Title
RNA-Mediated Transcriptional Activation Of Endogenous Pluripotency Gene Expression

Author
Voutila, Jon Mathew

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

RNA-Mediated Transcriptional Activation Of Endogenous Pluripotency Gene Expression

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular and Medical Pharmacology

by

Jon Mathew Voutila

2014

ABSTRACT OF THE DISSERTATION

RNA-Mediated Transcriptional Activation

Of Endogenous Pluripotency Gene Expression

by

Jon Mathew Voutila

Doctor of Philosophy in Molecular and Medical Pharmacology

University of California, Los Angeles, 2014

Professor Noriyuki Kasahara, Chair

The generation of induced pluripotent stem cells (iPSCs) by the forced overexpression of the Yamanaka factors—Klf4, Oct4, Sox2, and c-Myc—in terminally differentiated cells opened a new frontier in regenerative medicine. These iPSCs can provide a patient-specific source of cells that are able to differentiate into any cell type for possible therapeutic transplantation. However, the most common protocols for iPSC generation involve introduction of exogenous transgenes and genomic integration. Even with non-integrating protocols, iPSC quality and differentiation capacity varies on the reprogramming method used. It is now recognized that small activating RNAs (saRNAs) targeted to the promoter of certain genes can activate endogenous gene expression. Here we demonstrate the design and characterization of saRNAs that activate each of the endogenous Yamanaka genes, and a comprehensive analysis of KLF4 and MYC saRNA activity on global gene expression is presented. An optimized transfection protocol for reprogramming is developed, and we then investigate additional factors that enhance

pluripotency gene expression when combined with saRNA. Knockdown of the histone methyltransferase SUV39H1 with siRNA is shown to dramatically increase saRNA activity on downstream gene activation. This approach was able to successfully generate iPSCs, but the results were not reproducible and the negative control cells also acquired pluripotency. We ultimately conclude that this method is unreliable at producing iPSCs at this stage until additional factors can be identified to improve saRNA activity for successful reprogramming.

The dissertation of Jon Mathew Voutila is approved.

Lily Wu

James Byrne

Christopher Denny

Noriyuki Kasahara, Committee Chair

University of California, Los Angeles

2014

DEDICATION

For my parents, who made this possible with their endless support. Thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

The work for this dissertation was performed under the direction of Noriyuki Kasahara.

Chapter One

Chapter One is a reprint of Voutila J, Sætrom P, Mintz P, Sun G, Alluin J, Rossi JJ, Kasahara N, and Habib NA: Gene Expression Profile Changes After Short-activating RNA-mediated Induction of Endogenous Pluripotency Factors in Human Mesenchymal Stem Cells. *Molecular Therapy Nucleic Acids* 1:e35 (2012).

We thank the City of Hope Bioinformatics Core Facility as well as Ascia Eskin and the UCLA DNA Microarray Core Facility for expert technical assistance with microarray experiments and data analyses, the UCLA Vector Core and Shared Resource Facility for assistance with lentiviral vector construction and production, and the UCLA Virology Core Facility for assistance with p24 ELISA assays. We also would like to thank Kathrin Plath and James Byrne at the UCLA Broad Stem Cell Research Center for their expert advice and very helpful discussion.

Chapter Two

We thank James Byrne for providing primary adult skin cells and reprogramming expertise, as well as the members of the Byrne lab, especially Agustin Vega-Crespo, Patrick Lee, and Jason Awe, for comprehensive assistance and training in iPSC culture. We also thank the UCLA Vector Core and Shared Resource Facility for assistance with lentiviral vector

construction and production, and the UCLA Virology Core Facility for assistance with p24 ELISA assays.

Chapter Three

We thank James Byrne for providing primary adult skin cells and reprogramming expertise, as well as the members of the Byrne lab, especially Agustin Vega-Crespo, Patrick Lee, and Jason Awe, for comprehensive assistance and training in iPSC culture. We also thank the UCLA Vector Core and Shared Resource Facility for assistance with lentiviral vector construction and production, and the UCLA Virology Core Facility for assistance with p24 ELISA assays.

VITA

2007	B.S., Molecular, Cell, and Developmental Biology University of California, Los Angeles Los Angeles, California
2007-2008	Staff Research Associate Department of Pediatrics—Hematology/Oncology UCLA Medical Center Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

Voutila J, Sætrom P, Mintz P, Sun G, Alluin J, Rossi JJ, Kasahara N, and Habib NA: Gene Expression Profile Changes After Short-activating RNA-mediated Induction of Endogenous Pluripotency Factors in Human Mesenchymal Stem Cells. *Molecular Therapy Nucleic Acids* 1:e35 (2012).

Voutila, J, Sætrom P, Mintz P, Sun G, Alluin J, Rossi JJ, Kasahara N, and Habib NA: Induction of Endogenous Pluripotency Factor Expression in Human Mesenchymal Stem Cells by Small Activating RNAs. American Society of Gene and Cell Therapy Annual Meeting, 2011.

Pellegrini M, Cheng JC, Voutila J, Judelson D, Taylor J, Nelson SF, and Sakamoto KM: Expression profile of CREB knockdown in myeloid leukemia cells. *BMC Cancer* 8(1): 264 (2008).

Introduction

The field of regenerative medicine was revolutionized in 2006 with the successful reprogramming of terminally differentiated somatic cells to pluripotent stem cells[1]. These induced pluripotent stem cells (iPSCs) were first generated by retroviral expression of the so-called Yamanaka factors—Oct4, Klf4, Sox2, and c-Myc—and are similar to embryonic stem cells (ESCs) in morphology, gene expression, and differentiation capacity. This new technique overcame many of the pitfalls associated with previous methods of pluripotent cell isolation or generation. While reprogramming of nuclei by somatic cell nuclear transfer (SCNT), first demonstrated in *Xenopus*[2] and more recently in primates[3,4], is able to produce pluripotent cells, ethical issues involved in obtaining and manipulating human oocytes or embryos have limited the use of this method in regenerative medicine. An adult cell nucleus may also be reprogrammed by cell fusion with another pluripotent cell[5], but the resulting tetraploid cells also have restricted applications. The Yamanaka method sidesteps these issues by starting with somatic cells and producing karyotypically normal iPSCs, providing a true patient-specific and clinically relevant source of pluripotent cells.

The explosion of interest in factor-based reprogramming initially focused on understanding the molecular mechanisms at work and improving the efficiency of iPSC generation. Since the initial study using retroviral transduction, numerous additional approaches to force Yamanaka factor expression have been developed, including, but not limited to, adenoviral[6] and Sendai-viral[7] transduction, and plasmid[8], protein[9], and mRNA[10] transfection. Efficiency of iPSC generation depends on the methodology and varies widely, from the initial ~0.01-0.1% to recent reports of near-100% efficiency[11]. The most successful approaches to improve efficiency have focused on opening up chromatin to promote gene expression with

histone deacetylase inhibitors[12], and well as more specific targeted knockdown of histone modifying enzymes that repress gene expression, such as SUV39H1[13,14] and MBD3[11]. These studies have shown that epigenetics are the rate-limiting step in iPSC generation, highlighting the importance of promoting a pluripotent epigenetic state for efficient reprogramming.

As iPSC generation became more efficient, the field started to focus on improving their applicability to the clinic. In Yamanaka's follow-up study demonstrating germline-competency of mouse iPSCs, roughly 20% of the mice developed tumors, and it was determined that this was due to reactivation of the retroviral c-Myc transgene[15]. While it was later shown that c-Myc is not required for successful reprogramming (although, at a significant loss of efficiency)[16], the desire to generate iPSCs without any genomic alteration led to the development of many of the non-integrating methods mentioned previously. Researchers have also utilized Cre-recombinase excisable vectors[17] to combine the benefits of the efficiency of lentiviral-based reprogramming with the safety of non-integrating protocols, but short vector sequences remain after excision that may influence gene expression depending on the integration site. Work has also focused on identifying "safe harbor" sites in the genome where integrations are unlikely to induce oncogenesis[18], improving the potential for iPSCs in gene therapy.

Regardless of reprogramming methodology or efficiency, several issues continue to delay the safe translation of iPSCs to the clinic. Chromosomal aberrations have been reported in both ESCs and iPSCs, frequently due to replicative stress during reprogramming or selective pressure *in vitro*, such as duplications of chromosome 12 resulting in an upregulation of NANOG[19]. These chromosomal abnormalities are not always detected due to the limited resolution of some standard cytogenetic screens and the acquisition of new defects over extended passaging. Even in iPSCs with normal ploidy, incomplete epigenetic reprogramming has significant effects on gene

expression and differentiation potential. iPSCs are similar but not identical to ESCs in their genome-wide DNA methylation[20], histone modifications[21], and subtelomeric 5-hydroxymethylcytosine modifications[22]. This incomplete epigenetic reprogramming leaves an "epigenetic memory" that varies based on reprogramming method and donor cell type, and influences the differentiation potential of the resulting iPSCs[23,24]. A serious consequence of this aberrant gene expression is the potential for the expression of fetal antigens which may make iPSC-derived tissue transplants a target for immune rejection. Teratomas formed by injection of ESCs into syngeneic C57BL/6 mice are not immunogenic, whereas syngeneic iPSCs induced a T-cell-dependent immune response[25]. A later study using transplants of syngeneic terminally differentiated cells demonstrated an immune response due to the aberrant expression of the fetal antigens ZG16 and HORMAD1[26], while another group was unable to detect an immune response to iPSC-derived cells after transplantation[27]. These conflicting results are likely due to the variability of the aberrations obtained during reprogramming and the dependence of these aberrations on donor cell type and reprogramming method. SCNT has been shown to be more effective than factor-based reprogramming in establishing a pluripotent epigenetic state[23], indicating that expression of exogenous transgenes may not be the ideal method for generating clinically-relevant iPSCs.

In addition to the widely-used mechanism of post-transcriptional knockdown of target mRNA[28], there is growing evidence for the involvement of small RNAs in the regulation of transcription. Small promoter-targeted RNAs have also been shown to repress transcription and induce epigenetic changes in eukaryotic cells through a mechanism called transcriptional gene silencing[29,30]. Small RNAs have also been shown to activate gene expression at the transcriptional level[31,32]. This phenomenon has been called RNA activation (RNAa), and is

conserved in rodent and primate cells[33]. The mechanism of action is controversial, but these so-called small activating RNAs (saRNAs) recruit Argonaute proteins and target either the genomic DNA or naturally-occurring antisense transcripts arising from the gene locus[34]. Histone methyltransferases are then recruited to reverse epigenetic silencing and activate transcription of the targeted gene[35–37]. Because RNAa specifically activates endogenous gene expression, it presents an alternative to the forced expression of exogenous transgenes. Manipulation of endogenous cellular pathways and gene regulatory networks with saRNA could provide a more natural method to influence cell fate.

The work described herein explores the possibility of using RNAa as an alternative method for nuclear reprogramming. Using bioinformatic analysis and functional screening, saRNAs for each of the Yamanaka factors were designed and validated *in vitro*. A comprehensive analysis of KLF4 and c-MYC saRNA activity is presented, showing global gene expression changes with saRNA treatment consistent with exogenous gene expression, and ruling out interferon response or antisense RNA involvement in saRNA activity. These saRNAs were then used in combination with a variety of additional siRNAs or other factors to optimize pluripotency gene expression in human cells. In particular, the combined use of siRNA targeting SUV39H1, responsible for H3K9 trimethylation, was found to be highly effective in enhancing saRNA-mediated transcriptional activation levels. The ability of RNAa to successfully generate iPSCs is also demonstrated, but issues with reproducibility necessitate further investigation before this method can be adopted over other alternative protocols. The use of saRNA to substitute for individual pluripotency factors, in combination with vector-mediated delivery of reprogramming genes, may still have utility for mitigating the risk of insertional mutagenesis and re-activation of integrated oncogenic factor expression.

References

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

2. Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* **10,** 622–640 (1962).

3. Byrne, J. A. *et al.* Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* **450,** 497–502 (2007).

4. Tachibana, M. *et al.* Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153,** 1228–1238 (2013).

5. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309,** 1369–1373 (2005).

6. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322,** 945–949 (2008).

7. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **85,** 348–362 (2009).

8. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322,** 949–953 (2008).

9. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4,** 381–384 (2009).

10. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

11. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).

12. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **26,** 795–797 (2008).

13. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).

14. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151,** 994–1004 (2012).

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

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

17. Soldner, F. *et al.* Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136,** 964–977 (2009).

18. Papapetrou, E. P. *et al.* Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat. Biotechnol.* **29,** 73–78 (2011).

19. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7,** 521–531 (2010).

20. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41,** 1350–1353 (2009).

21. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).

22. Wang, T. *et al.* Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. *Nat. Cell Biol.* **15,** 700–711 (2013).

23. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290 (2010).

24. Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat. Biotechnol.* **29,** 1117–1119 (2011).

25. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. Immunogenicity of induced pluripotent stem cells. *Nature* **474,** 212–215 (2011).

26. Araki, R. *et al.* Negligible immunogenicity of terminally differentiated cells derived from induced pluripotent or embryonic stem cells. *Nature* **494,** 100–104 (2013).

27. Guha, P., Morgan, J. W., Mostoslavsky, G., Rodrigues, N. P. & Boyd, A. S. Lack of immune response to differentiated cells derived from syngeneic induced pluripotent stem cells. *Cell Stem Cell* **12,** 407–412 (2013).

28. Hannon, G. J. & Rossi, J. J. Unlocking the potential of the human genome with RNA interference. *Nature* **431,** 371–378 (2004).

29. Matzke, M. *et al.* Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim Biophys Acta* **1677,** 129–41 (2004).

30. Hawkins, P. G. & Morris, K. V. RNA and transcriptional modulation of gene expression. *Cell Cycle* **7,** 602–7 (2008).

31. Li, L. C. *et al.* Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci.* **103,** 17337–17342 (2006).

32. Janowski, B. A. *et al.* Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol.* **3,** 166–173 (2007).

33. Huang, V. *et al.* RNAa Is Conserved in Mammalian Cells. *PLoS ONE* **5,** e8848–e8848 (2010).

34. Portnoy, V., Huang, V., Place, R. F. & Li, L.-C. Small RNA and transcriptional upregulation. *Wiley Interdiscip. Rev. RNA* **2,** 748–760 (2011).

35. Morris, K. V., Santoso, S., Turner, A.-M., Pastori, C. & Hawkins, P. G. Bidirectional Transcription Directs Both Transcriptional Gene Activation and Suppression in Human Cells. *PLoS Genet.* **4,** e1000258–e1000258 (2008).

36. Schwartz, J. C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat. Struct. Mol. Biol.* **15,** 842–848 (2008).

37. Modarresi, F. *et al.* Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol* (2012). doi:10.1038/nbt.2158

Citation: Molecular Therapy–Nucleic Acids (2012) 1, e35; doi:10.1038/mtna.2012.20

www.nature.com/mtna

Gene Expression Profile Changes After Short-activating RNA-mediated Induction of Endogenous Pluripotency Factors in Human Mesenchymal Stem Cells

Jon Voutila[1], Pål Sætrom[2,3], Paul Mintz[4], Guihua Sun[5], Jessica Alluin[5], John J Rossi[5], Nagy A Habib[4] and Noriyuki Kasahara[1,6]

It is now recognized that small noncoding RNA sequences have the ability to mediate transcriptional activation of specific target genes in human cells. Using bioinformatics analysis and functional screening, we screened short-activating RNA (saRNA) oligonucleotides designed to target the promoter regions of the pluripotency reprogramming factors, Kruppel-like factor 4 (KLF4) and c-MYC. We identified KLF4 and c-MYC promoter-targeted saRNA sequences that consistently induced increases in their respective levels of nascent mRNA and protein expression in a time- and dose-dependent manner, as compared with scrambled sequence control oligonucleotides. The functional consequences of saRNA-induced activation of each targeted reprogramming factor were then characterized by comprehensively profiling changes in gene expression by microarray analysis, which revealed significant increases in mRNA levels of their respective downstream pathway genes. Notably, the microarray profile after saRNA-mediated induction of endogenous KLF4 and c-MYC showed similar gene expression patterns for stem cell- and cell cycle-related genes as compared with lentiviral vector-mediated overexpression of exogenous KLF4 and c-MYC transgenes, while divergent gene expression patterns common to viral vector-mediated transgene delivery were also noted. The use of promoter-targeted saRNAs for the activation of pluripotency reprogramming factors could have broad implications for stem cell research.

