two data sets: adjusted p-value < 0.001 for the NF45 RNA-seq data and adjusted p-value < 0.05 for the NF90/NF110 RIP-seq data. In addition, we further required \log_2 fold change (as calculated by DESeq2) to be at least 1 in NF45 RNA-seq data.

Clustering analysis

The hierarchical clustering in Figure 23B was performed as follows. We identified 1670 genes differentially expressed in at least one of the following comparisons: NF45 KO vs. NF45 WT, NF90+NF110 KO vs. NF110 KO, NF110 KO vs. NF90/NF110 WT, and NF90+NF110 KO vs. NF90/NF110 WT. Note that the input samples from the NF90/NF110 RIP-seq experiments were used as RNA-seq data. By further requiring the genes to be robustly expressed (FPKM \geq 10) in at least one condition, we retained 971 genes as candidate regulatory targets. These 971 genes were used for hierarchical clustering in Figure 23B. In order to remove batch effects and facilitate comparison between NF45 and NF90/NF110 RNA-seq data (RIP-seq input data), we separately standardized the FPKM value of each gene by subtracting the mean and dividing by the standard deviation across NF45 RNA-seq samples and NF90/NF110 RNA-seq samples. Hierarchical clustering of the genes was performed using the standardized expression values with Euclidean metric and average linkage. Expression modules were then extracted by accounting for differential gene expression analysis.

Gene ontology analysis

Gene ontology analysis was performed with the functional annotation tool from DAVID 6.8 (Huang et al., 2009a; Huang et al., 2009b).

Pairwise expression correlation analysis

The pairwise expression correlation analysis is similar to the analysis performed as before (Bulut-Karslioglu et al., 2016). Figure 24A plots the \log_2 fold change of FPKM values between indicated conditions for 7995 genes that are robustly expressed (mean FPKM ≥ 10) in at least one of the NF45 KO, NF45 WT, NF90/NF110 WT, NF110 KO, and NF90+NF110 KO RNA-seq samples. The Spearman correlation coefficients and associated p-values are also indicated. In Figure 24B, we defined pathway expression as the mean FPKM values of genes associated with each of the 281 KEGG pathways (Kanehisa and Goto, 2000; Kanehisa et al., 2016) that contain at least 10 genes. The \log_2 fold change of pathway expression values between indicated conditions were plotted, with Spearman correlation coefficients and associated p-values indicated. In Figure 24C, we defined GO term expression as the mean FPKM values of genes associated with each of the 3516 GO terms (Huang et al., 2009a; Huang et al., 2009b) that contain at least 10 genes. The \log_2 fold change of GO term expression values between indicated conditions were plotted, with Spearman correlation coefficients and associated p-values indicated.

Identifying binding targets from RIP-seq data

NF90/NF110 RIP-seq was performed in NF110 KO, NF90+NF110 KO, and NF90+NF110 WT ESCs. Each experiment contains paired input (total RNA) and IP samples, and experiments were repeated in biological triplicate. We consider a theoretical RIP-seq experiment as consisting of 4 conditions: WT input, WT IP, KO input, and KO IP. In this theoretical RIP-seq experiment, genes enriched in KO IP compared to KO input contain unspecific targets of the antibody and experimental noise, while genes enriched in WT IP compared to WT input contain true binding

targets of the protein of interest as well as unspecific targets of the antibody and experimental noise. Thus, comparing WT IP/WT input with KO IP/KO input yields the true binding targets. Our NF90/NF110 RIP-seq can be regarded as three realizations of the theoretical experiment by replacing the 4 conditions in the theoretical experiment with the conditions of the real experiment, as summarized in the following table (also see Figure 25B).

Experiment	Conditions				Results
Theoretical experiment	WT input	WT IP	KO input	KO IP	True binding targets of protein of interest
1	WT input	WT IP	NF110 KO input	NF110 KO IP	NF110 only targets
2	WT input	WT IP	NF90+NF110 KO input	NF90+NF110 KO IP	NF90+NF110 targets
3	NF110 KO input	NF110 IP input	NF90+NF110 KO input	NF90+NF110 KO IP	NF90 targets (including targets redundantly bound by NF110)

Combinatorial gene set expression analysis

This analysis aimed to infer the first order approximation of the functional interactions between NF45, NF90, and NF110 by using the RNA-seq data from our genetic knockout models. Various modes may exist in the regulatory network of these three proteins, as they can function both cooperatively as a complex and independently by themselves. Furthermore, when functioning independently of NF45, NF90 and NF110 may regulate the same genes redundantly or distinct sets of genes. The same possibilities also exist for NF45-NF90 and NF45-NF110 complexes. Employing a combinatorial gene set expression analysis framework, we were able to rule out certain regulatory modes inconsistent with our transcriptomic data, while confirming the

existence of other possible modes. The results provide novel insights about the gene regulatory network associated with NF45 and NF90/NF110 in ESCs (Figure 26B).