Molecular Therapy–Nucleic Acids (2012) **1**, e35; doi:10.1038/mtna.2012.20; published online 7 August 2012
Subject Category: siRNAs, shRNAs, and miRNAs

INTRODUCTION

Small RNA molecules are important in the regulation of various molecular and biological activities in the cell, and it is now well known that short RNA sequences play a critical role in regulating the expression levels of specific genes in a targeted manner. RNA interference (RNAi) has become widely recognized to be an important gene regulatory mechanism that causes sequence-specific downregulation of mRNAs.[1] Acting through this mechanism, double-stranded short interfering RNAs (siRNAs) can knock down expression levels *via* the RNA-induced silencing complex, which mediates degradation or translational inhibition of the targeted mRNAs.[2] Moreover, in addition to RNA-induced silencing complex-mediated regulation at the post-transcriptional level, RNAi can also modulate gene transcription itself. In fission yeast, homologs of the RNA-induced silencing complex can regulate chromatin through recruitment of histone-modifying proteins to loci transcribing small noncoding RNA,[3] a mechanism also seen in plants, ciliates, nematodes, and flies. Small promoter-targeted RNAs have also been shown to repress transcription and induce epigenetic changes in eukaryotic cells through a mechanism called transcriptional gene silencing.[4,5]

In human cells, it has recently been reported that short RNAs targeted to the promoter regions of certain genes can activate expression at the transcriptional level.[6,7] This phenomenon has been called RNA activation (RNAa), and has been shown to be conserved in other mammalian species, including mouse, rat, and nonhuman primates.[8] Promoter-targeted small hairpin RNAs have also been shown to efficiently upregulate genes *in vivo*.[9] While the mechanism is not completely understood, it appears that naturally occurring antisense transcripts arising from or near the same genetic locus are able to direct recruitment of Argonaute proteins and histone methyltransferases. Short-activating RNAs (saRNA) may regulate transcription by targeting these antisense transcripts for degradation, resulting in a reversal of this epigenetic silencing and upregulation of sense mRNA.[10–12]

Based on this approach, we have developed a method to design saRNAs for upregulation of specific cellular genes. In the present study, we focused on the feasibility and genetic consequences of upregulating important target genes involved in stem cell regulation and reprogramming. Combined expression of the transcription factors Kruppel-like factor 4 (KLF4), POU5F1 (also called OCT3/4), SOX2, and c-MYC has been shown to reprogram mouse and human

The first two authors contributed equally to this work.
[1]Department of Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, California, USA; [2]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway; [3]Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway; [4]Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK; [5]Division of Molecular Biology, Beckman Research Institute of City of Hope, Duarte, California, USA; [6]Department of Medicine, UCLA School of Medicine, Los Angeles, California, USA. Correspondence: Noriyuki Kasahara, Departments of Medicine and Molecular & Medical Pharmacology, Director, JCCC Vector Shared Resource & CURE Vector Core Facility, University of California, Los Angeles (UCLA), California 90095, USA. E-mail: NKasahara@mednet.ucla.edu
Keywords: gene expression profiling; mesenchymal stem cells; pluripotency genes; RNA activation; saRNA

Received 1 May 2012; revised 1 May 2012; accepted 1 May 2012; advance online publication 7 August 2012. doi:10.1038/mtna.2012.20

fibroblasts into induced pluripotent stem (iPS) cells.[13–16] In particular, KLF4 is important for maintenance of embryonic stem cells,[17] and has been reported to be a master regulator in embryonic stem cells that controls the expression of other pluripotency factors including POU5F1, SOX2, c-MYC, and NANOG.[18] It has previously been reported that direct reprogramming can be achieved in murine fibroblasts with only Klf4, Oct4, and Sox2.[19] However, it has recently been shown that c-Myc is critical for efficient induction of pluripotency in the early phases of reprogramming, by altering the metabolic state of the cell.[20]

Accordingly, in the present study, we employed a genomic–bioinformatic approach to design saRNAs that specifically target two of these key reprogramming genes, KLF4 and MYC. These saRNAs were tested for their ability to upregulate the targeted reprogramming factors, as well as their respective downstream genes, in human mesenchymal stem cells (MSCs), adult bone marrow-derived tissue-specific stem cells that already have multilineage differentiation potential. The effects of saRNA transfection on endogenous gene expression profiles were compared with those resulting from lentiviral vector-mediated overexpression of the exogenous KLF4 and c-MYC transgenes. To date, there have been no studies comprehensively examining the cellular gene expression profiles after saRNA-mediated gene activation of endogenous genes, particularly pluripotency-related genes, and few studies comparing how different reprogramming methods might differentially affect various cellular pathways. These results indicate that the use of saRNA shows significant potential, both as a tool for studying stem cell biology, as well as a safe method to manipulate stem cell gene expression without altering the genome.

RESULTS

Design of saRNA candidate sequences targeting KLF4 and c-MYC

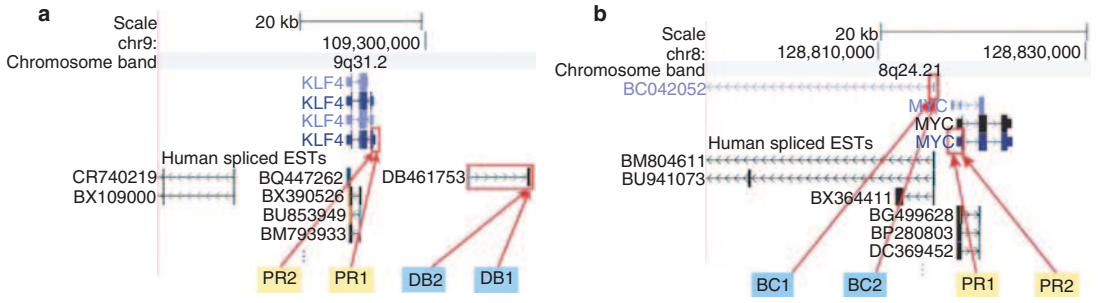
To design saRNA candidates for activation of stem cell factors, we developed a novel bioinformatic approach. The KLF4 gene, located on chromosome 9 (9q31.2), and c-MYC gene, located on chromosome 8 (8q24.21), were our initial targets for activation (**Figure 1a,b**). To identify potential antisense transcripts from the KLF4 and c-MYC loci, we searched the genomic region surrounding each locus for spliced expressed sequence tags (ESTs) which mapped to the positive strand. Although it is usually difficult to determine the transcriptional orientation of ESTs, orientation can be determined by using splice site signatures of spliced ESTs.[21] We found no spliced ESTs that overlapped KLF4, but the scan identified one antisense EST (DB461753) ~15 kb upstream of KLF4's annotated transcription start site (TSS). We were also unable to find a spliced ESTs that overlapped c-MYC, but identified one antisense EST (BC042052) ~2 kb upstream of c-MYC's TSS. These ESTs were then further investigated as potential candidates (**Figure 1c**).

Recent deep sequencing experiments have revealed that antisense RNAs often are found in the region surrounding TSSs.[22–25] Therefore, this region was chosen to design saRNA sequences that targeted potential antisense transcripts from the promoter region. We used the antisense sequence 500 nts upstream and downstream from the TSS (abbreviated KLF4_AS_TSS+/–500 and MYC_AS_TSS+/–500) as a second target candidate. Our goal was to design short sense RNAs that could potentially bind and degrade antisense RNAs generated from the two candidate sequences (ESTs and the regions surrounding TSSs) with the hypothesis that this binding and degradation of the antisense RNAs would activate gene expression. Candidate saRNAs targeting the 500 nt upstream of the TSS were designated PR1, whereas those targeting the 500 nt downstream of the TSS were designated PR2. To give effective antisense targeting and degradation and to minimize off-target effects, we used the GPboost siRNA design algorithm[26] to identify potential short RNAs for downregulating the two candidate sequences.

From the lists of predicted siRNA candidates, we selected the two most promising non-overlapping siRNA target sites on the antisense EST DB461753 and BC042052, and the most promising siRNA target site on each side of the KLF4 and c-MYC TSS within the antisense promoter sequence (KLF4_AS_TSS+/–500 and MYC_AS_TSS+/–500). The candidate siRNAs were selected based on predicted efficacy score from GPboost.[26] We found four potential saRNA candidates for activating each gene (**Figure 1c**).

Upregulation of target gene expression by saRNAs

Reasoning that activation of pluripotency factor gene expression might be more readily achieved in adult tissue-derived stem cells that retain restricted multilineage potential, we then tested whether these saRNA candidates could upregulate KLF4 or c-MYC expression, respectively, in primary human MSCs derived from adult bone marrow. Target gene expression levels were examined by quantitative PCR of reverse-transcribed mRNA from MSC cultures after transfection of each individual saRNA candidate oligonucleotide, as compared with transfection with an Alexa Fluor 555- or FAM-labeled negative control oligo. A transfection efficiency of >95% was determined by flow cytometry and by knockdown with control siRNAs using the same transfection conditions (See **Supplementary Materials** and **Methods** and **Supplementary Figures S1** and **S2**). Initially, none of the saRNA oligos appeared to show any effect on target gene expression after 48 hours. However, upon continued transfection with KLF4-PR1 saRNA every other day, significant upregulation of KLF4 mRNA (**Figure 2a**) was observed, on the order of 2.5-fold over controls by day 4, and reaching approximately fourfold by day 6 of treatment ($P < 0.01$). Target gene mRNA levels after treatment with KLF4-PR1 were significantly higher when MSCs were exposed to saRNA at concentrations of 25 or 50 nmol/l, as compared with 5 nmol/l (**Figure 2b**). Increased Klf4 protein was confirmed in MSCs treated with KLF4-PR1 saRNA (**Figure 2c**, left panel) by western blot analysis, and densitometric quantitation of these blots showed over threefold upregulation of Klf4 protein relative to β-actin internal control (**Figure 2c**, right panel), correlating closely with the level of mRNA upregulation. This time-dependent upregulation of KLF4 gene expression by the KLF4 promoter-targeted PR1 saRNA was observed consistently over multiple experiments. None of the other three KLF4-targeted saRNA candidate sequences was able to upregulate target



c

ID	Target	Pos	Exon	Sense (passenger)	Antisense (guide)
KLF4 -DB1	DB461753	416	2	GACCAUAUUUCUCUUGAAU	AUUCAAGAGAAAUAUGGUC
KLF4 -DB2	DB461753	313	2	ACAAGGCUUCCAUUAAAGA	UCUUUAAUGGAAGCCUUGU
KLF4 -PR1	AS TSS -500	514	n/a	GCGCGUUCCUUACUUAUAA	UUAUAAGUAAGGAACGCGC
KLF4 -PR2	AS TSS+500	26	n/a	CUUCUUUGGAUUAAAUAUA	UAUAUUUAAUCCAAAGAAG
MYC -BC1	BC042052	63	1	GUGACUAUUCAACCGCAUA	UAUGCGGUUGAAUAGUCAC
MYC -BC2	BC042052	31	1	GAGGAGUUACUGGAGGAAA	UUUCCUCCAGUAACUCCUC
MYC -PR1	AS TSS -500	787	n/a	AGCAGUACUGUUUGACAAA	UUUGUCAAACAGUACUGCU
MYC -PR2	AS TSS+500	322	n/a	GAAUUACUACAGCGAGUUA	UAACUCGCUGUAGUAAUUC

Figure 1 Design of KLF4 and c-MYC saRNA candidates. (**a**) KLF4 locus and potential antisense target candidates. The schematic shows the genomic location of KLF4, the structure of the KLF4 transcript, and the spliced ESTs reported from various cell types in the surrounding regions. Red boxes outline the KLF4 promoter region and the closest antisense EST upstream of KLF4 (DB461753). The antisense EST DB461753 initiates roughly 15 kb from KLF4's transcription start site (TSS) and terminates more than 25 kb away. Red arrows indicate the target sites for the short-activating RNA (saRNA) candidates. (**b**) MYC locus and potential antisense target candidates. The schematic shows the genomic location of MYC, the structure of the MYC transcript, and the spliced ESTs reported from various cell types in the surrounding regions. Red boxes outline the MYC promoter region and the closest antisense EST upstream of KLF4 (BC042052). The antisense EST BC042052 initiates roughly 2 kb from MYC's TSS and terminates 50 kb away. Red arrows indicate the target sites for the saRNA candidates. (**c**) saRNA candidates for KLF4 and MYC genes. The list shows the most promising saRNAs against the antisense EST DB461753, BC042052, and saRNAs targeting KLF4 or MYC sequences within a stretch of 500 bp either upstream or downstream of the TSS for each gene. EST, expressed sequence tag; n/a, not applicable.

gene expression, although conversely, KLF4-DB1 appeared to reduce KLF4 mRNA levels in MSCs to about 60% that of scrambled oligo-treated controls by day 6.

Similarly, we found that among the four saRNA candidate sequences targeted to the c-MYC promoter and antisense ESTs, both MYC-PR1 and MYC-PR2 were able to induce consistent upregulation of MYC mRNA (**Figure 2d**) by up to 1.8-fold ($P < 0.01$) and 1.6-fold ($P < 0.01$), respectively, during the 6-day interval of treatment, as compared with scrambled sequence controls. While MYC-BC1 and MYC-BC2 also appeared to affect MYC mRNA levels to some extent, these effects were more modest, and were not consistent between days 4 and 6. Again, target gene mRNA levels after treatment with MYC-PR1 and MYC-PR2 were significantly higher when MSCs were exposed to saRNA at concentrations of 25 or 50 nmol/l, as compared with 5 nmol/l (**Figure 2e**). Image densitometry analysis of western blots from MSCs treated with MYC-PR1 or MYC-PR2 again confirmed upregulation of c-Myc protein by over 2.5-fold relative to β-actin internal control (**Figure 2f**). All western blots were performed in triplicate with replicates shown in **Supplementary Figure S3**.

To determine whether this upregulation of KLF4 and MYC is due to true transcriptional activation, expression levels of nascent RNA were assessed. MSCs were pulsed with ethynyl uridine (EU) during saRNA treatment and total RNA was isolated. Newly transcribed EU-RNA was separated from total RNA by biotinylation of EU in a copper-catalyzed "click" reaction, followed by purification on streptavidin magnetic beads. Quantitative PCR of reverse-transcribed nascent RNA showed that the level of nascent KLF4 mRNA was significantly upregulated with KLF4-PR1 treatment across a 6-day timecourse (**Figure 3a**). The level of nascent MYC mRNA was significantly upregulated with MYC-PR2 treatment, whereas no significant change in nascent MYC mRNA expression was seen with MYC-PR1 treatment (**Figure 3b**).

To verify that these saRNAs do not act through interferon response pathways, interferon response gene expression levels were assayed. No significant upregulation of interferon response genes was detected 24 hours after saRNA treatment (**Figure 3c**). In contrast, overexpression of KLF4 and c-MYC by lentiviral transduction showed dose-dependent induction of interferon response genes (**Supplementary Figure S4**).

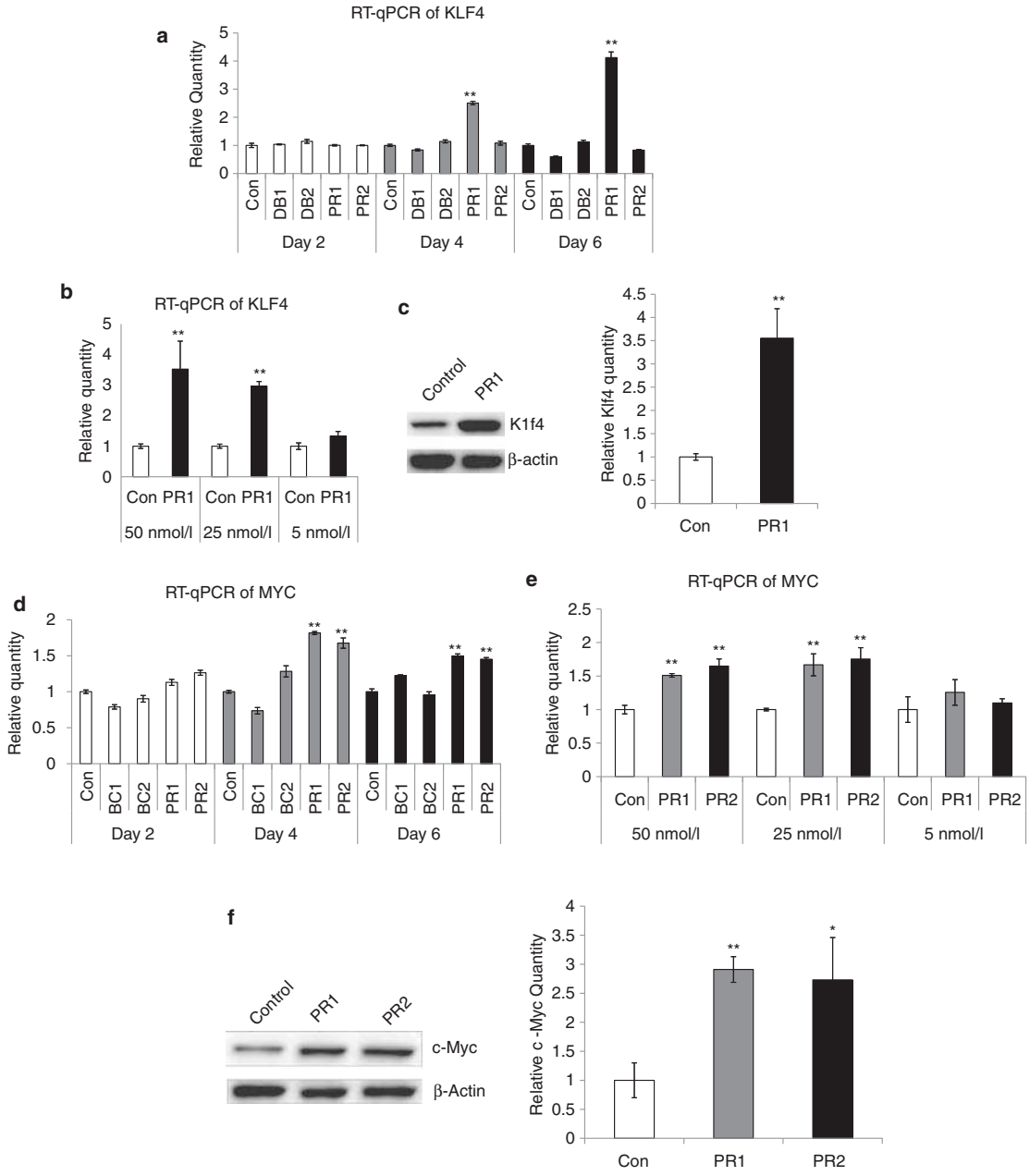


Figure 2 Screening of KLF4 and c-MYC short-activating RNA (saRNA) candidates in mesenchymal stem cells (MSCs). (**a**) RT-qPCR of KLF4 saRNAs treated in MSCs showing increase of KLF4 in KLF4-PR1. (**b**) RT-qPCR of KLF4 for KLF4-PR1–treated MSCs at the indicated saRNA concentrations for 6 days. (**c**) Western blot for Klf4 and β-actin protein and relative quantitation in control- and KLF4-PR1–treated MSCs. (**d**) RT-qPCR of MYC saRNAs treated in MSCs showing increase of c-MYC in MYC-PR1 and MYC-PR2. (**e**) RT-qPCR of MYC for MYC-PR1 and MYC-PR2–treated MSCs at the indicated saRNA concentrations for 6 days. (**f**) Western blot for c-Myc and β-actin protein and relative quantitation of c-Myc in control-, MYC-PR1-, and MYC-PR2–treated MSCs. Asterisks indicate statistical significance: **P* < 0.05; ***P* < 0.01. con, control; RT-qPCR, reverse transcription-quantitative PCR.

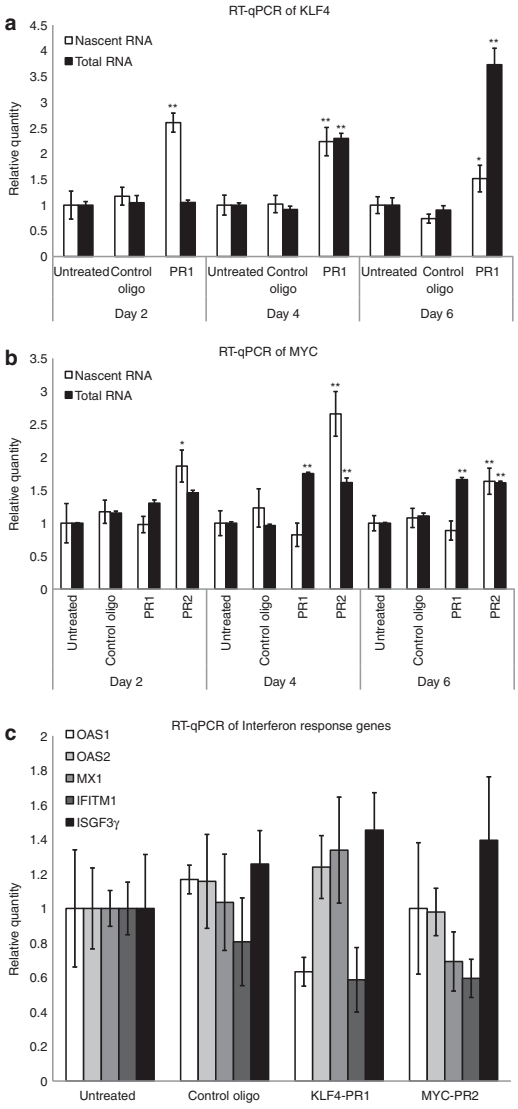


Figure 3 Mechanism of activation by KLF4- and c-MYC–targeted short-activating RNAs (saRNAs). (**a**) RT-qPCR results from KLF4-PR1 saRNA-treated mesenchymal stem cells (MSCs), showing increases in newly transcribed KLF4 mRNA compared to total RNA. (**b**) RT-qPCR results from MYC-PR1- and MYC-PR2–treated MSCs showing increases in newly transcribed MYC mRNA compared to total RNA for MYC-PR2. Asterisks indicate statistical significance: $*P < 0.05$; $**P < 0.01$. (**c**) RT-qPCR results from KLF4- and MYC- saRNA-treated MSCs, showing no significant increase in mRNA levels of key interferon response genes. RT-qPCR, reverse transcription-quantitative PCR.

To further elucidate the mechanism of activation by saRNA, the presence of promoter-associated antisense RNAs was investigated using 5′ Rapid Amplification of cDNA Ends (RACE). Since any antisense RNAs involved in regulating gene expression may not be polyadenylated,[27] random hexamers were used to prime cDNA synthesis from total MSC RNA. Antisense strand-specific primers matching the KLF4-PR1, MYC-PR1, and MYC-PR2 saRNA sequences were used for 5′ RACE to amplify potential antisense RNAs that are targeted by each saRNA and may be involved in regulation of expression. RACE reactions were run on an agarose gel, followed by purification, cloning, and sequencing of any products (See **Supplementary Materials** and **Methods**). Despite the presence of bands from each saRNA primer, each was identified to be a known mRNA with homology to the 3′ end of the saRNA sequence, likely a result of mispriming at permissive annealing temperatures (**Supplementary Figure S5**). Repeats of this assay with more restrictive annealing temperatures yielded no products (data not shown).

To further rule out involvement of possible antisense RNAs arising downstream of each saRNA target sequence, cDNA was synthesized from total MSC RNA using antisense strand-specific primers for KLF4-PR1, MYC-PR1, and MYC-PR2. This cDNA was used as template for PCRs using primers downstream of the saRNA target sequence, which would be in the direction of the 5′ end of any putative antisense transcript. No antisense transcripts were amplified up to 525 bp from the KLF4-PR1 target site (**Supplementary Figure S6a**). Of the priming sites up to 667 bp from MYC-PR1, including and surpassing the MYC-PR2 target site, two products were amplified, cloned, and sequenced corresponding to the 63 and 136 bp downstream of MYC-PR1 (**Supplementary Figure S6b**). The expression level of this antisense transcript was significantly upregulated with MYC-PR2 treatment, whereas its upregulation with MYC-PR1 was not statistically significant ($P = 0.0682$) (**Supplementary Figure S6c**).

To assess the biological significance of saRNA activation of KLF4 and MYC in MSCs, cells were monitored for morphological changes during oligo treatment with KLF4-PR1 and MYC-PR2, which had been confirmed to be associated with increased nascent mRNA transcribed from their respective target genes. After 6 days of treatment, KLF4-PR1–treated MSCs showed marked differences in cell morphology compared with control, whereas MYC-PR2 had modest changes (**Figure 4**). In contrast to scrambled sequence oligo-treated controls, which appeared as progressively more densely packed fibroblastic cells over time, KLF4-PR1–treated MSCs were less confluent and predominantly arranged in clusters with epitheloid cell-like morphology. MYC-PR2–treated cells were less confluent than controls but appeared more heterogeneous with numerous epitheloid cells containing enlarged nuclei.

Thus, target gene mRNA levels after treatment with KLF-PR1 and MYC-PR2, were significantly higher when MSCs were exposed to these saRNA oligos at concentrations of 25 or 50 nmol/l over a 6-day treatment period. Hence, these promoter-targeted saRNA oligos consistently induced increases in both mRNA and protein expression of KLF4 and c-MYC, respectively, in a time- and dose-dependent manner. Neither of these saRNAs activated the interferon response, and both appear to act through specific activation of transcription. We therefore focused on these two positive saRNAs, using the 6-day treatment protocol, in further experiments examining how saRNA-mediated upregulation might affect downstream targets.

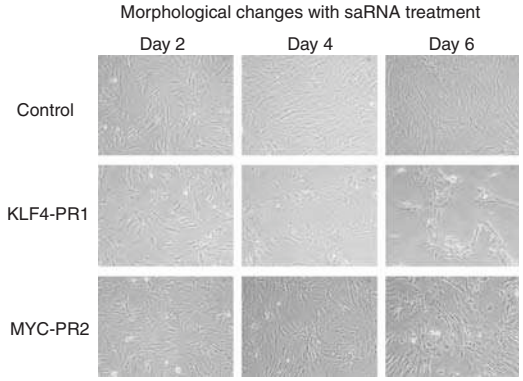


Figure 4 Morphological changes with saRNA treatment. Phase contrast images of MSCs transfected with the indicated saRNA on the indicated day of a 6-day time course. MSC, mesenchymal stem cell; saRNA, short-activating RNA.

Gene expression profile analysis after saRNA-mediated upregulation of KLF4 and c-MYC

Microarray analysis was performed to determine the global gene expression profile and to investigate possible off-target effects with saRNA treatment. Differential gene expression profiles after upregulation of endogenous KLF4 and c-MYC by treatment with their respective saRNAs versus control oligo were examined, and compared with that after overexpression of exogenous KLF4 and c-MYC delivered by viral gene transfer using commercially available second-generation lentivirus vectors. MSCs were transduced with 0.1 pg p24 protein per cell as determined by p24 ELISA, a viral dose that does not result in significant induction of interferon response genes (**Supplementary Figure S4**). Normalized expression data for individual replicates are included in **Supplementary Table S1**. Only those cellular genes showing upward or downward changes in their expression levels at a significance level of at least $P < 0.1$, as compared with their levels in scrambled sequence oligo-treated controls, were included in these analyses.

Interestingly, analysis of the overall gene expression profile in human MSCs after treatment with KLF4 PR-1 (**Figure 5a**) and MYC-PR2 (**Figure 5b**) showed that the majority of cellular genes exhibited concordant changes in expression, but with some notable differences, between upregulation by saRNA and overexpression by lentiviral transduction. For MSCs treated with KLF4-PR1 saRNA, 68% of the cellular genes showing significant changes in their expression levels (971 out of 1,429 genes) showed the same pattern of regulation as Klf4 lentivirus-transduced MSCs. For MSCs treated with MYC-PR2, 64% (273 out of 429 genes) exhibited the same pattern of regulation.

However, this indicates that roughly a one-third of the cellular genes with significant changes in expression after saRNA transfection or lentiviral gene transfer showed discordant regulation between these two groups. To determine whether these differences were due to off-target effects or resulted from the different methods used to activate expression, we used MetaCore pathway analysis from GeneGo (Carlsbad, CA) to determine what pathways were significantly enriched in those genes that were differentially regulated between saRNA and virus samples. Notably, the pathways that were most significantly differentially regulated between KLF4-PR1 saRNA-treated MSCs and KLF4 virus-treated MSCs were those involved in cytoskeletal remodeling, macropinocytosis, and regulation of epithelial-to-mesenchymal transition (EMT), including transforming growth factor (TGF)-β induction of EMT (**Table 1**). The pathways that were most significantly differentially regulated between MYC-PR2 saRNA-treated MSCs and the c-MYC virus-treated MSCs included cell survival and proliferation pathways such as granzyme A signaling, TGF-β regulation, and telomere length, as well as macropinocytosis and EMT pathways (**Table 2**). That is, for KLF4-targeted saRNA versus KLF4-virus, as well as MYC-targeted saRNA versus MYC-virus, in both cases genes involved in the same pathways were found to be discordantly regulated in the same manner. The high degree of similarity in differentially regulated pathways observed with two different sets of saRNAs versus two different lentivirus vectors suggests that the discordant gene expression patterns primarily arise due to characteristic cellular changes in response to oligo transfection versus viral transduction, rather than off-target effects that are shared by each set of saRNAs.

To determine whether narrowing the focus of the differential expression analysis to those types of genes we expect to be regulated by KLF4 and c-MYC would yield greater similarity, we generated lists of genes involved in stem cell maintenance, development, and proliferation, as well as cell cycle-related genes from the AmiGO gene ontology database.[28] Lists of genes used in this analysis are included in **Supplementary Tables S2 and S3**. Heatmaps visualizing the differential expression profiles of KLF4 saRNA-treated and KLF4 virus-treated MSC compared with scrambled oligo-treated control MSC for stem cell-related genes (**Figure 5c**) and cell cycle-related genes (**Figure 5d**) showed much greater similarity, with 74% (26 out of 35) exhibiting concordant regulation for stem cell-related genes, and 80% (132 out of 165) exhibiting concordant regulation for cell cycle-related genes. Heatmaps focused on cell cycle-related genes in MYC-PR2 (**Figure 5e**) saRNA-treated MSC compared to MYC virus-treated MSC also showed a higher degree of similarity, with 67% of cell cycle-related genes (8 out of 12) for MYC-PR2 exhibiting concordant regulation.

Further DAVID gene ontology (GO) analysis[29,30] of all genes revealed that the most significantly enriched GO terms in the KLF4-PR1 saRNA-treated samples were the same as those in the KLF4 virus-treated samples (**Table 3**). Similarly, the majority of the most significantly enriched GO terms in the MYC-PR2 saRNA-treated samples were also significantly enriched in the c-MYC virus-treated samples (**Table 4**).

To further validate the gene expression profile, we chose several well-known transcriptional gene targets of Klf4[31] and c-Myc[32] to verify the results seen in the microarray data by real-time PCR. We found that KLF4-PR1 saRNA transfection resulted in significantly increased expression of the Klf4 target genes cyclin D1, ornithine decarboxylase (ODC1), p21, and p53 (**Figure 6a**); KLF4 virus-transduced cells also showed significantly increased expression of cyclin D1 and p21 to a similar degree, although not of ODC1 or p53. Similarly,

Table 1 Top 10 most significantly enriched pathways among those genes that were differentially regulated between KLF4-PR1 and KLF4 virus samples

Pathway	*P* value
Neurophysiological process: receptor-mediated axon growth repulsion	1.64×10^{-5}
Development: HGF-dependent inhibition of TGF-β–induced EMT	2.23×10^{-5}
Cytoskeleton remodeling: TGF, WNT, and cytoskeletal remodeling	3.68×10^{-5}
Transport: macropinocytosis regulation by growth factors	1.51×10^{-4}
Development: regulation of EMT	1.67×10^{-4}
Cell cycle: ESR1 regulation of G1/S transition	3.18×10^{-4}
Cell adhesion: α-4 integrins in cell migration and adhesion	3.68×10^{-4}
Translation: regulation of EIF4F activity	4.09×10^{-4}
Cell adhesion: plasmin signaling	4.23×10^{-4}
Development: TGF-β–dependent induction of EMT *via* SMADs	4.23×10^{-4}

Abbreviations: EMT, epithelial-to-mesenchymal transition; HGF, hepatocyte growth factor; TGF-β, transforming growth factor-β; SMAD, Drosophila Sma/Mad ortholog.

Table 2 Top 10 most significantly enriched pathways among those genes that were differentially regulated between MYC-PR2 and c-MYC virus samples

Pathway	*P* value
Transcription: role of Akt in hypoxia-induced HIF1 activation	4.66×10^{-4}
Apoptosis and survival: granzyme A signaling	6.40×10^{-4}
Development: PDGF signaling *via* STATs and NF-κB	7.75×10^{-4}
Normal and pathological TGF-β–mediated regulation of cell proliferation	8.49×10^{-4}
Development: regulation of telomere length and cellular immortalization	1.01×10^{-3}
Some pathways of EMT in cancer cells	3.02×10^{-3}
Cell adhesion: ECM remodeling	3.19×10^{-3}
GTP metabolism	3.55×10^{-3}
Transcription: PPAR pathway	5.01×10^{-3}
Transport: macropinocytosis regulation by growth factors	5.48×10^{-3}

Abbreviations: ECM, extracellular matrix; EMT, epithelial-to-mesenchymal transition; GTP, guanosine triphosphate; NF-κB, nuclear factor-κB; PDGF, platelet-derived growth factor; PPAR, peroxisome proliferator-activated receptor; TGF-β, transforming growth factor-β.

Table 3 Top 10 most significantly enriched gene ontology terms among KLF4-PR1 and KLF4 virus samples from all genes with corrected *P* value <0.05 and absolute fold change >2

KLF4-PR1	*P* value	Fold enrichment	KLF4 virus	*P* value	Fold enrichment
Nuclear division	2.16×10^{-17}	6.160865987	M phase of mitotic cell cycle	1.15×10^{-14}	4.594941095
Mitosis	2.16×10^{-17}	6.160865987	Mitotic cell cycle	2.25×10^{-14}	3.574354207
M phase of mitotic cell cycle	3.32×10^{-17}	6.077611041	Cell cycle phase	2.50×10^{-14}	3.396337559
Organelle fission	7.66×10^{-17}	5.917673909	Nuclear division	4.12×10^{-14}	4.535309559
M phase	1.34×10^{-16}	4.848239763	Mitosis	4.12×10^{-14}	4.535309559
Cell cycle phase	7.24×10^{-16}	4.220739263	Organelle fission	1.46×10^{-13}	4.356284182
Mitotic cell cycle	1.97×10^{-15}	4.399661906	M phase	3.72×10^{-13}	3.62313405
Cell cycle process	3.21×10^{-12}	3.223883762	Cell cycle process	1.20×10^{-12}	2.81815595
Cell division	6.05×10^{-12}	4.312307844	Cell cycle	1.84×10^{-11}	2.402392917
Cell cycle	1.62×10^{-10}	2.649960034	Cell division	1.18×10^{-9}	3.275893776

MYC-PR2 saRNA transfection resulted in significantly increased expression of c-Myc target genes ODC1, p21, and p53, as did transduction with c-MYC virus (**Figure 6b**).

Finally, to assess the ability of the KLF4-PR1 saRNA to activate the expression of other reprogramming factors, we analyzed mRNA expression of OCT4, SOX2, and MYC, as well as the stem cell marker NANOG, by real-time PCR. We found that KLF4 activation by saRNA was also able to activate expression of OCT4, SOX2, MYC, and NANOG (**Figure 6c**). Further analysis of OCT4 mRNA isoforms showed that OCT4A was being significantly upregulated, whereas OCT4B showed no difference (**Figure 6d**).