Chapter 5—Concluding remarks and future directions

In this thesis, we explored the roles of miRNAs and RBPs as post-transcriptional regulators in two experimental assays of pluripotency. In this final chapter, we address lingering questions directly stemming from our studies as well as broader implications for future directions of investigation.

Are miR-290 and miR-302 absolutely dispensable in reprogramming?

In chapters 2 and 3, we studied the *mir-290~295* (*mir-290*) and *mir-302~367* (*mir-302*) miRNA clusters in somatic cell reprogramming, first using them as a tool to shed light on the path through which somatic cells are able to reach an induced pluripotent state and then examining whether these same miRNAs are required for the process for which they are so informative. We found that while *mir-290* and *mir-302* are expressed in a sequential order during forward differentiation and embryonic development, their activation is unordered during reprogramming, though the pattern is affected by the specific cocktail of reprogramming factors used. Fascinatingly, the ESCC miRNA family members of *mir-290* and *mir-302* are not required for reprogramming, as we were able to successfully generate iPSCs from *mir-290* KO and *mir-302a-d* KO MEFs. A major caveat to these experiments is that we could not generate *mir-290/mir-302a-d* double KO MEFs to test in our reprogramming assay, as these embryos are embryonic lethal before E13.5 when MEFs are harvested (Parchem et al., 2015). We strove to overcome this experimental hurdle by introducing a *GFP-miR-302-sponge* into *mir-290* KO MEFs (Figure 13). Excitingly, these MEFs were able to reprogram, but although we were able to demonstrate sponge functionality in the iPSCs, we could not prove whether the sponge fully

removed all ESCC miRNAs and whether or not it was functional throughout the entire de-differentiation assay.

Therefore, our current efforts are directed at genetically ablating the *mir-302* locus in *mir-290* KO MEFs. Our approach is using CRISPR guide RNAs that target the 5' and 3' ends of the locus. MEFs that took up the guide RNAs will be bulk-sorted by FACS, as the CRISPR constructs contain GFP and BFP markers. We will not be able to isolate and reprogram individual *mir-290/mir-302* double KO MEF clones, as primary MEFs senesce with prolonged cell culture; however, the results of the de-differentiation assay will be telling. If, for example, we obtain no iPSC colonies relative to wt or single KO MEFs, we will conclude that the *mir-290* and *mir-302* miRNAs are indeed required for reprogramming, but that each cluster can compensate for the other, as we observed in our experiments in chapter 3. If, however, we do obtain iPSC colonies, we will isolate and genotype individual iPSC clones to determine whether any are *mir-290-/mir-302-*. The presence of any such double KO iPSCs would be proof of principle that these two miRNA clusters are not required for reprogramming. Genotyping the iPSCs from this experiment would also help us definitively determine whether miR-367, which does not share a seed sequence with miR-302a-d and which has been found to be variably important to the de-differentiation process (Anokye-Danso et al., 2011; Liao et al., 2011), is truly required for generating iPSCs. If all *mir-290/mir-302* double KO iPSC clones retain *mir-367* in the genomic locus, then we would conclude that miR-367 is indeed necessary for reprogramming.

However, the cleanest approach to addressing this question is to create an inducible, conditional KO of the *mir-302* locus in *mir-290* KO mice. This way, *mir-302* can be specifically removed from the MEFs upon addition of doxycycline or another chemical signal. Regardless,

though, we will still face the challenge of being able to demonstrate whether or not the locus was fully deleted in all MEF clones at the beginning of the reprogramming process. If we do not observe any decreases in reprogramming efficiency in *mir-290/mir-302* double KO cells even when the resulting iPSCs genotype correctly, we are left with the possibility that *mir-302* was not fully deleted until midway through the reprogramming assay. However, given that endogenous *mir-302* does not activate until the late stages of reprogramming (chapter 2, Parchem et al., 2014), it is conceptually unlikely that delayed removal of the cluster would confound the interpretation of whether *mir-302* is required for iPSC generation.

A binding motif for NF90/NF110?