Conversely, the ability of MYC-PR2 saRNA to activate endogenous KLF4 gene expression was also analyzed. In response to MYC-PR2 transfection, there was an approximately twofold increase in KLF4 mRNA levels (**Figure 6e**). Notably, this effect of MYC-PR2 saRNA was corroborated by the microarray analysis (**Supplementary Table S1**).

Table 4 Top 10 most significantly enriched gene ontology terms among MYC-PR2 samples from all genes with corrected *P* value <0.05 and absolute fold change >1.5

MYC-PR2	*P* value	Fold enrichment
Response to nutrient levels[a]	0.00284	8.209112294
Response to extracellular stimulus[a]	0.00422	7.350886918
Response to lipid	0.05296	35.93766938
Tube morphogenesis[a]	0.05605	7.640291915
Embryonic morphogenesis	0.06378	4.241823271
Response to nutrient[a]	0.06653	6.930836237
Regulation of hydrolase activity	0.08013	3.850464576
Collagen metabolic process[a]	0.08119	23.10278746
Multicellular organismal macromolecule metabolic process[a]	0.08949	20.86703383
Response to retinoic acid	0.09499	19.60236511
Response to hormone stimulus[a]	0.09875	3.515641569

[a]Indicate those that are also enriched in c-MYC virus samples.

DISCUSSION

We have been able to identify and characterize two saRNAs, KLF4-PR1 and MYC-PR2, that specifically activate transcription of endogenous KLF4 and MYC genes, respectively, in primary human MSCs. Interestingly, in both cases, it was a promoter-targeted saRNA that gave the desired effect. Given that studies implicate antisense RNAs that overlap the gene of interest as the targets for degradation in RNAa,[10,11] it is conceivable that these promoter-targeted saRNAs are targeting an antisense RNA that has not yet been discovered. However, we were unable to identify any antisense transcripts

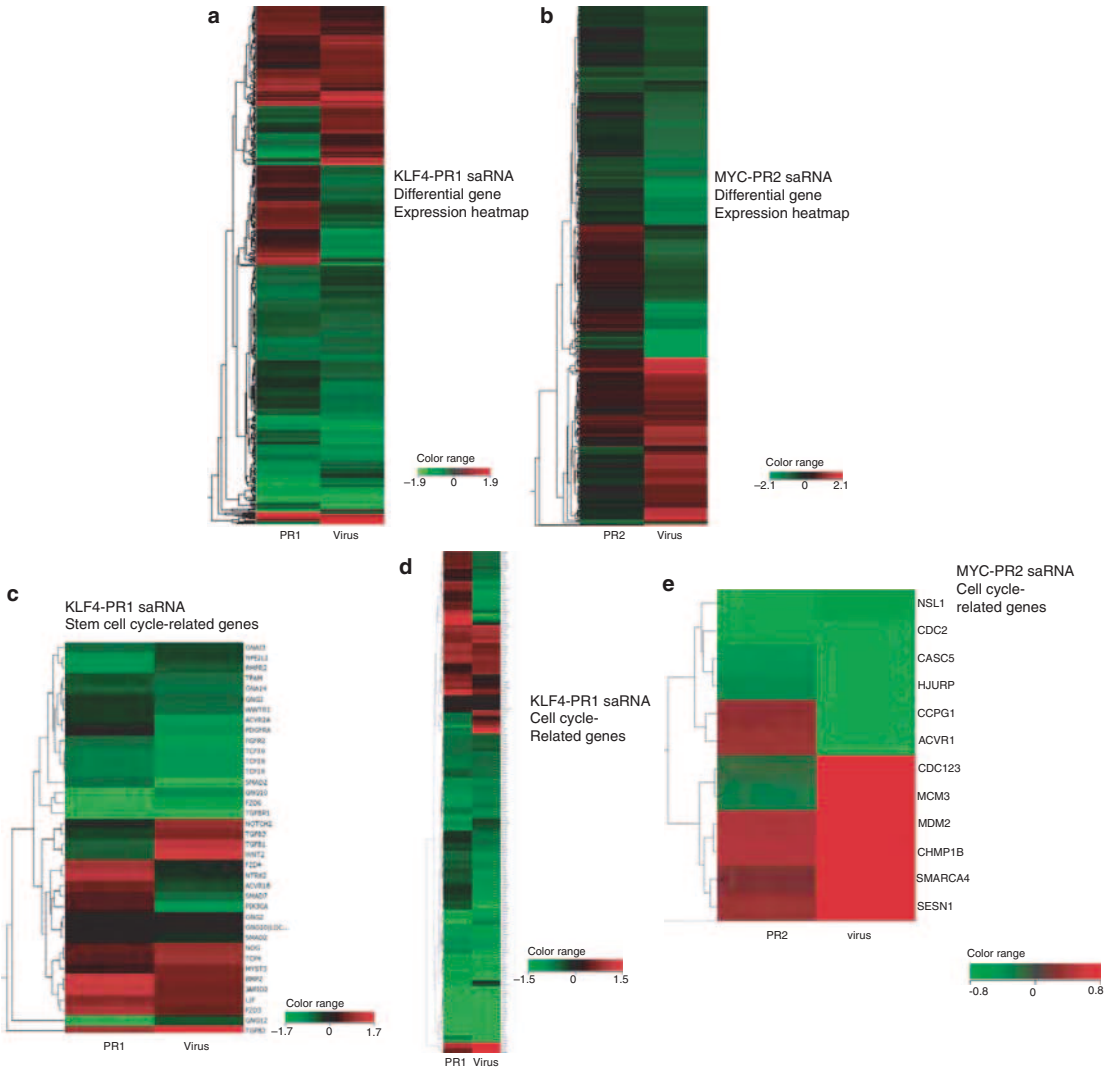


Figure 5 Differential gene expression of saRNA-transfected and virus-transduced samples compared to control. (**a**) Expression of all genes with corrected *P* value <0.05 and absolute fold change >1.5 for KLF4-PR1 and KLF4 virus samples. (**b**) Expression of all genes with corrected *P* value <0.1 and absolute fold change >1.5 for MYC-PR2 and c-MYC virus samples. (**c**) Expression of genes with stem cell-related gene ontology, corrected *P* value <0.1, and absolute fold change >1.5 for KLF4-PR1 and KLF4 virus samples. (**d**) Expression of genes with cell cycle-related gene ontology, corrected *P* value <0.1, and absolute fold change >1.5 for KLF4-PR1 and KLF4 virus samples. (**e**) Expression of genes with cell cycle-related gene ontology, corrected *P* value <0.1 and absolute fold change >1.5 for MYC-PR2 and c-MYC virus samples. saRNA, short-activating RNA.

arising from the region of the KLF4-PR1 target site. Antisense transcripts have been reported to arise from the vicinity of the c-Myc promoter in human cells such as prostate cancer cell lines,[33,34] and we were able to identify one such antisense RNA in primary human MSCs. However, this antisense RNA does not overlap the MYC-PR2 target site and is therefore unlikely to be the RNA target of MYC-PR2. Interestingly, this antisense RNA does overlap the MYC-PR1 target site, but cells transfected with MYC-PR1 saRNA did not show increased transcription of nascent MYC mRNA. Hence it is difficult to ascertain whether MYC-PR1 does target this antisense RNA, perhaps resulting in increased accumulation of MYC sense mRNA at a post-transcriptional level, or if the observed upregulation of MYC mRNA levels by MYC-PR1 is an off-target effect. Due to the relatively small activation of MYC, it would be difficult to completely rule out off-target

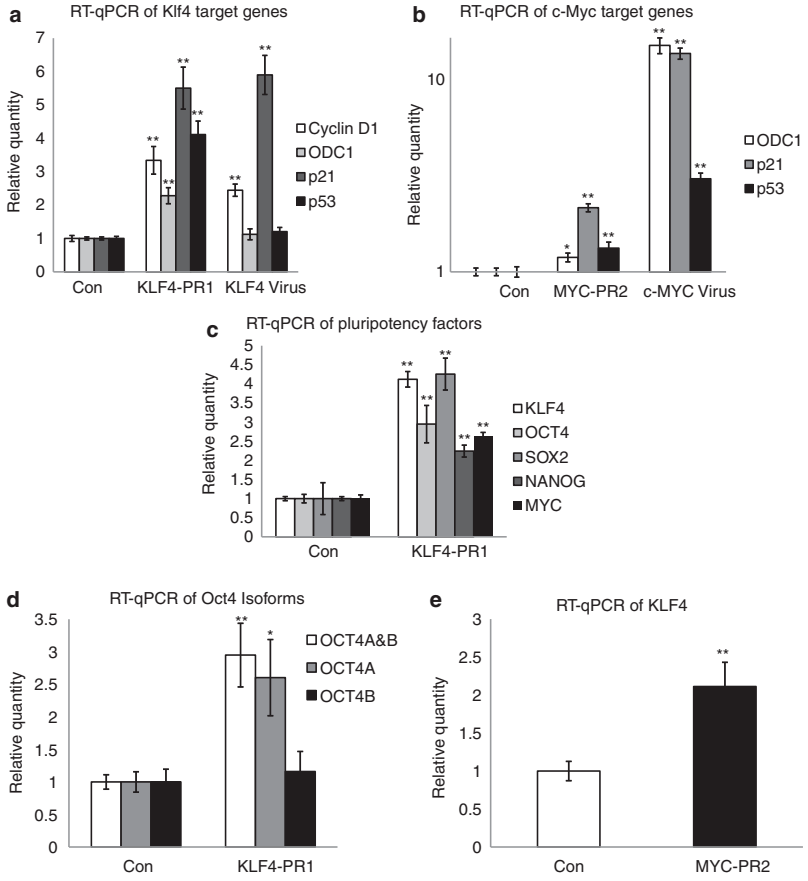


Figure 6 Validation of microarray gene expression of Klf4 and c-Myc target genes. (**a**) RT-qPCR of Klf4 target genes in KLF4-PR1 and Klf4 virus samples relative to control gene expression. (**b**) RT-qPCR of c-Myc target genes in MYC-PR2 and c-MYC virus samples relative to control gene expression. (**c**) RT-qPCR for KLF4-PR1 activation of stem cell and reprogramming factors relative to control gene expression. (**d**) RT-qPCR for KLF4-PR1 activation of OCT4 isoforms relative to control gene expression. (**e**) RT-qPCR for MYC-PR2 activation of KLF4. Asterisks indicate statistical significance: $^{*}P < 0.05$; $^{**}P < 0.01$. con, control; RT-qPCR, reverse transcription-quantitative PCR.

effects even with the use of more sophisticated techniques such as chromatin immunoprecipitation analysis. Furthermore, specific saRNA-targeted effects on MYC expression by transcriptional or post-transcriptional mechanisms, and coexisting off-target effects, are not necessarily mutually exclusive. Although it may not be possible for every gene to be activated by RNAa, our results suggest that this method could be used to generate saRNA candidates for activation of other endogenous genes for which a promoter-associated antisense RNA has not yet been defined.

Focusing on stem cell- and cell cycle-related genes, which were expected to be altered by KLF4 and c-MYC expression, we found the similarities between saRNA-transfected and virus-transduced samples to be quite striking. This was confirmed by GO analysis, as the most significantly enriched terms were nearly identical in KLF4-PR1 oligo and KLF4 virus samples, and most of the top MYC-PR2 terms were also significantly enriched in the c-MYC virus samples. In addition, several well-known KLF4 and c-MYC target genes were validated by real-time PCR, and showed a similar pattern of expression in the saRNA-treated samples as in the virus-transduced samples. Furthermore, we have shown that KLF4 saRNA is able to activate expression of endogenous pluripotency factors including OCT4, SOX2, NANOG, and MYC. Activation of OCT4 was specific to OCT4A, which has been reported to be essential for stemness in human embryonic stem cells.[35] The highly specific nature of OCT4A upregulation in response to transfection of saRNA targeting KLF4 makes this likely to be a true downstream event caused by specific activation of KLF4. Furthermore, the saRNA-treated MSCs showed marked differences in cell morphology, supporting our conclusion that this activation is biologically relevant.

Interestingly, for some downstream gene targets, our results were different from what was expected, as Klf4 is expected to downregulate expression of cyclin D1,

ornithine decarboxylase, and p53,[31] whereas c-Myc is expected to downregulate p21.[32] However, these previously observed results are often cell- and tissue type-dependent, as evidenced by more recent reports showing activation of cyclin D1 and p53 by Klf4,[36,37] and activation of p21 through p53 by c-Myc[38] in different cell types. In addition, our data are consistent across both saRNA- and virus-treated samples in independent microarray and real-time PCR experiments.

As with RNAi, a primary concern for using RNAa to study biological processes is the minimization of potential off-target effects. To evaluate possible off-target effects, we compared the gene expression profile of saRNA activation of KLF4 and MYC to lentiviral-mediated expression. Some significant differences were observed in the total gene expression profile, but this is perhaps not surprising, as one method employed transfection of saRNA and the other lentiviral transduction. Indeed, the use of any transduction method involving introduction of long stretches of exogenous nucleic acids, including both viral vectors as well as plasmids, will likely activate a variety of innate signaling mechanisms.[39,40] This may result in unwanted upregulation of interferons and related genes that may not only cause protein synthesis shutdown and effects on cell proliferation, but may also impair normal stem cell function, as has been described in hematopoietic stem cells.[41] Notably, our interferon response gene expression analysis as well as our microarray analysis of gene functional annotation and differentially regulated pathways showed no evidence of interferon response upregulation after saRNA oligo transfection. In contrast, we observed significant induction of interferon responses upon viral transduction at a dose of 1 pg p24 per cell. This corresponds roughly to an multiplicity of infection of 10, which is within the range of multiplicity of infections typically used in reprogramming protocols. Hence, it may be highly advantageous that RNA duplexes <23 bp in length will not cause induction of interferon,[42] while effecting transcriptional gene activation of endogenous pluripotency factors.

In fact, when we focused on the genes that were differentially expressed between the saRNA and virus samples, the most significantly enriched pathways implicated the use of lentiviral vector-mediated gene transfer as the cause of many of the expression profile changes that were discrepant. Notably, cytoskeletal remodeling, including TGF-mediated remodeling, were upregulated in both KLF4 virus- and c-MYC virus-treated samples. It is well known that retroviruses manipulate the host cytoskeleton to facilitate virus entry and integration.[43] Furthermore, the presence of TGF signaling pathways is also not surprising, considering that carryover of HIV Tat can occur after packaging of second-generation lentivirus vectors, and that Tat protein has been associated with the induction of TGF-β,[44] which likely serves multiple functions for the virus, including immunosuppression and facilitation of cytoskeletal remodeling.[45,46] The differential regulation of macropinocytosis in both KLF4 virus- and c-MYC virus-treated samples is also likely due to viral entry, as both native HIV- and vesicular stomatitis virus-G (VSV-G)–pseudotyped HIV vectors have been reported to use macropinocytosis for entry.[47,48] HIV Tat protein has also been shown to enter cells by macropinocytosis.[49]

Notably, several EMT regulation pathways were observed to be differentially regulated between saRNA- and virus-treated samples, irrespective of whether it was KLF4 or c-MYC being targeted or transduced. This was quite striking, as it has been reported that suppression of EMT signals is required for reprogramming mouse fibroblasts.[8] Klf4 serves to activate the mesenchymal-to-epithelial transition (MET) that is required for reprogramming, whereas Oct4, Sox2, and c-Myc suppress TGF-β–induced EMT. The differential regulation of these pathways in our experiments fits well into this model, as in both cases Klf4 and c-Myc are acting in competition with the TGF-β induced by lentiviral infection. This suggests that using lentiviral vectors to activate reprogramming factors may actually hinder the reprogramming process, as these vectors have been shown to activate TGF-β signaling. This underscores the need for alternative methods of gene activation in reprogramming.

In recent years, invaluable information has been obtained from the use of RNAi to study stem cell biology by inhibiting expression of specific genes involved in the regulation of pluripotency within embryonic as well as somatic stem cells. Here, we have shown the potential of using saRNAs that, conversely, upregulate expression of endogenous genes in stem cells. As each gene can be selectively targeted for activation, the use of saRNAs may also provide a highly useful tool in studying the contribution of individual factors in iPS cell reprogramming. Several studies have used inducible systems to study the reprogramming process,[20,50,51] but these have been limited by the inability to activate or repress the activity of each factor individually. Since RNAa is a transient process, it may be possible to develop an optimized protocol for iPS cell production wherein each factor can be activated when needed by the transfection of its specific saRNA, and similarly removed when it is no longer necessary.