In chapter 4, we identified NF45 and NF90/NF110 as physically and functionally interconnected regulators of ESC pluripotency, proliferation, and differentiation. Although these proteins have been implicated to be RBPs through various genome-wide and directed molecular studies, we were unable to conclusively identify binding sites for them in our ESCs. Initially, we attempted to perform photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP, Hafner et al., 2010) for NF45 and NF90/NF110, but the low crosslinking efficiency and low read diversity prevented us from performing in depth analyses (data not shown). RIP-seq allowed us to identify transcripts that co-immunoprecipitated with NF45, NF90, and NF110 (Figure 27); however, it cannot identify the exact location of binding on the RNA molecule of interest. Recently, a group developed an improved method for detecting the binding sites of RBPs called enhanced crosslinking and immunoprecipitation (eCLIP, Van Nostrand et al., 2016b). Using eCLIP data generated in K562 human leukemia cells, we found preferential binding of NF90/NF110 in introns (data not shown). However, we were unable to

detect any significant sequence motif using HOMER (Heinz et al., 2010, data not shown), which may either suggest that NF90/NF110 might recognize directing signals like RNA structure or may be due to the technical limitations of reliably capturing and identifying RBP-bound RNAs using the various variations on CLIP (reviewed in Van Nostrand et al., 2016a; Wang et al., 2015). Whether NF90/NF110 exhibits similar binding patterns in ESCs and K562 cells remains to be determined by applying eCLIP or other more efficient protocols on NF90/NF110 in ESCs. If NF90/NF110 does indeed bind to introns in ESCs, it would be of interest to determine whether and how it might be involved in transcript splicing and whether it accomplishes this in complex with NF45 or other cofactors.

NF45 and NF90/NF110 interactions with the larger regulatory network?

The physical and functional interactions we discovered between NF45 and NF90/NF110 combined with the observation that a number of RBPs are involved in different regulatory activities (Ye and Blelloch, 2014) raise potentially interesting directions for future research. Specifically, it would be intriguing to discover how NF45 and NF90/NF110 interact with other regulatory factors, and how these combinations of factors change to effect changes in cell state. Indeed, NF45, NF90, and NF110 have been identified over the years in a number of protein complexes involved in a myriad of cellular activities. Some notable examples include Drosha (Gregory et al., 2004), which regulates miRNA biogenesis; components of the exon junction complex (Singh et al., 2012) and the splicing factor Rbfox2 (Damianov et al., 2016), which regulate alternative splicing; and precursors of the 60S ribosomal subunit (Wandrey et al., 2015), which regulate translation.

Therefore, it would be interesting to take expand upon these known interactions and discover the full network of factors that NF45 and NF90/NF110 form complexes with. We could perform NF45 or NF90/NF110 IP-mass spectrometry in ESCs, EpiCs, and other unrelated cell lines to identify all of their protein cofactors in different cellular contexts. We could also conduct such experiments in different intracellular compartments of a single cell type. The proteins interactors detected in each of these various situations could provide insight into how NF45 and NF90/NF110 exact their effects under different conditions.

A complementary approach could be to detect changes in all RBP binding upon loss of NF45, NF90/NF110, or both, since our NF90/NF110 RIP-seq data suggested that NF90/NF110 regulate targets that are involved in different steps of RNA processing (Figure 27D). A number of groups, including our lab, are working on optimizing experimental approaches for such a global RBP “footprinting” assay, which broadly consists crosslinking RBPs to their target RNAs *in vivo*, isolating the crosslinked RNAs, and identifying them by RNA-seq. Such an experiment could not only allow us to identify RBPs whose activity depends on NF45 or NF90/NF110 but also help us determine the mechanism by which these RBP coregulators function on their RNA targets.

Importantly, the approaches proposed here for NF45 and NF90/NF110 could be applied for studying other RBPs, as well, many of which very likely also function in fluid combinations of various protein and RNA co-regulators.

Lessons for the future

It is without a question that post-transcriptional mechanisms are critical for regulating pluripotent stem cells, and this thesis contributes to the field’s understanding of how specific

miRNAs and RBPs fit into these regulatory networks. In addition to opening up the new research directions discussed in the sections above, the insights we make here highlight the skepticism and care that must be exercised when extrapolating and making assumptions from data not directly addressing the question at hand. For example, our work on miR-290 and miR-302 showed that factors that are highly expressed and functionally important in a cell state may counterintuitively not actually be necessary for the acquisition and maintenance of that cell state. As another example, we found that NF90/NF110 targets identified in HeLa, 293T, and ESCs are mostly non-overlapping (Kuwano et al., 2010; Neplioueva et al., 2010), which shows that the same factor may have different targets or altogether different functions in different cellular contexts. These lessons and principles apply not just to miRNAs and RBPS but also to any molecule or factor being investigated in pluripotent cells or another biological system. Keeping these complexities in mind for future studies will help bring us closer to fully understanding stem cells, development, and other biological systems and to being able to effectively utilize this knowledge in clinical situations.

References

- Aeschimann, F., Kumari, P., Bartake, H., Gaidatzis, D., Xu, L., Ciosk, R., and Grosshans, H. (2017). LIN41 post-transcriptionally silences mRNAs by two distinct and position-dependent mechanisms. *Mol. Cell* *65*, 1-14.
- Amano, H., Itakura, K., Maruyama, M., Ichisaka, T., Nakagawa, M., and Yamanaka, S. (2006). Identification and targeted disruption of the mouse gene encoding ESG1 (PH34/ECAT2/DPPA5). *BMC Dev. Biol.* *6*, 11.
- Anders, S., Pyl, P.T., and Huber, W. (2014). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166-169.
- Anokye-Danso, F., Trivedi, C.M., Jühr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., and Morrisey, E.E. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* *8*, 376–388.
- Baltz, A.G., Munschauer, M., Schwanhaüusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* *46*, 674–690.
- Bar, M., Wyman, S.K., Fritz, B.R., Qi, J., Garg, K.S., Parkin, R.K., Kroh, E.M., Bendoraite, A., Mitchell, P.S., Nelson, A.M. et al. (2008). MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells* *26*, 2496-2505.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* *136*, 215-233.
- Barker, N., Bartfeld, S., and Clevers, H. (2010). Tissue-resident adult stem cell populations of rapidly self-renewing organs. *Cell Stem Cell* *7*, 656-670.
- Bellin, M., Marchetto, M.C., Gage, F.H., and Mummery, C.L. (2012). Induced pluripotent stem cells: the new patient? *Nat. Rev. Mol. Cell Biol.* *13*, 713-726.
- Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T, Klatt, P., et al. (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat. Struct. Mol. Biol.* *15*, 268-279.
- Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. *Nat. Genet.* *35*, 215–217.
- Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* *16*, 513-525.

- Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* 2, 151–159.
- Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, 191-195.
- Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209–1222.
- Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B.J., Porubsky, D., et al. (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell Stem Cell* 15, 295-309.
- Bulut-Karslioglu, A., Biechele, S., Jin, H., Macrae, T.A., Hejna, M., Gertsenstein, M., Song, J.S., and Ramalho-Santos, M. (2016). Inhibition of mTOR induces a paused pluripotent state. *Nature* 540, 119-123.
- Card, D.A., Hebbar, P.B., Li, L., Trotter, K.W., Komatsu, Y., Mishina, Y., and Archer, T.K. (2008). Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol. Cell. Biol.* 28, 6426–6438.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406.
- Chan, E.M., Ratanasirintraooot, S., Park, I.H., Manos, P.D., Loh, Y.H., Huo, H., Miller, J.D., Hartung, O., Rho, J., Ince, T.A., et al. (2009). Live cell imaging distinguishes bona fide human iPSC cells from partially reprogrammed cells. *Nat. Biotechnol.* 27, 1033–1037.
- Chang, H.M., Martinez, N.J., Thornton, J.E., Hagan, J.P., Nguyen, K.D., and Gregory, R.I. (2012). Trim71 cooperates with microRNAs to repress Cdkn1a expression and promote embryonic stem cell proliferation. *Nat. Commun.* 3, 923.
- Chen, C., Ridzon, D., Lee, C.T., Blake, J., Sun, Y., and Strauss, W.M. (2007). Defining embryonic stem cell identity using differentiation-related microRNAs and their potential targets. *Mamm. Genome* 18, 316-327.
- Chen, Y., and Varani, G. (2013). Engineering RNA-binding proteins for biology. *FEBS J.* 280, 3734-3754.
- Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* 151, 765–777.

- Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell* *161*, 404-416.
- Ciaudo, C., Servant, N., Cognat, V., Sarazin, A., Kieffer, E., Viville, S., Colot, V., Barillot, E., Heard, E., and Voinnet, O. (2009). Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. *PLoS Genet.* *5*, e1000620.
- Corthesy, B., and Kao, P.N. (1994). Purification by DNA affinity chromatography of two polypeptides that contact the NF-AT DNA binding site in the interleukin 2 promoter. *J. Biol. Chem.* *269*, 20682-20690.
- Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* *309*, 1369-1373.
- Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A. et al. (2016). Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. *Cell* *165*, 606-619.
- De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., et al. (2015). Hallmarks of pluripotency. *Nature* *525*, 469-478.
- Diaz, A.A., Qin, H., Ramalho-Santos, M. and Song, J.S. (2015). HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res.* *43*, e16-e16.
- Eaves, C.J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood* *23*, 2605-2613.
- Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* *4*, 721-726.
- Ebert, M.S., and Sharp, P.A. (2010). MicroRNA sponges: progress and possibilities. *RNA* *16*, 2043-2050.
- Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proc. Natl. Acad. Sci. USA* *103*, 16319–16324.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* *292*, 154-156.
- Fabian, M.R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* *19*, 586-593.
- Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* *11*, 477–490.

Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Genet.* 9, 102-114.

Flynn, R.A., and Chang, H.Y. (2014). Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* 14, 752–761.

Gabardella, G., Carissimo, A., Chen, A., Cutillo, L., Nowakowski, T.J., di Bernardo, D., and Blueloch, R. (2017). The impact of microRNAs on transcriptional heterogeneity and gene co-expression across single embryonic stem cells. *Nat. Commun.* 8, 14126.

Ghosal, S., Das, S., and Chakrabarti, J. (2013). Long noncoding RNAs: new players in the molecular mechanism for maintenance and differentiation of pluripotent stem cells. *Stem Cells Dev.* 22, 2240–2253.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Treka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* 11, 769–782.

Graber, T.E., Baird, S.D., Kao, P.N., Mathews M.B., and Holcik, M. (2010). NF45 functions as an IRES trans-acting factor that is required for translation of cIAP1 during the unfolded protein response. *Cell Death Differ.* 17, 719-729.

Graf, T., and Stadtfeld, M. (2008). Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* 3, 480-483.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235-240.

Greve, T.S., Judson, R.L., and Blueloch, R. (2013). microRNA control of mouse and human pluripotent stem cell behavior. *Annu. Rev. Cell Dev. Biol.* 29, 213-239.

Gruber, A.J., Grandy, W.A., Balwierz, P.J., Dimitrova, Y.A., Pachkov, M., Ciaudo, C., Nimwegen, Ev, and Zavolan, M. (2014). Embryonic stem cell-specific microRNAs contribute to pluripotency by inhibiting regulators of multiple differentiation pathways. *Nucleic Acids Res.* 42, 9313-9326.

Guan, D., Altan-Bonnet, N., Parrott, A.M., Arrigo, C.J., Li, Q., Khaleduzzaman, M., Li, H., Lee, C.G., Pe'ery, T., and Mathews, M.B. (2008). Nuclear factor 45 (NF45) is a regulatory subunit of complexes of NF90/NF110 involved in mitotic control. *Mol. Cell. Biol.* 28, 4629-4641.

Guo, C., Xue, Y., Yang, G., Yin, S., Shi, W., Cheng, Y., Yan, X, Fan, S., Zhang, H., and Zeng, F. (2016). Nanog RNA binding proteins YBX1 and ILF3 affect pluripotency of embryonic stem cells. *Cell Biol. Int.* 8, 847-860.

- Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* *136*, 1063-1069.
- Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* *10*, 622–640.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr., Jungkamp, A.C., Munschauer, M. et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* *141*, 129-141.
- Hagan, J.P., Piskounova, E., and Gregory, R.I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* *16*, 1021–1025.
- Han, D.W., Greber, B., Wu, G., Tapia, N., Arauzo-Bravo, M.J., Ko, K., Bernemann, C., Stehling, M., and Scholer, H.R. (2010). Direct reprogramming of fibroblasts into epiblast stem cells. *Nat. Cell Biol.* *13*, 66-71.
- Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* *3*, 391-401.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* *38*, 576-589.
- Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* *32*, 276–284.
- Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H., and Kim, V.N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* *151*, 521–532.
- Hochedlinger, K., and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. *Development* *136*, 509–523.
- Horie, K., Kokubu, C., Yoshida, J., Akagi, K., Isotani, A., Oshitani, A., Yusa, K., Ikeda, R., Huang, Y., Bradley, A., and Takeda, J. (2011). A homozygous mutant embryonic stem cell bank applicable for phenotype-driven genetic screening. *Nat. Methods*, *8*, 1071-1077.
- Houbaviy, H.B., Dennis, L., Jaenisch, R., and Sharp, P.A. (2005). Characterization of a highly variable eutherian microRNA gene. *RNA* *11*, 1245-1257.

- Houbaviy, H.B., Murray, M.F., and Sharp, P.A. (2003). Embryonic stem cell-specific microRNAs. *Dev. Cell* 5, 351-358.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc*, 4, 44-57.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1-13.
- Jankowsky, E., and Harris, M.E. (2015). Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 16, 533-544.
- Jouneau, A., Ciaudo, C., Sismeiro, O., Brochard, V., Jouneau, L., Vandormael-Pournin, S., Coppee, J.-Y., Zhou, Q., Heard, E., Antoniewski, C., and Cohen-Tannoudji, M. (2012). Naive and primed murine pluripotent stem cells have distinct miRNA expression profiles. *RNA* 18, 253–264.
- Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* 27, 459–461.
- Judson, R.L., Greve, T.S., Parchem, R.J., and Blelloch, R. (2013). MicroRNA-based discovery of barriers to dedifferentiation of fibroblasts to pluripotent stem cells. *Nat. Struct. Mol. Biol.* 20, 1227–1235.
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27-30.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457-D462.
- Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* 19, 489–501.
- Kao, P.N., Chen, L., Brock, G., Ng, J., Kenny, J., Smith, A.J., and Corthesy, B. (1994). Cloning and expression of cyclosporine A- and FK506-sensitive Nuclear Factor of Activated T-Cells: NF45 and NF90. *J. Biol. Chem.* 269, 20691-20699.
- Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* 1, 302-307.
- Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.* 8, 533–543.