In this context, current methods for upregulating KLF4 and c-MYC require transfection[52] or viral transduction[13–16] of KLF4 or c-MYC expression vectors into cells. As noted above, this study suggests that the TGF-β induced by lentiviral vectors may actually be detrimental to reprogramming. Further, oncogenic reactivation of stably integrated c-MYC transgenes poses a serious safety issue to the use of iPS cells.[53] In addition, evidence that latent viral expression of reprogramming factors impairs normal differentiation of iPS cells,[54] and intolerance to genomic damage caused by exogenous DNA or transposon integration[39,40] further emphasizes the need for a method of iPS cell generation that uses endogenous cellular processes and requires no foreign DNA. In this regard, while reprogramming to full pluripotency has not, to date, been demonstrated with this method, other groups have recently shown saRNA-mediated upregulation of endogenous OCT4 in a breast cancer cell line,[27] and endogenous KLF4 in prostate cancer cell lines.[55] Notably, downstream gene expression and phenotypic changes induced by saRNA-mediated upregulation of KLF4 in prostate cancer cell lines were reported to be comparable to those obtained by retroviral vector transduction. This is consistent with our results obtained in primary human MSCs, and suggests that the use of synthetic saRNA oligos may prove highly advantageous as a safe and efficient alternative for upregulation of endogenous reprogramming genes.

MATERIALS AND METHODS

Bioinformatics and saRNA design. The genes KLF4 and c-MYC were analyzed to design saRNA molecules. Four parameters were used: (i) download target gene annotations; (ii) identify antisense RNA target sequences; (iii) select promoter antisense sequences; and (iv) identify candidate saRNAs. First, the method downloads information about the target's genomic location, orientation, and transcriptional structure from available databases such as the RefSeq database at UCSC (University of California, Santa Cruz). Second, given a database of RNA transcripts with known read direction, such as the UCSC Spliced EST track, our method searches the database for transcripts that are antisense to and in the vicinity of the target gene. More specifically, the method identifies antisense transcripts that (i) overlap the target's promoter and the target mRNA's 5′ end; (ii) overlap the target mRNA; (iii) are at most 20–100 kb upstream of the target's TSS; or (iv) are at most 20–100 kb downstream of the target's polyadenylation site. The method uses these four criteria as hierarchical filters such that if it finds antisense transcripts that for example satisfy criterion (i), the method does not consider the three other criteria. Third, based on the target's TSS, the method downloads the antisense genomic sequence from a fixed size region upstream and downstream of the TSS. The typical region size used by the method is 500 nts upstream and downstream of TSS, but larger or smaller sizes can also be used. Fourth, the method designs siRNAs that give effective and specific downregulation of the antisense target sequence. The method (i) uses a siRNA design algorithm, such as GPboost,[26] to identify candidate effective siRNAs; (ii) removes all candidate siRNAs with aaaa, cccc, gggg, or uuuu motifs and GC content <20% or >55%; (iii) removes all candidates that have Hamming distance <2 to all potential off-target transcripts; and (iv) returns a given number of remaining non-overlapping siRNAs sorted by their predicted siRNA knockdown efficacy. The method returns the two highest scoring saRNAs for a given antisense target sequence.

Cell culture. Bone marrow-derived adult human mesenchymal stem cells (Lonza, Basel, Switzerland) were cultured in the manufacturer's media as instructed. The KLF4, MYC, and control duplex RNA oligonucleotides were transfected into MSCs using Lipofectamine RNAiMAX reagent (Invitrogen, Carlsbad, CA) following the manufacturer's protocol with 30 pmol oligo to 1 µl reagent in a 24-well plate to a final oligo concentration of 50 nmol/l. Transfections were performed every other day for the duration of each experiment. The BLOCK-iT Alexa Fluor Red Fluorescent Control (Invitrogen) and Silencer FAM labeled Negative Control #1 siRNA (Applied Biosystems, Carlsbad, CA), which have no homology to any known gene, were used as negative controls and to assess transfection efficiency by fluorescence microscopy and flow cytometry. Images were taken at ×100 magnification on a Nikon TS100 microscope (Nikon Instruments, Melville, NY).

Plasmids and lentivirus vector production. The plasmids pSin-EF2-KLF4-Pur and pSin-EF2-c-MYC-Pur were generated by cloning human KLF4 and c-MYC transgenes from plasmids pMXs-hKLF4 or pMXs-hcMYC,[56] respectively, into the pSin-EF2-Pur lentiviral vector backbone.[57] VSV-G–pseudotyped second-generation lentivirus preparations were produced using standard protocols; briefly, packaging plasmids pMD2.G, psPAX2, and transfer vector were cotransfected into 293T cells with jetPRIME reagent (Polyplus-transfection, New York, NY), and 48 hours later virus-containing supernatant was collected, filtered, and concentrated by ultracentrifugation. Vector titers were determined by p24 ELISA, performed by the UCLA Virology Core.

Quantitative reverse transcription-PCR. Total RNA was isolated from MSCs using the RNeasy Micro Plus Kit to remove gDNA (QIAGEN, Valencia, CA). RNA was reverse-transcribed to cDNA using the High Capacity cDNA Kit (Applied Biosystems). For nascent RNA analysis, experiments were performed using the Click-iT Nascent RNA Capture Kit (Invitrogen) according to the manufacturer's protocol with a 1 hour EU pulse before sample collection on each day of the experiment. Quantitative real-time PCR was performed using Taqman Gene Expression Master Mix (Applied Biosystems) on a MyiQ2 thermal cycler (Bio-Rad, Hercules, CA) according to the manufacturer's standard protocols. The Taqman primer sets used were as follows: KLF4, Hs00358836_m1; POU5F1 (OCT4A and OCT4B isoform), Hs00999632_g1; POU5F1 (OCT4A isoform), Hs01895061_u1; POU5F1 (OCT4B isoform), Hs00742896_s1; SOX2, Hs00602736_s1; NANOG, Hs02387400_g1; MYC, Hs00153408_m1; CCND1, Hs00277039_m1; CDKN1A, Hs00355782_m1; ODC1, Hs00159739_m1; TP53, Hs99999147_m1; ACTB, Hs00357333_g1 (Applied Biosystems). For interferon response gene expression, primers from the Interferon Response Detection Kit (System Biosciences, Mountain View, CA) were used for SYBR Green real-time PCR with SsoFast EvaGreen Supermix (Bio-Rad). As suggested by the manufacturer's protocol, samples were collected for expression analysis 24 hours after saRNA transfection or viral transduction. Experiments were performed in triplicate wells with at least three replicate reactions per PCR. Expression of β-actin mRNA was used as an internal control and samples were normalized to the scrambled sequence control oligonucleotide or untreated samples. Statistical significance was determined by Student's *t*-test, with P values <0.05 considered significant.

Western blot. Cells were lysed and protein concentration was determined using Coomassie Plus Assay Reagent (Thermo Scientific, Waltham, MA). Each sample was loaded onto a NuPAGE Bis-Tris Gel (Invitrogen) at 30 µg/well and electrophoresed and transferred according to the manufacturer's specifications. Primary antibodies used were GKLF (sc-20691; Santa Cruz Biotechnology, Santa Cruz, CA) and c-Myc (sc-764; Santa Cruz Biotechnology). β-Actin primary antibody (ab8227; Abcam, Cambridge, MA) was used as an internal control and for quantitation. Protein was detected using anti-rabbit HRP conjugated secondary antibody (HAF008; R&D Systems, Minneapolis, MN), developed using Immun-Star WesternC Reagent (Bio-Rad), and visualized on a ChemiDoc XRS+ (Bio-Rad). Blots shown are representative from three replicates. Protein quantitation was performed using Image Lab software (Bio-Rad). Statistical significance was determined by Student's *t*-test, with P values <0.05 considered significant.

Microarray and data analysis. Total RNA was isolated from treated MSCs as described. RNA was processed and hybridized to a GeneChip Human Gene 1.0 ST array (Affymetrix, Santa Clara, CA) in triplicate by the City of Hope Microarray Core Facility (Duarte, CA). Data analysis was performed by the UCLA DNA Microarray Core Facility. Samples were normalized using the ExonRMA16 summarization algorithm and filtered on expression percentile in the raw data (20–100%). Differential expression analysis compared to control samples was performed using an unpaired *t*-test with asymptotic *P* value computation and Benjamini-Hochberg multiple testing correction. Heatmaps were generated using hierarchical clustering using centroid linkage and Euclidean similarity measure. For pathway analysis, lists of genes differentially regulated between saRNA and virus samples were used to generate significantly enriched pathways in MetaCore (version 6.3 build 25177 by GeneGo). For GO analysis, lists of differentially expressed genes with the indicated adjusted *P* values and absolute fold changes were generated for the saRNA- and virus-treated samples versus control. These lists were used to generate functional annotation charts using DAVID bioinformatic analysis with the GOTERM_BP_FAT category on a HuGene-1_0-st-v1 background.[29,30]

Acknowledgments The authors thank the City of Hope Bioinformatics Core Facility as well as Ascia Eskin and the UCLA DNA Microarray Core Facility for expert technical assistance with microarray experiments and data analyses, the UCLA Vector Core and Shared Resource Facility for assistance with lentiviral vector construction and production, and the UCLA Virology Core Facility for assistance with p24 ELISA assays. We also would like to thank Kathrin Plath and James Byrne at the UCLA Broad Stem Cell Research Center for their expert advice and very helpful discussion. The authors declared no conflict of interest.

Supplementary Material

Figure S1. Transfection efficiency of red fluorescent control oligonucleotide by flow cytometry in MSCs.
Figure S2. Transfection efficiency of positive control siRNAs as evaluated by RT-qPCR for expression levels of targeted genes.
Figure S3. Replicates of western blots from **Figure 2c,f**.
Figure S4. RT-qPCR of Klf4 and c-Myc virus transduction in MSCs for interferon response genes at the indicated p24 amount per cell.
Figure S5. 5′-RACE for identification of promoter-associated antisense RNAs.
Figure S6. PCR to identify promoter-associated antisense RNAs.
Table S1. Normalized microarray expression data for each individual replicate used in this study.
Table S2. List of genes with stem cell-related gene ontology, used to generate **Figure 3c**.
Table S3. List of genes with cell cycle-related gene ontology, used to generate **Figure 3d–f**.
Materials and Methods.

REFERENCES

1. Hannon, GJ and Rossi, JJ (2004). Unlocking the potential of the human genome with RNA interference. *Nature* **431**: 371–378.
2. Castanotto, D and Rossi, JJ (2009). The promises and pitfalls of RNA-interference-based therapeutics. *Nature* **457**: 426–433.
3. Moazed, D (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457**: 413–420.
4. Matzke, M, Aufsatz, W, Kanno, T, Daxinger, L, Papp, I, Mette, MF *et al.* (2004). Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim Biophys Acta* **1677**: 129–141.
5. Hawkins, PG and Morris, KV (2008). RNA and transcriptional modulation of gene expression. *Cell Cycle* **7**: 602–607.
6. Li, LC, Okino, ST, Zhao, H, Pookot, D, Place, RF, Urakami, S *et al.* (2006). Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci USA* **103**: 17337–17342.
7. Janowski, BA, Younger, ST, Hardy, DB, Ram, R, Huffman, KE and Corey, DR (2007). Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* **3**: 166–173.
8. Huang, V, Qin, Y, Wang, J, Wang, X, Place, RF, Lin, G *et al.* (2010). RNAa is conserved in mammalian cells. *PLoS ONE* **5**: e8848.
9. Turunen, MP, Lehtola, T, Heinonen, SE, Assefa, GS, Korpisalo, P, Girnary, R *et al.* (2009). Efficient regulation of VEGF expression by promoter-targeted lentiviral shRNAs based on epigenetic mechanism: a novel example of epigenetherapy. *Circ Res* **105**: 604–609.
10. Morris, KV, Santoso, S, Turner, AM, Pastori, C and Hawkins, PG (2008). Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet* **4**: e1000258.
11. Schwartz, JC, Younger, ST, Nguyen, NB, Hardy, DB, Monia, BP, Corey, DR *et al.* (2008). Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol* **15**: 842–848.
12. Modarresi, F, Faghihi, MA, Lopez-Toledano, MA, Fatemi, RP, Magistri, M, Brothers, SP *et al.* (2012). Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol* **30**: 453–459.
13. 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.
14. Takahashi, K, Tanabe, K, Ohnuki, M, Narita, M, Ichisaka, T, Tomoda, K *et al.* (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**: 861–872.
15. Park, IH, Zhao, R, West, JA, Yabuuchi, A, Huo, H, Ince, TA *et al.* (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**: 141–146.
16. Wernig, M, Meissner, A, Foreman, R, Brambrink, T, Ku, M, Hochedlinger, K *et al.* (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**: 318–324.
17. Li, Y, McClintick, J, Zhong, L, Edenberg, HJ, Yoder, MC and Chan, RJ (2005). Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood* **105**: 635–637.
18. Kim, J, Chu, J, Shen, X, Wang, J and Orkin, SH (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**: 1049–1061.
19. Wernig, M, Meissner, A, Cassady, JP and Jaenisch, R (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2**: 10–12.
20. Sridharan, R, Tchieu, J, Mason, MJ, Yachechko, R, Kuoy, E, Horvath, S *et al.* (2009). Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**: 364–377.
21. Yelin, R, Dahary, D, Sorek, R, Levanon, EY, Goldstein, O, Shoshan, A *et al.* (2003). Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* **21**: 379–386.
22. Core, LJ, Waterfall, JJ and Lis, JT (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845–1848.
23. He, Y, Vogelstein, B, Velculescu, VE, Papadopoulos, N and Kinzler, KW (2008). The antisense transcriptomes of human cells. *Science* **322**: 1855–1857.
24. Preker, P, Nielsen, J, Kammler, S, Lykke-Andersen, S, Christensen, MS, Mapendano, CK *et al.* (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851–1854.
25. Seila, AC, Calabrese, JM, Levine, SS, Yeo, GW, Rahl, PB, Flynn, RA *et al.* (2008). Divergent transcription from active promoters. *Science* **322**: 1849–1851.
26. Saetrom, P and Snøve, O Jr (2004). A comparison of siRNA efficacy predictors. *Biochem Biophys Res Commun* **321**: 247–253.
27. Hawkins, PG and Morris, KV (2010). Transcriptional regulation of Oct4 by a long noncoding RNA antisense to Oct4-pseudogene 5. *Transcription* **1**: 165–175.
28. Ashburner, M, Ball, CA, Blake, JA, Botstein, D, Butler, H, Cherry, JM *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.
29. Dennis, G Jr, Sherman, BT, Hosack, DA, Yang, J, Gao, W, Lane, HC *et al.* (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**: P3.
30. Huang, da W, Sherman, BT and Lempicki, RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
31. Evans, PM and Liu, C (2008). Roles of Krüpel-like factor 4 in normal homeostasis, cancer and stem cells. *Acta Biochim Biophys Sin* (Shanghai) **40**: 554–564.

32. Zeller, KI, Jegga, AG, Aronow, BJ, O'Donnell, KA and Dang, CV (2003). An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol* **4**: R69.
33. Napoli, S, Pastori, C, Magistri, M, Carbone, GM and Catapano, CV (2009). Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J* **28**: 1708–1719.
34. Celano, P, Berchtold, CM, Kizer, DL, Weeraratna, A, Nelkin, BD, Baylin, SB *et al.* (1992). Characterization of an endogenous RNA transcript with homology to the antisense strand of the human c-myc gene. *J Biol Chem* **267**: 15092–15096.
35. Cauffman, G, Liebaers, I, Van Steirteghem, A and Van de Velde, H (2006). POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. *Stem Cells* **24**: 2685–2691.
36. Zhu, S, Tai, C, MacVicar, BA, Jia, W and Cynader, MS (2009). Glutamatergic stimulation triggers rapid Krüpple-like factor 4 expression in neurons and the overexpression of KLF4 sensitizes neurons to NMDA-induced caspase-3 activity. *Brain Res* **1250**: 49–62.
37. Wassmann, S, Wassmann, K, Jung, A, Velten, M, Knuefermann, P, Petoumenos, V *et al.* (2007). Induction of p53 by GKLF is essential for inhibition of proliferation of vascular smooth muscle cells. *J Mol Cell Cardiol* **43**: 301–307.
38. Felsher, DW, Zetterberg, A, Zhu, J, Tlsty, T and Bishop, JM (2000). Overexpression of MYC causes p53-dependent G2 arrest of normal fibroblasts. *Proc Natl Acad Sci USA* **97**: 10544–10548.
39. Marión, RM, Strati, K, Li, H, Murga, M, Blanco, R, Ortega, S *et al.* (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**: 1149–1153.
40. Wang, W, Lin, C, Lu, D, Ning, Z, Cox, T, Melvin, D *et al.* (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci USA* **105**: 9290–9295.
41. Sato, T, Onai, N, Yoshihara, H, Arai, F, Suda, T and Ohteki, T (2009). Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. *Nat Med* **15**: 696–700.
42. Reynolds, A, Anderson, EM, Vermeulen, A, Fedorov, Y, Robinson, K, Leake, D *et al.* (2006). Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* **12**: 988–993.
43. Fackler, OT and Kräusslich, HG (2006). Interactions of human retroviruses with the host cell cytoskeleton. *Curr Opin Microbiol* **9**: 409–415.
44. Poggi, A and Zocchi, MR (2006). HIV-1 Tat triggers TGF-β production and NK cell apoptosis that is prevented by pertussis toxin B. *Clin Dev Immunol* **13**: 369–372.
45. Reinhold, D, Wrenger, S, Kähne, T and Ansorge, S (1999). HIV-1 Tat: immunosuppression via TGF-beta1 induction. *Immunol Today* **20**: 384–385.
46. Mandal, S, Johnson, KR and Wheelock, MJ (2008). TGF-beta induces formation of F-actin cores and matrix degradation in human breast cancer cells via distinct signaling pathways. *Exp Cell Res* **314**: 3478–3493.
47. Mercer, J and Helenius, A (2009). Virus entry by macropinocytosis. *Nat Cell Biol* **11**: 510–520.
48. Maréchal, V, Prevost, MC, Petit, C, Perret, E, Heard, JM and Schwartz, O (2001). Human immunodeficiency virus type 1 entry into macrophages mediated by macropinocytosis. *J Virol* **75**: 11166–11177.
49. Wadia, JS, Stan, RV and Dowdy, SF (2004). Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat Med* **10**: 310–315.
50. Stadtfeld, M, Maherali, N, Breault, DT and Hochedlinger, K (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**: 230–240.
51. Brambrink, T, Foreman, R, Welstead, GG, Lengner, CJ, Wernig, M, Suh, H *et al.* (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**: 151–159.
52. Okita, K, Nakagawa, M, Hyenjong, H, Ichisaka, T and Yamanaka, S (2008). Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**: 949–953.
53. Okita, K, Ichisaka, T and Yamanaka, S (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**: 313–317.
54. Papapetrou, EP, Tomishima, MJ, Chambers, SM, Mica, Y, Reed, E, Menon, J *et al.* (2009). Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc Natl Acad Sci USA* **106**: 12759–12764.
55. Wang, J, Place, RF, Huang, V, Wang, X, Noonan, EJ, Magyar, CE *et al.* (2010). Prognostic value and function of KLF4 in prostate cancer: RNAa and vector-mediated overexpression identify KLF4 as an inhibitor of tumor cell growth and migration. *Cancer Res* **70**: 10182–10191.
56. Lowry, WE, Richter, L, Yachechko, R, Pyle, AD, Tchieu, J, Sridharan, R *et al.* (2008). Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci USA* **105**: 2883–2888.
57. Yu, J, Vodyanik, MA, Smuga-Otto, K, Antosiewicz-Bourget, J, Frane, JL, Tian, S *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**: 1917–1920.