- Kim, B.M., Thier, M.C., Oh, S., Sherwood, R., Kanellopoulou, C., Edenhofer, F., and Choi, M.Y. (2012). MicroRNAs are indispensable for reprogramming mouse embryonic fibroblasts into induced stem cell-like cells. *PLoS One* 7, e39239.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Kluiver, J., Slezak-Prochazka, I., Smigielska-Czepiel, K., Halsema, N., Kroessen, B.J., and van den Berg, A. Generation of miRNA sponge constructs. *Methods* 58, 113-117.
- Krishnakumar, R., Chen, A.F., Pantovich, M.G., Danial, M., Parchem, R.J., Labosky, P.A., and Belloch, R. (2016). FOXD3 regulates pluripotent stem cell potential by simultaneously initiating and repressing enhancer activity. *Cell Stem Cell* 18, 104-117.
- Kuwano, Y., Pullmann, R. Jr., Marasa, B.S., Abdelmohsen, K., Lee, E.K., Yang, X, Martindale, J.L., Zhan, M., and Gorospe, M. (2010). NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res.* 38, 225-238.
- Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsvelde, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1122–1130.
- Lee, J.W., Liao, P.C., Young, K.C., Chang, C.L., Chen, S.S.L., Chang, T.T., Lai, M.D., and Wang, S.W. (2011). Identification of hnRNPH1, NF45, and C14orf166 as novel host interacting partners of the mature hepatitis C virus core protein. *J. Proteome Res.* 10, 4522-4534.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15-20.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787-798.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 7, 51-63.
- Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y, Cai, J., Guo, X., Qin, B. et al. (2011). MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J. Biol. Chem.* 286, 17359-17364.
- Lim, C.Y., Tam, W.L., Zhang, J., Ang, H.S., Jia, H., Lipovich, L., Ng, H.H., Wei, C.L., Sung, W.K., Robson, P., et al. (2008). Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. *Cell Stem Cell* 3, 543–554.

- Lin, S.L., Chang, D.C., Lin, C.-H., Ying, S.Y., Leu, D., and Wu, D.T.S. (2011). Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nuc. Acids Res.* *39*, 1054-1065.
- Liu, Z., Sall, A., and Yang, D. (2008). MicroRNA: an emerging therapeutic target and intervention tool. *Int. J. Mol. Sci.* *9*, 978-999.
- Liu, Z., Skamagki, M., Kim, K., and Zhao, R. (2015). Canonical microRNA facilitates but may be dispensable for transcription factor-mediated reprogramming. *Stem Cell Reports* *5*, 1119-1127.
- Loedige, I., Gaidatzis, D., Sack, R., Meister, G., and Filipowicz, W. (2013). The mammalian TRIM-NHL protein TRIM71/LIN-41 is a repressor of mRNA function. *Nucleic Acids Res.* *41*, 518–532.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Lu, R., Markowitz, F., Unwin, R.D., Leek, J.T., Airoidi, E.M., MacArthur, B.D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L.A., et al. (2009). Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* *462*, 358–362.
- Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* *8*, 479-490.
- Madan, B., Madan, V., Weber, O., Tropel, P., Blum, C., Kieffer, E., Viville, S., and Fehling, H.J. (2009). The pluripotency-associated gene *Dppa4* is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. *Mol. Cell. Biol.* *29*, 3186–3203.
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* *1*, 55-70.
- Malik, N. and Rao, M.S. (2013). A review of the methods for human iPSC derivation. *Methods Mol. Biol.* *997*, 23-33.
- Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* *149*, 590-604.
- Maronna, R.A., and Zamar, R.H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* *44*, 307-317.

- Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* *134*, 521-533.
- Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Gottgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* *11*, 491-504.
- Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA* *78*, 7634-7638.
- McKee, A., Minet, E., Stern, C., Riahi, S., Stiles, C.D., and Silver, P.A. (2005). A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev. Biol.* *5*, 14.
- Medeiros, L.A., Dennis, L.M., Gill, M.E., Houbaviy, H., Markoulaki, S., Fu, D., White, A.C., Kirak, O., Sharp, P.A., Page, D.C., and Jaenisch, R. (2011). Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proc. Natl. Acad. Sci. U S A* *108*, 14163-14168.
- Melton, C., Judson, R.L., and Belloch, R. (2010). Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* *463*, 621-626.
- Merrill, M.K., and Gromeier, M. (2006). The double-stranded RNA binding protein 76:Nf45 heterodimer inhibits translation initiation at the rhinovirus type 2 internal ribosome entry site. *J. Virol.* *80*, 6936-6942.
- Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2003). *Manipulating the mouse embryo: A laboratory manual.* (Cold Spring Harbor Press)
- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochizuki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* *26*, 101-106.
- Neplyoueva, V., Dobrikova, E.Y., Mukherjee, N., Keene, J.D., and Gromeier, M. (2010). Tissue type-specific expression of the dsRNA-binding protein 76 and genome-wide elucidation of its target mRNAs. *PLoS One* *5*, e11710.
- Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. (2013). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell* *155*, 121-134.
- Nichols, J., and Smith, A. (2009). Naïve and primed pluripotent states. *Cell Stem Cell* *4*, 487-492.

- Ng, H.H., and Surani, A. (2011). The transcriptional and signaling networks of pluripotency. *Nat. Cell Biol.* *13*, 490-496.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* *99*, 247–257.
- Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* *448*, 313-317.
- Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* *139*, 2288-2298.
- Parchem, R.J., Moore, N., Fish, J.L., Parchem, J.G., Braga, T.T., Shenoy, A., Oldham, M.C., Rubenstein, J.L., Schneider, R.A., and Belloch, R. (2015). miR-302 is required for timing of neural differentiation, neural tube closure, and embryonic viability. *Cell Rep.* *12*, 760-773.
- Parchem, R.J., Ye, J., Judson, R.L., LaRussa, M.F., Krishnakumar, R., Belloch, A., Oldham, M.C., and Belloch, R. (2014). Two miRNA clusters reveal alternative paths in late-stage reprogramming. *Cell Stem Cell* *14*, 617–631.
- Peng, S., Chen, L.L., Lei, X.X., Yang, L., Lin, H., Carmichael, G.G., and Huang, Y. (2011). Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells* *29*, 496–504.
- Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eiriksdottir, E., Langel, U., and Eberwine J. (2006). Immunoprecipitation of mRNA-protein complexes. *Nat. Protoc.* *1*, 577-580.
- Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* *151*, 1617–1632.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* *8*, 2281-2308.
- Rodolfa, K.T., and Eggan, K. (2006). A transcriptional logic for nuclear reprogramming. *Cell* *126*, 652-655.
- Rybak, A., Fuchs, H., Hadian, K., Smirnova, L., Wulczyn, E.A., Michel, G., Nitsch, R., Krappmann, D., and Wulczyn, F.G. (2009). The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. *Nat. Cell Biol.* *11*, 1411–1420.
- Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N., and Nishinakamura, R. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic

stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* 133, 3005–3013.

Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T., and Agata, Y. (2009). The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol. Cell. Biol.* 29, 3754-3769.

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

Shaw, J.P., Utz, P.J., Durand, D.B., Toole, J.J., Emmel, E.A., and Crabtree, G.R. (1988). Identification of a putative regulator of early T cell activation genes. *Science* 241, 202-205.

Shi, L., Zhao, G., Qiu, D., Godfrey, W.R., Vogel, H., Rando, T.A., Hu, H., and Kao, P.N. (2005). NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3'-untranslated region of MyoD and p21WAF/CIP1 mRNAs. *J. Biol. Chem.* 280, 18961-18989.

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

Shim, J., Lim, H., Yates, J.R., and Karin, M. (2002). Nuclear export of NF90 is required for interleukin-2 mRNA stabilization. *Mol. Cell* 10, 1331-1344.

Shyh-Chang, N., and Daley, G.Q. (2013). Lin28: primal regulator of growth and metabolism in stem cells. *Cell Stem Cell* 12, 395–406.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151, 750-764.

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* 12, 645-653.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 15, 259-267.

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

Stadler, B.M., Ivanovska, I., Mehta, K., Song, S., Nelson, A., Tan, Y., Mathieu, J., Darby, C., Blau, C.A., Ware, C., et al. (2010). Characterization of microRNAs involved in embryonic stem cell states. *Stem Cells Dev.* 19, 935–950.

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

Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y., et al. (2004). Human embryonic stem cells express a unique set of microRNAs. *Dev Biol.* *270*, 488–498.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr. Biol.* *11*, 1553–1558.

Tahmasebi, S., Alain, T., Rajasekhar, V.K., Zhang, J.P., Prager-Khoutorsky, M., Khoutorsky, A., Dogan, Y., Gkogkas, C.G., Petroulakis, E., Sylvestre, A., et al. (2014). Multifaceted regulation of somatic cell reprogramming by mRNA translational control. *Cell Stem Cell* *14*, 606–616.