CHAPTER TWO

Optimization of saRNA Transfection and Assessment of Additional Factors to Improve Pluripotency Gene Expression for Cellular Reprogramming

A successful small activating RNA (saRNA)-based approach for induced pluripotent stem cell (iPSC) generation will require additional saRNAs activating OCT4 and SOX2. In addition, many of the factors that have been reported to improve reprogramming efficiency, including small molecule compounds and small interfering RNAs (siRNAs), may provide synergistic effects on pluripotency factor gene expression when combined with saRNA. Here we design and screen saRNAs for OCT4 and SOX2, and optimize saRNA transfection to a traditional reprogramming protocol. We also show that the most successful method to improve saRNA activity on downstream gene activation is in combination with a siRNA targeting SUV39H1, a histone methyltransferase responsible for H3K9 trimethylation. Combining SUV39H1 siRNA with the four Yamanaka factor saRNAs results in a peak NANOG mRNA activation of 60-fold over control. This demonstrates that chromatin accessibility is a critical factor reducing saRNA activity, and removing repressive histone marks dramatically improves saRNA function.

Introduction

Since the initial report of exogenous factor-based cellular reprogramming, considerable effort has gone into optimizing the induced pluripotent stem cell (iPSC) generation protocol. Aside from the original Yamanaka factors, additional factors facilitate reprogramming, such as NANOG and LIN28[1,2], and have been shown to increase efficiency when used in combination. Targeting cellular mechanisms that inhibit reprogramming with siRNAs, such as the p53 pathway[3,4], can promote iPSC generation by preventing p53-mediated cell cycle arrest or apoptosis. Similarly, microRNAs (miRNAs), with their ability to regulate the expression of hundreds of genes[5], are also good candidates for modulating pluripotency. By inhibiting miRNAs that negatively regulate the embryonic stem cell (ESC) transcriptional network like miR-145, Yamanaka factor expression is enhanced[6]. Conversely, exogenous expression or transfection of ESC-specific miRNAs can generate iPSCs[7,8]. Small molecules can also improve reprogramming efficiency. Histone deacetylase (HDAC) and DNA methyltransferase inhibitors enhance iPSC generation by promoting an ESC-like transcriptional state, and can even replace KLF4 and/or MYC during reprogramming[9,10].

Given the importance of complete epigenetic reprogramming[11], modulation of specific chromatin-modifying enzymes has demonstrated the most dramatic effects on the efficiency of iPSC generation. Inhibition of DOT1L, a histone methyltransferase (HMT) responsible for H3K79 dimethylation (a mark of active genes), accelerates iPSC generation by facilitating the removal of H3K79me2 on fibroblast-specific genes[12]. Similarly, inhibition of SUV39H1/2, HMTs responsible for H3K9 trimethylation (a mark of repressed genes), accelerates iPSC generation by facilitating the removal of H3K9me3 on pluripotency genes[12,13]. While targeting these genes improved both the speed and efficiency of iPSC generation, inhibition of MBD3, a

subunit of the nucelosome remodeling and deacetylation (NuRD) repressor complex, enabled near-100% reprogramming efficiency within 7 days[14,15]. The importance of epigenetic modifiers was further demonstrated by the report that the relative efficiency of viral and modified mRNA reprogramming approaches was due to the activation of innate immune responses that activate global gene expression changes to enhance chromatin remodeling[16].

In this study, SOX2 and OCT4 small activating RNAs (saRNAs) were designed and screened, and combinatorial effects on gene expression were shown when used in combination with the previously validated KLF4 and MYC saRNAs. These transfection conditions were then adapted to a reprogramming protocol to optimize saRNA activity and limit toxicity with repeated transfections at low seeding densities on feeder cells. To further enhance pluripotency factor gene expression for iPSC generation, these saRNAs were tested in combination with several approaches shown to enhance reprogramming. The most successful improvement of saRNA activity was seen with inhibition of SUV39H1, resulting in a peak NANOG mRNA upregulation of 60-fold compared to control. This finding demonstrates the importance of chromatin accessibility for saRNA activity, and presents a straightforward method to improve the downstream gene expression effects of saRNA.

Results

Design and screen of OCT4 and SOX2 saRNA. As previously described[17], candidate saRNAs were designed for activation of the two remaining Yamanaka factors, OCT4 and SOX2 (Figure 2-1A, B). Given that most cellular reprogramming protocols use fibroblasts and these cells are readily available from patients[18], we decided to use a primary adult human skin-derived fibroblast line, HUF1[19], for screening of candidate saRNA activity. Of the screened oligos, we identified one, OCT4-PR2, that upregulated OCT4 mRNA 4-fold (Figure 2-2A), and another, SOX2-PR1, that upregulated SOX2 mRNA 4.5-fold (Figure 2-2B) by day 6 after 3 transfections of saRNA. The previously identified KLF4-PR1 and MYC-PR2 saRNAs were confirmed to work in these fibroblasts (Figure 2-2C, D), although KLF4-PR1 was only able to upregulate KLF4 mRNA by 1.5-fold, compared to 4-fold in mesenchymal stem cells[17].

Optimization of transfections for reprogramming protocols. When used individually with the previously used RNAiMAX reagent, each saRNA is transfected at a 50nM final concentration in culture. Since these saRNAs would presumably need to be used in combination for successful iPSC generation, we hypothesized that repeated transfections of 200nM total saRNA would eventually induce toxicity and increase the likelihood of off-target effects[20]. We then tested an alternative small RNA transfection reagent, INTERFERin, that is reported to be effective with as little as 1nM RNA. Use of this reagent demonstrated equivalent saRNA activity compared to RNAiMAX at concentrations ranging from 1-50nM (Figure 2-3A). When each saRNA was used in combination for a total saRNA concentration of 4nM (1nM each) over 6 days, a combinatorial effect on gene expression was observed, with upregulation of KLF4 mRNA by 5-fold, SOX2 mRNA by 6.5-fold, and MYC mRNA by 6-fold (Figure 2-3B). Since a stoichiometric ratio of 3:1 of OCT4 expression relative to the other factors has been shown to

improve reprogramming efficiency[21], we tried a combination of 3nM OCT4 saRNA with 1nM of each other saRNA (6nM total), but observed no significant increase in gene expression of any factor (Figure 2-3B).

We next attempted reprogramming HUF1 cells with saRNA using an adapted lentiviral reprogramming protocol with 4 conditions: transfection of control oligo or saRNA, transduction with the four factor stem cell cassette (STEMCCA) lentivirus[22], or transfection of saRNA and transduction of STEMCCA (Figure 2-4A). However, by day 14 it was clear that the saRNA-transfected cells were experiencing toxicity (Figure 2-4B). As no previous toxicity had been observed, we reasoned that the stress of re-seeding the cells on feeders at a low density or repeated transfections for longer than 6 days was responsible. We found that repeated transfections of 4nM oligo were tolerated at a low seeding density only when the culture medium was changed 4 hours after transfection (Figure 2-5A). Attempting this reprogramming protocol with the addition of transfection medium changes solved the toxicity issue, but yielded no iPSC colonies. We decided to use activation of NANOG expression, a master regulatory transcription factor of ESC and iPSCs[23], as a standard to measure and compare the relative effectiveness of any further additions to the reprogramming protocol. With this optimized saRNA transfection regimen alone, we were able to achieve 4-fold activation of NANOG mRNA compared to control (Figure 2-5B).

Evaluation of additional factors to improve pluripotency factor gene expression. The additional factors used to improve pluripotency factor expression to help achieve cellular reprogramming are summarized on Table 2-1. The HDAC inhibitor valproic acid (VPA) was tested at a range of concentrations from 0.1-5.0mM in combination with OCT4-saRNA transfection and no increase in mRNA expression compared to control was observed. A small-

RNA inhibitor of miR-145, which downregulates KLF4, OCT4, and SOX2[6], was also tested with saRNA co-transfection but provided no increase in pluripotency factor gene expression. Given that saRNA transfection does not induce an interferin response[17] and the global epigenetic changes associated with innate immune activation are actually beneficial to reprogramming[16], we next sought to similarly activate a TLR3-mediated innate immune response with transduction of an integrase-deficient lentivirus (IDLV) or treatment with polyinosinic-polycytidylic acid (poly(I:C)). Transduction of a GFP IDLV with saRNA transfection provided only a 2-fold increase in NANOG expression compared to control, while treatment with poly(I:C) actually negated the effect of saRNA, returning NANOG mRNA to the same level as control (Table 2-1).

Since inhibition of chromatin-modifying enzymes provide dramatic increases in reprogramming efficiency[12,13], we next tried inhibition of the HMTs SUV39H1 and DOT1L with siRNA. To improve the chances of successful iPSC generation, we also switched to a primary newborn human fibroblast cell line which was specifically shown to benefit from this reprogramming methodology[13]. While pretreatment of the cells with DOT1L siRNA had no effect on saRNA activity, SUV39H1 siRNA pretreatment dramatically augmented saRNA activity, activating NANOG mRNA 60-fold compared to control (Figure 2-6).

Discussion

The results presented here provide a unique insight into the factors that regulate saRNA activity. The ability of saRNA to activate OCT4 and SOX2 mRNA expression in terminally differentiated fibroblasts suggests that RNAa may be able to activate the expression of genes for which there is a very strong negative regulation. However, fibroblasts do have very weak expression of OCT4, SOX2, and even NANOG mRNA[24], so it is unclear if a basal level of gene expression and chromatin accessibility is required for saRNA to activate a promoter. It should be noted that as with the KLF4 and MYC saRNAs, it was the promoter-targeted sequences that yielded a successful saRNA. While we have targeted the nearest antisense RNA in our design methodology, with the exception of OCT4, there are no known antisense RNAs within 2kb of the transcription start site. It is possible that these RNAs are not in close enough proximity to the promoter to be involved in the regulation of these genes. Interestingly, an antisense RNA has been shown to regulate OCT4 expression, but it is arising from one of OCT4's many pseudogenes[25]. We do not investigate in this study if our OCT4 saRNA is targeting this reported antisense RNA. The possibility that antisense RNAs arising from separate chromosomes can regulate distant genes may necessitate expansion of our saRNA design algorithm parameters.

While our KLF4 saRNA activated KLF4 mRNA 4-fold compared to control in mesenchymal stem cells, only 1.5-fold mRNA overexpression was seen in fibroblasts, contrasting with the MYC saRNA which showed the same activity in both cell types. This indicates that saRNA activity is dependent on cell type, likely due to the different epigenetic and transcriptional regulatory networks active in each cell. In the absence of additional factors, there is likely a limit to the level of transcriptional activation that is possible for a given gene in a given cell type by RNAa. This is supported by the observation of higher KLF4 and MYC mRNA

expression when all four saRNAs were co-transfected, as OCT4, SOX2, and KLF4 have been shown to positively regulate expression of themselves and each other, with KLF4 also activating MYC[26]. While a dose-response with varying levels of saRNA was previously observed[17], it is unclear if this is true at the cellular level or if it is due to the inefficiencies of transfection with smaller amounts of oligo with different transfection reagents. The reagent used in this study demonstrated equivalent saRNA activity to the previous reagent with 25-fold less oligo, with no benefit seen with increasing concentration. A 3:1 stoichiometric ratio of OCT4 saRNA showed no benefit, indicating that at the current levels of oligo concentration are above a threshold where a dose response can be observed. Furthermore, adding more saRNA is not functionally equivalent to adding more virus or mRNA, as the saRNA presumably only has the two endogenous copies of its target gene to act upon.

Many additional factors were tested that have been previously shown to enhance pluripotency factor expression or iPSC generation, with mixed results. VPA was identified early on as an enhancer of iPSC generation[9,10], but some groups have reported little or no benefit to reprogramming or pluripotency factor gene expression from VPA treatment depending on the methodology[27,28]. In the range of VPA concentrations tested here, lower concentrations had no effect on saRNA-induced gene expression, while higher concentrations actually downregulated expression. These inconsistencies could possibly be explained due to VPA-induced production of reactive oxygen species (ROS)[29], which in turn upregulate expression of p53[30], a known inhibitor of reprogramming[3,4]. Indeed, suppression of ROS by the addition of antioxidants during reprogramming enhances mouse and human iPSC generation[31]. Another HDAC inhibitor, sodium butyrate (NaB), also causes an increase in ROS[32], but has been shown to be more effective than VPA in promoting human iPSC generation by upregulation of the ESC-specific

miR302/367 cluster[33]. As both VPA and NaB are non-specific in that they inhibit all of class-I and class-IIb HDACs[34], any perceived benefits of opening chromatin for reprogramming factor accessibility must be weighed against the risks of simultaneously upregulating genes that are detrimental to pluripotency. The genes that may be most affected by HDAC inhibitor treatment are likely to vary based on cell type, culture conditions, and reprogramming methodology.

Indeed, the method used to deliver the Yamanaka factors may have concomitant effects on gene expression that influence reprogramming efficiency. iPSC generation by transfection of *in vitro* transcribed mRNA required base modifications to downregulate the innate immune response to exogenous ssRNA[35]. It was later shown that the lowered, but still present interferon response to the modified mRNAs was actually facilitating reprogramming by a TLR3-mediated induction of global gene expression, including upregulation of HATs and downregulation of HDACs[16]. As we had previously shown that our approach does not activate an interferon response[17], we sought to similarly activate TLR3 with the synthetic dsRNA poly(I:C) or an IDLV. Poly(I:C) treatment with saRNA co-transfection negated the ability of saRNA to activate NANOG mRNA. This could be due to poly(I:C)-induction of an NLRP3 inflammasome response[36], which has been shown to prevent siRNA escape from the endosome[37]. Activation of TLR3 by IDLV transduction halved saRNA-induced NANOG upregulation. The significance of this finding is not clear; as with VPA treatment, a non-specific opening of chromatin may allow reprogramming factor access to repressed pluripotency genes, but genes detrimental to reprogramming may also be activated inadvertently. Furthermore, a nonspecific activation of NANOG mRNA with VPA, poly(I:C), or IDLV treatment may be occurring in the control cells, diminishing the perceived benefits of saRNA without a mRNA copy number analysis. Regardless of whether these attempts improved baseline NANOG expression, they were not

sufficient to generate iPSCs when combined with saRNA, indicating that a more specific approach will help elucidate specific factors that will augment saRNA activity.

In agreement with the common theme that chromatin access is critical for efficient iPSC generation, the most effective approach we tested to improve NANOG activation was specific inhibition of the HMT SUV39H1 with siRNA. This demonstrates that promoting an open chromatin state by inhibiting repressive H3K9me3 modifications significantly raises the level of downstream gene activation by saRNA. Inhibition of DOT1L provided no benefit (and even impaired the effect of SUV39H1 siRNA), contradicting reports that removing H3K79me2 on fibroblast-specific genes aids reprogramming[13]. However, H3K79me2 is a general mark of active genes, not specifically fibroblast genes, so downregulation of this HMT could prevent pluripotency-related genes from being efficiently activated. Similar to treatment with HDAC inhibitors, the benefits of turning off fibroblast specific genes must be weighed against the possibility that DOT1L inhibition may prevent genes necessary for reprogramming from being turned on. It is also difficult to determine at this stage if these various attempts to improve chromatin accessibility are only be helpful in combination with certain reprogramming methodologies or cell types.

We believe one of the greatest strengths of a saRNA-based approach to iPSC generation is its specificity and relative lack of off-target effects. However, there is growing evidence in the field that the off-target effects induced by the introduction of exogenous transgenes activate pathways that are ultimately necessary for efficient reprogramming. Specifically, it is becoming clear that promoting an oocyte- or ESC-like epigenetic state is critical for improving the efficiency and quality of iPSC generation. Until the field determines which particular oocyte-specific factors contribute to the efficiency of nuclei reprogrammed by SCNT[28,38], an saRNA-

based approach may not be able to provide and safe and clinically-relevant method for iPSC generation.

Materials and Methods

Bioinformatics and short activating RNA design

saRNAs were designed as previously described[17]. The genes POU5F1/OCT4 and SOX2 were analyzed to design short activating RNA molecules. Four parameters were used: 1) download target gene annotations; 2) identify antisense RNA target sequences; 3) select promoter antisense sequences; and 4) identify candidate short activating RNAs. First, the method downloads information about the target's genomic location, orientation, and transcriptional structure from available databases such as the RefSeq database at UCSC. Second, given a database of RNA transcripts with known read direction, such as the UCSC Spliced EST track, our method searches the database for transcripts that are antisense to and in the vicinity of the target gene. More specifically, the method identifies antisense transcripts that (a) overlap the target's promoter and the target mRNA's 5' end; (b) overlap the target mRNA; (c) are at most 20 – 100kb upstream of the target's transcription start site (TSS); or (d) are at most 20 – 100 kb downstream of the target's poly-adenylation site. The method uses these four criteria as hierarchical filters such that if it finds antisense transcripts that for example satisfy criterion (a), the method does not consider the three other criteria. Third, based on the target's TSS, the method downloads the antisense genomic sequence from a fixed size region upstream and downstream of the TSS. The typical region size used by the method is 500 nts upstream and downstream of TSS, but larger or smaller sizes can also be used. Fourth, the method designs siRNAs that give effective and specific down-regulation of the antisense target sequence. The method (a) uses a siRNA design algorithm, such as GPboost[39], to identify candidate effective siRNAs; (b) removes all candidate siRNAs with aaaa, cccc, gggg, or uuuu motifs and GC content less than 20% or greater than 55%; (c) removes all candidates that have Hamming distance less than two to all potential off-target transcripts;

and (d) returns a given number of remaining non-overlapping siRNAs sorted by their predicted siRNA knockdown efficacy. The method returns the two highest scoring saRNAs for a given antisense target sequence.

Cell culture

Primary adult human skin-derived fibroblasts ("HUF1")[19] and neonatal human skin-derived fibroblasts ("BJ", ATCC, Manassas, Virginia) were cultured in fibroblast medium containing DMEM supplemented with 10% fetal bovine serum and 1X penicillin/streptomycin. For reprogramming optimization experiments, human embryonic stem cell medium contained DMEM/F12 supplemented with 20% Knockout Serum Replacement (Invitrogen, Carlsbad, California), 1X non-essential amino acids, penicillin/streptomycin, β-mercaptoethanol (Millipore, Billerica, Massachusetts), and 4ng/mL bFGF (Invitrogen). Mitomycin C-inactivated CF-1 mouse embryonic fibroblasts (Globalstem, Rockville, Maryland) were used as a feeder layer when indicated. RNA oligonucleotides were transfected using Lipofectamine RNAiMAX (Invitrogen, Carlsbad, California) or INTERFERin reagent (Polyplus-transfection, New York, New York), where indicated, following the manufacturer's protocol with a final oligo concentration of 1-50nM depending on the experiment. Transfections were performed every other day unless otherwise indicated. The BLOCK-iT Alexa Fluor Red Fluorescent Control (Invitrogen), which has no homology to any known gene, was used as a negative control.

Plasmids and lentivirus production

The stem cell cassette (STEMCCA)[22] lentiviral reprogramming vector was used to make VSV-G pseudotyped second-generation lentivirus preparations using standard protocols; briefly,

packaging plasmids pMD2.G, psPAX2, and transfer vector were cotransfected into 293T cells (ATCC) with jetPRIME reagent (Polyplus-transfection), and 48 hours later virus-containing supernatant was collected and filtered. Vector titers were determined by p24 ELISA, performed by the UCLA Virology Core, and immunocytochemistry for Oct4 protein expression (sc-5279, Santa Cruz Biotechnology, Dallas, Texas). Integrase-deficient lentivirus was produced by replacing psPAX2 with the integrase mutant pCMVdeltaR8.2int(-) packaging plasmid.

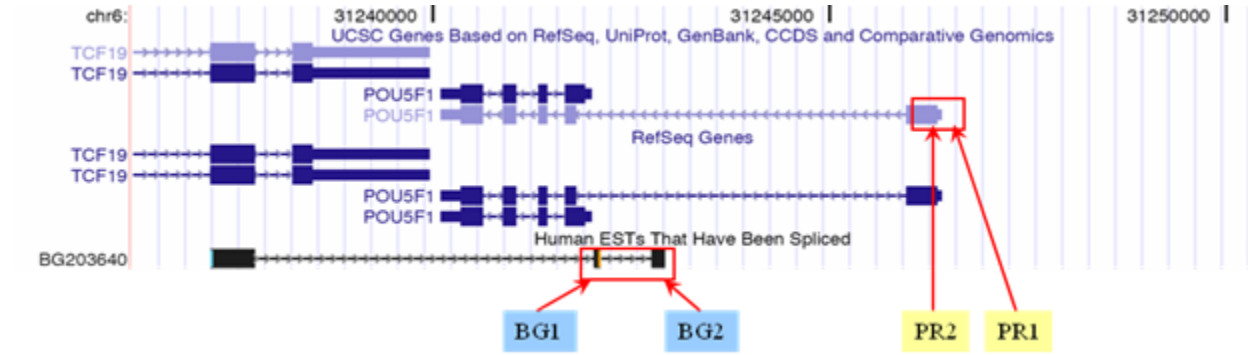
iPSC protocol optimization

Unless otherwise specified, the general reprogramming protocol was as follows: fibroblasts were seeded at a density of $1x10^4$ cells/cm^2 and were transfected with saRNA starting the following day and every other day for the duration of the experiment. STEMCCA lentivirus was used for a positive control and these cells were transduced the day after seeding in medium containing polybrene at 4µg/mL. On the 5^{th} day following the initial transfection or transduction, cells were replated at a density of $1.7x10^3$ cells/cm^2 on a MEF feeder layer. On the day following re-seeding, medium was changed to hESC medium with transfections continuing every other day. Poly(I:C) was added to the culture medium at 300ng/mL during the first three transfections[16]. Cells were transduced with GFP IDLV 4 hours after the initial transfection. HMT siRNAs targeting DOT1L (SASI_Hs01_00161195, Sigma-Aldrich, St. Louis, Missouri) and SUV39H1 (SASI_Hs02_00335192, Sigma-Aldrich) were transfected at 1nM three days before and again one day before the first saRNA transfection.

Quantitative RT-PCR

Total RNA was isolated using the Aurum Total RNA Mini Kit (Bio-Rad, Hercules, California), which included a DNase I digestion step to remove genomic DNA. RNA was reverse-transcribed to cDNA using the High Capacity cDNA Kit (Applied Biosystems). Quantitative real-time PCR was performed using SsoFast Probes Supermix (Bio-Rad) on a MyiQ2 thermal cycler (Bio-Rad) according to the manufacturer's standard protocols. The Taqman primer sets used were as follows: KLF4, Hs00358836_m1; OCT4 (POU5F1), Hs00999632_g1; SOX2, Hs00602736_s1; NANOG, Hs02387400_g1; MYC, Hs00153408_m1; ACTB, Hs00357333_g1 (Applied Biosystems). Experiments were performed in triplicate wells with at least 3 replicate reactions per PCR. Expression of β-actin mRNA was used as an internal control and samples were normalized to the scrambled sequence control oligonucleotide-treated samples.

A



B

Figure 2-1. Design of OCT4 and SOX2 saRNA candidates. (**A**) This panel shows the genomic location of OCT4 (POU5F1) and spliced expressed sequence tags (ESTs) from the surrounding regions (image from the UCSC genome browser[40]). Red boxes outline OCT4 (POU5F1) and the antisense EST with the most overlap with OCT4 (BG203640). The antisense EST BG203640 originates downstream of OCT4 and overlaps with about half the OCT4 primary transcript. (**B**) This panel shows the genomic location of SOX2 and spliced ESTs from the surrounding regions (image from the UCSC genome browser[40]). Red boxes outline SOX2 and the closest antisense EST upstream of SOX2 (BG220229).

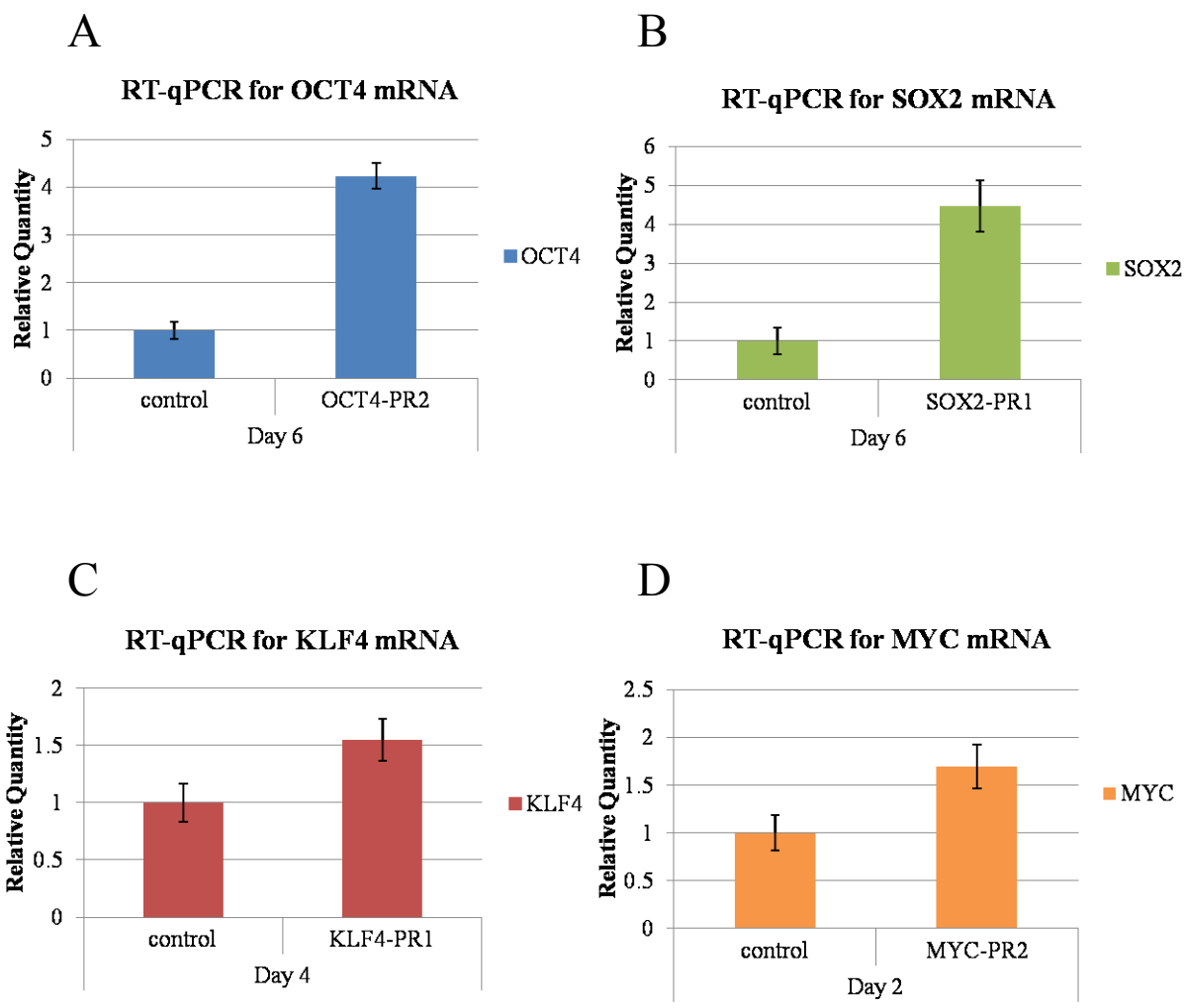
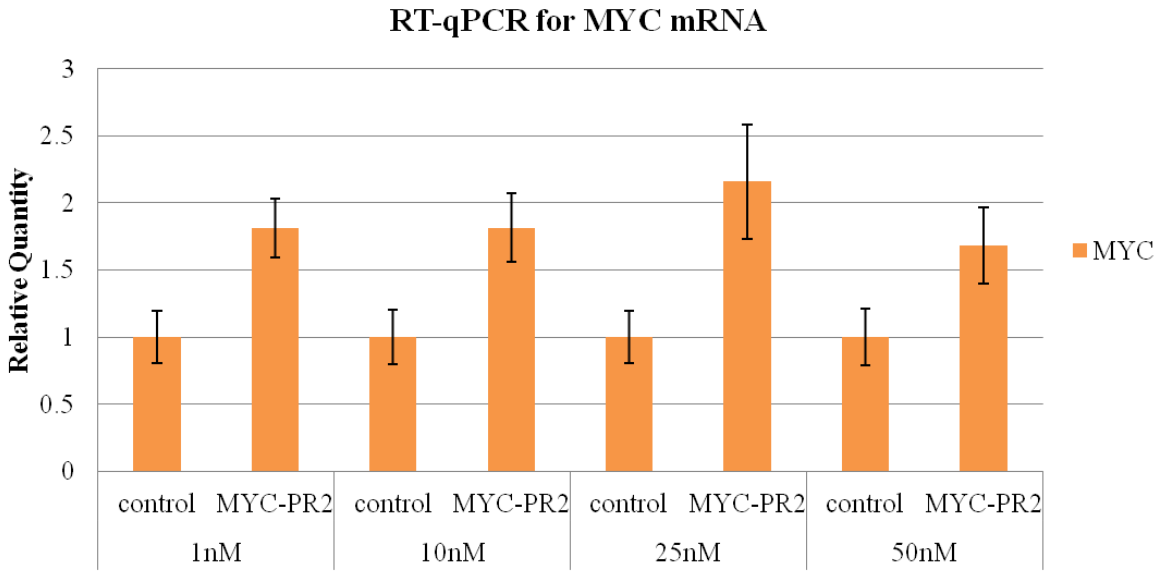


Figure 2-2. Screen of saRNA in primary adult human fibroblasts. (A) RT-qPCR for OCT4 mRNA showing an increase with OCT4-PR2 transfection. **(B)** RT-qPCR for SOX2 mRNA showing an increase with SOX2-PR1 transfection. **(C)** RT-qPCR for KLF4 mRNA showing an increase with KLF4-PR1 transfection. **(D)** RT-qPCR for MYC mRNA showing an increase with MYC-PR2 transfection.

A



B

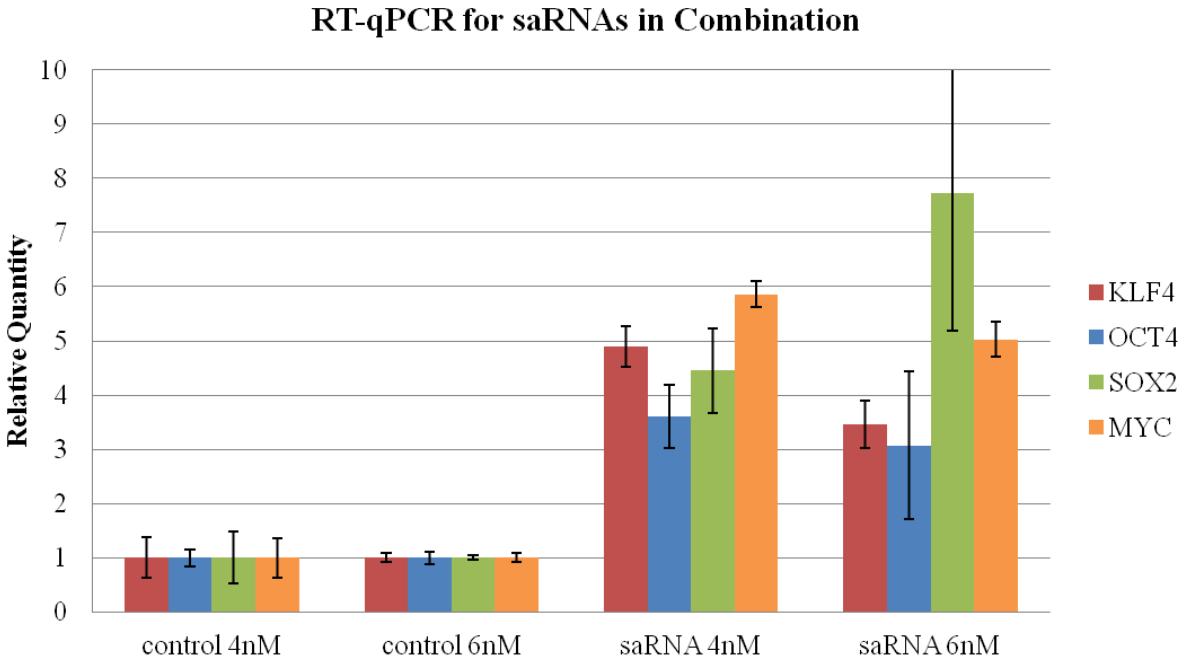


Figure 2-3. Testing saRNA activity with INTERFERin transfection reagent. (A) RT-qPCR for MYC mRNA two days after MYC-PR2 transfection at the indicated saRNA concentration showing equivalent MYC activation at each dose. **(B)** RT-qPCR for each factor on day 6 after 3 cotransfections of either 1nM each saRNA (4nM total) or 3nM OCT4-PR2 saRNA with 1nM of each other saRNA (6nM total). Similar activation is shown regardless of stoichiometry.