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

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

Takahashi, K., and Yamanka, S. (2012). Induced pluripotent stem cells in medicine and biology. *Development* *140*, 2457–2461.

Tashiro, F., Kanai-Azuma, M., Miyazaki, S., Kato, M., Tanaka, T., Toyoda, S., Yamato, E., Kawakami, H., Miyazaki, T., and Miyazaki, J. (2010). Maternal- effect gene *Ces5/Ooep/Moep19/Floped* is essential for oocyte cytoplasmic lattice formation and embryonic development at the maternal-zygotic stage transition. *Genes Cells* *15*, 813–828.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* *448*, 196–199.

Trapnell, C., Hendrickson, D., Sauvageau, M., Loyal, G., Rinn, J.L., and Pachter, L. (2012). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* *31*, 46–53.

Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M., and Yamanaka, S. (2009). Roles of *Sall4* in the generation of pluripotent stem cells from blastocysts and fibroblasts. *Genes Cells* *14*, 683–694.

Van Nostrand, E.L., Huelga, S.C., and Yeo, G.W. (2016a). Experimental and computational considerations in the study of RNA-binding protein-RNA interactions. *Adv. Exp. Med. Biol.* *907*, 1–28.

- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., and Yeo, G.W. (2016b). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* *13*, 508-514.
- Vidigal, J.A., and Ventura, A. (2012). Embryonic stem cell miRNAs and their roles in development and disease. *Semin. Cancer Biol.* *22*, 428-436.
- Wandrey, F., Montellese, C., Koos, K., Badertscher, L., Bammert, L., Cook, A.G., Zemp, I., Horvat, P., and Kutay, U. (2015). The NF45/NF90 heterodimer contributes to the biogenesis of 60S ribosomal subunits and influences nucleolar morphology. *Mol. Cell. Biol.* *35*, 3491-3503.
- Wang, J, Rao, S., Chu, J., Shen X, Lavasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* *44*, 364-368.
- Wang, T., Xiao, G., Chu, Y., Zhang, M.Q., Corey, D.R., and Xie, Y. (2015). Design and bioinformatics analysis of genome-wide CLIP experiments. *Nucleic Acids Res.* *43*, 5263-5274.
- Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L., and Blelloch, R. (2008). Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* *40*, 1478–1483.
- Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat. Genet.* *39*, 380-385.
- Watanabe, A., Yamada, Y., and Yamanaka, S. (2012). Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *368*, 20120292.
- Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naïve to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* *17*, 155-169.
- Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* *448*, 318-324.
- Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol. Cell* *48*, 195–206.
- Wolkowicz, U.M., and Cook, A.G. (2012). NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res.* *40*, 9356-9368.
- Worringer, K.A., Rand, T.A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., and Yamanaka, S. (2014). The let-7/LIN-41 pathway regulates reprogramming to

human induced pluripotent stem cells by controlling expression of pro-differentiation genes. *Cell Stem Cell* *14*, 40–52.

Wright, J.E., and Ciosk, R. (2013). RNA-based regulation of pluripotency. *Trends Genet.* *29*, 99–107.

Xu, B., Zhang, K., and Huang, Y. (2009). Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA* *15*, 357–361.

Ye, J., and Blelloch, R. (2014). Regulation of pluripotency by RNA binding proteins. *Cell Stem Cell* *15*, 271–280.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* *453*, 519–523.

You, K.T., Park, J., and Kim, V.N. (2015). Role of the small subunit processome in the maintenance of pluripotent stem cells. *Genes Dev.* *29*, 2004–2009.

Young, R.A. (2011). Control of the embryonic stem cell state. *Cell* *144*, 940–954.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* *318*, 1917–1920.

Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L., et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat. Cell Biol.* *8*, 1114–1123.

Zhang, Z., Xiang, D., Heriyanto, F., Gao, Y., Qian, Z., and Wu, W.S. (2013). Dissecting the roles of miR-302/367 cluster in cellular reprogramming using TALE-based repressor and TALEN. *Stem Cell Reports* *1*, 218–225.

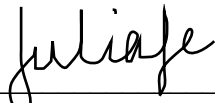
Zheng, G.X., Ravi, A., Calabrese, J.M., Medeiros, L.A., Kirak, O., Dennis, L.M., Jaenisch, R., Burge, C.B., and Sharp, P.A. (2011). A latent pro-survival function for the Mir-290-295 cluster in mouse embryonic stem cells. *PLoS Genet.* *7*, e1002054.

Publishing Agreement

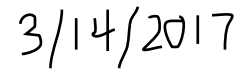
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible 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



Date