A

	STEMCCA lentivirus	**STEMCCA lentivirus + 4nM saRNA**	**4nM saRNA**	**Negative control**
Day -1	Seed cells	Seed cells	Seed cells	Seed cells
Day 0	Transduce	Transduce, transfect saRNA	Transfect saRNA	Change medium
Day 1	Change medium	Change medium	Change medium	Change medium
Day 2	Change medium	Transfect saRNA	Transfect saRNA	Change medium
Day 3				
Day 4	Change medium	Transfect saRNA	Transfect saRNA	Change medium
Day 5	Replate on feeders	Replate on feeders	Replate on feeders	Replate on feeders
Day 6, every other day	Change to ESC medium	Change to ESC medium, transfect saRNA	Change to ESC medium, transfect saRNA	Change to ES medium

B

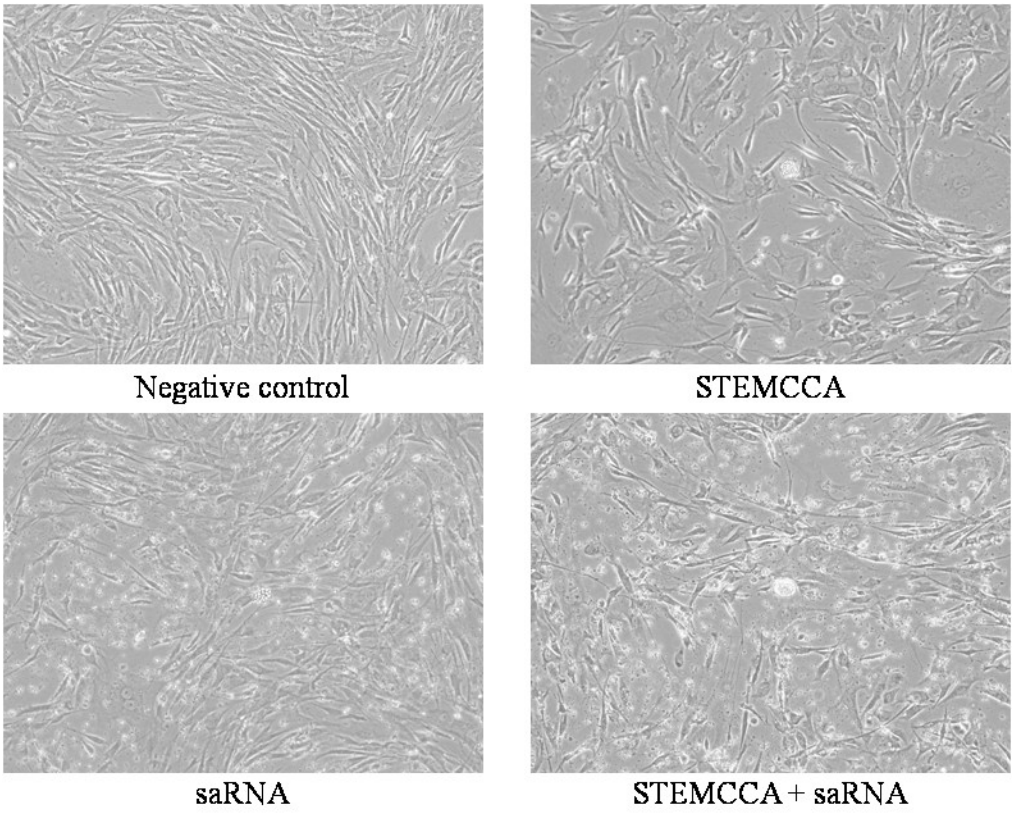
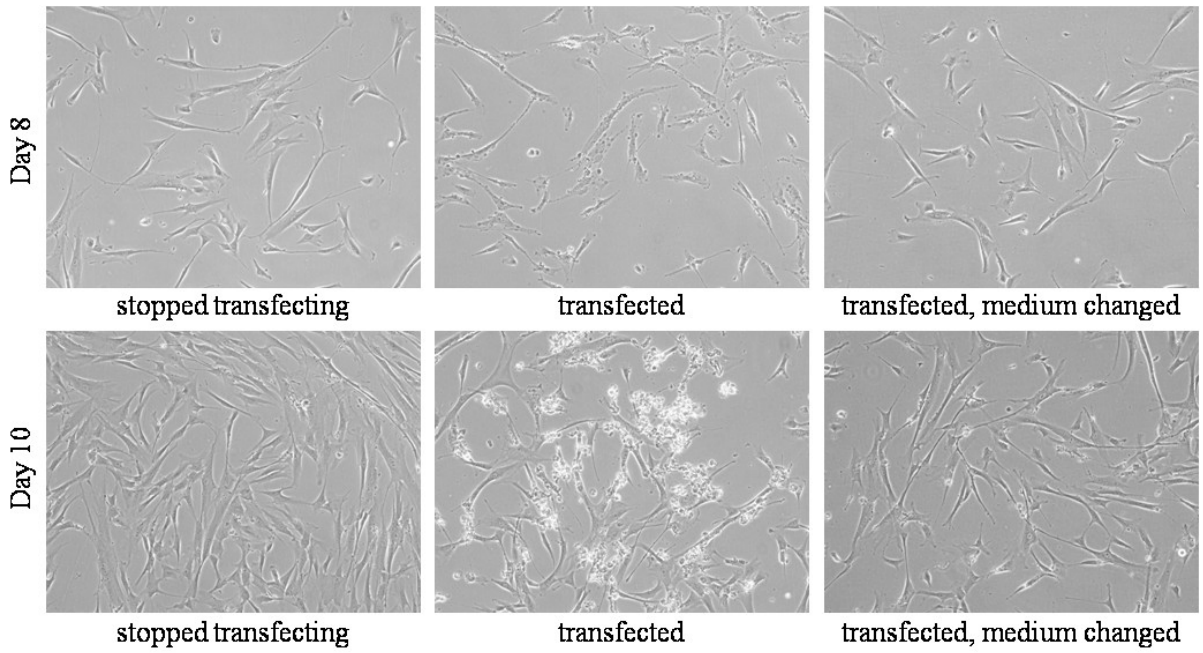


Figure 2-4. Initial saRNA reprogramming experiment. (A) Outline of experimental protocol with each condition tested. **(B)** Phase contrast micrographs of each condition on day 14 of the experiment. Significant cell debris can be seen in the transfected cells.

A



B

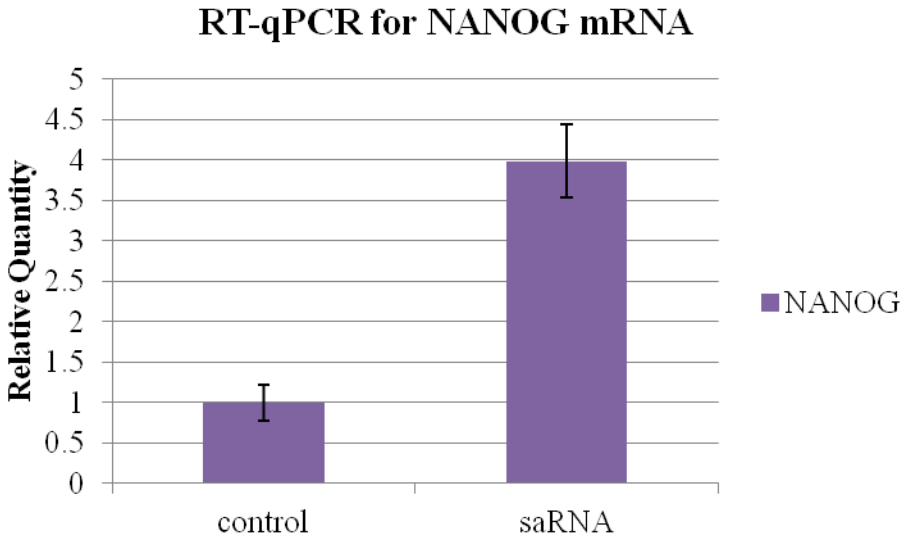


Figure 2-5. Optimization of transfection conditions for reprogramming. (A) Phase contrast micrographs of fibroblasts on day 8 and 10. Cells were transfected at day 0, 2, and 4, replated at low density on day 5, and were transfected with no medium change, transfected and medium changed after 4 hours, or were no longer transfected. Significant cell death is seen without medium change at low density. **(B)** Repeat of previous reprogramming experiment with medium changes. RT-qPCR on day 14 showing a peak NANOG mRNA activation of 4-fold.

Factor Tested	Effect
Valproic acid	No change in saRNA activity from 0.1-1.5mM, downregulation at 2-5mM
miR-145 inhibition	No change in saRNA activity
Activation of innate immunity	
Poly(I:C)	Loss of saRNA-mediated NANOG activation
GFP IDLV	No change in saRNA activity
HMT Inhibition	
DOT1L siRNA	No change in saRNA activity
DOT1L siRNA + SUV39H1 siRNA	40-fold NANOG mRNA activation
SUV39H1 siRNA	60-fold NANOG mRNA activation

Table 2-1. Summary of factors tested to improve saRNA activity. Shown are the results of attempted improvements to the reprogramming protocol to enhance saRNA activity as assayed by RT-qPCR for NANOG mRNA. No change indicates 4-fold NANOG mRNA activation was observed, similar to saRNA transfection alone. Loss of NANOG activation indicates NANOG mRNA level was equal to negative control.

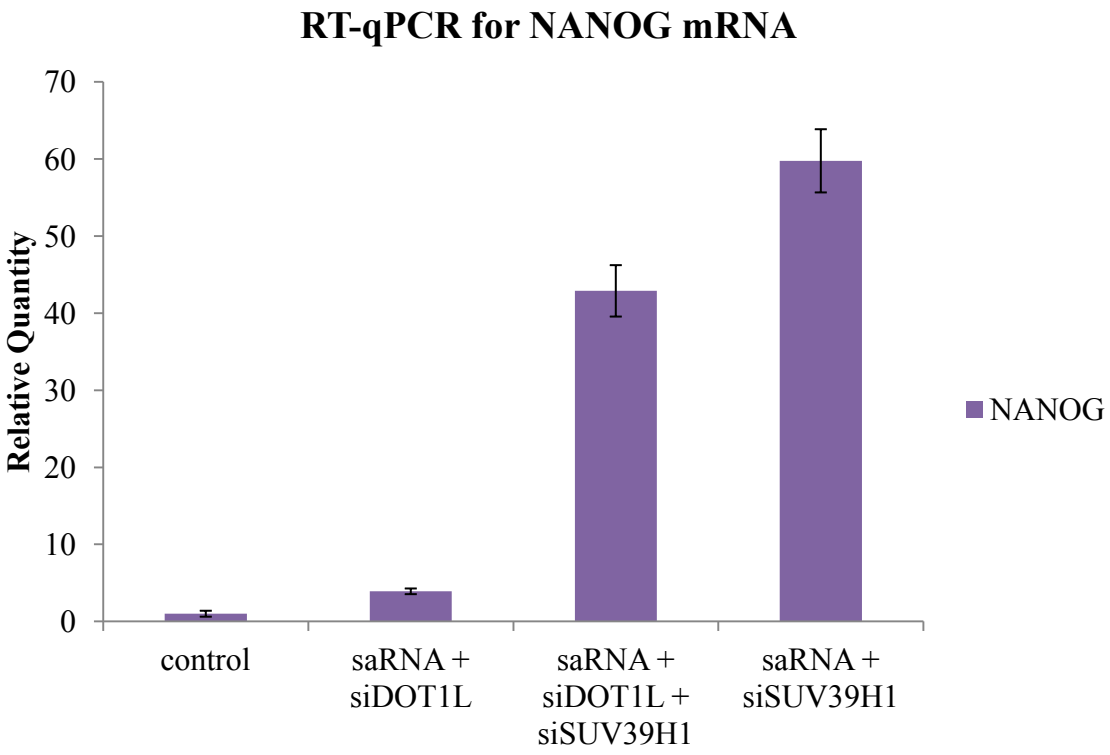


Figure 2-6. Activation of NANOG mRNA when saRNA is combined with HMT siRNA. RT-qPCR for NANOG mRNA on day 14 of a reprogramming experiment showing improvement in NANOG activation when fibroblasts are pretreated with the indicated HMT siRNAs.

References

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

2. Liao, J. *et al.* Enhanced efficiency of generating induced pluripotent stem (iPS) cells from human somatic cells by a combination of six transcription factors. *Cell Res.* **18,** 600–603 (2008).

3. Zhao, Y. *et al.* Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* **3,** 475–479 (2008).

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

5. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37,** 495–500 (2005).

6. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. & Kosik, K. S. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* **137,** 647–658 (2009).

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

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

9. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **26,** 795–797 (2008).

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

11. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290 (2010).

12. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).

13. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151,** 994–1004 (2012).

14. Luo, M. *et al.* NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells Dayt. Ohio* **31,** 1278–1286 (2013).

15. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).

16. Lee, J. *et al.* Activation of innate immunity is required for efficient nuclear reprogramming. *Cell* **151,** 547–558 (2012).

17. Voutila, J. *et al.* Gene Expression Profile Changes After Short-activating RNA-mediated Induction of Endogenous Pluripotency Factors in Human Mesenchymal Stem Cells. *Mol. Ther. Nucleic Acids* **1,** e35 (2012).

18. Vierbuchen, T. & Wernig, M. Molecular roadblocks for cellular reprogramming. *Mol. Cell* **47,** 827–838 (2012).

19. Awe, J. P. *et al.* Generation and characterization of transgene-free human induced pluripotent stem cells and conversion to putative clinical-grade status. *Stem Cell Res. Ther.* **4,** 87 (2013).

20. Caffrey, D. R. *et al.* siRNA off-target effects can be reduced at concentrations that match their individual potency. *PloS One* **6,** e21503 (2011).

21. Papapetrou, E. P. *et al.* Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc. Natl. Acad. Sci.* **106,** 12759–12764 (2009).

22. Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells Dayt. Ohio* **27,** 543–549 (2009).

23. Li, Y.-Q. Master stem cell transcription factors and signaling regulation. *Cell. Reprogramming* **12,** 3–13 (2010).

24. Page, R. L. *et al.* Induction of Stem Cell Gene Expression in Adult Human Fibroblasts without Transgenes. *Cloning Stem Cells* **11,** 417–426 (2009).

25. Hawkins, P. & Morris, K. V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* **1,** 1–11 (2010).

26. Kim, J. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* **132,** 1049–1061 (2008).

27. Hong, J., Park, S. K. & Roh, S. Effect of Valproic acid, a Histone Deacetylase Inhibitor, on the Expression of Pluripotency and Neural Crest Specific Marker Genes in Murine Multipotent Skin Precursor Cells. *Int. J. Oral Biol.* **35,** 209–214 (2010).

28. Shinagawa, T. *et al.* Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. *Cell Stem Cell* **14,** 217–227 (2014).

29. Kawai, Y. & Arinze, I. J. Valproic acid-induced gene expression through production of reactive oxygen species. *Cancer Res.* **66,** 6563–6569 (2006).

30. Macip, S. *et al.* Influence of induced reactive oxygen species in p53-mediated cell fate decisions. *Mol. Cell. Biol.* **23,** 8576–8585 (2003).

31. Esteban, M. A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6,** 71–79 (2010).

32. Gospodinov, A., Popova, S., Vassileva, I. & Anachkova, B. The inhibitor of histone deacetylases sodium butyrate enhances the cytotoxicity of mitomycin C. *Mol. Cancer Ther.* **11,** 2116–2126 (2012).

33. Zhang, Z. & Wu, W.-S. Sodium butyrate promotes generation of human induced pluripotent stem cells through induction of the miR302/367 cluster. *Stem Cells Dev.* **22,** 2268–2277 (2013).

34. Spiegel, S., Milstien, S. & Grant, S. Endogenous modulators and pharmacological inhibitors of histone deacetylases in cancer therapy. *Oncogene* **31,** 537–551 (2012).

35. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

36. Rajan, J. V., Warren, S. E., Miao, E. A. & Aderem, A. Activation of the NLRP3 inflammasome by intracellular poly I:C. *FEBS Lett.* **584,** 4627–4632 (2010).

37. Van den Boorn, J. G., Schlee, M., Coch, C. & Hartmann, G. SiRNA delivery with exosome nanoparticles. *Nat. Biotechnol.* **29,** 325–326 (2011).

38. Awe, J. P. & Byrne, J. A. Identifying candidate oocyte reprogramming factors using cross-species global transcriptional analysis. *Cell. Reprogramming* **15,** 126–133 (2013).

39. Sætrom, P. & Snøve Jr, O. A comparison of siRNA efficacy predictors. *Biochem. Biophys. Res. Commun.* **321,** 247–253 (2004).

40. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).

CHAPTER THREE

Induction of Pluripotency by saRNA

The previous study identified a siRNA targeting the histone methyltransferase SUV39H1 as a potent enhancer of small activating RNA (saRNA) activity. In a repeat of that experiment, induced pluripotent stem cell (iPSC)-like colonies were observed with saRNA transfection. However, these colonies were also seen in the negative control oligonucleotide transfected cells. Subsequent repeats of this experiment failed to reproduce this result in either the control or saRNA transfected cells. In this study, we demonstrate the pluripotency of the transfected cells and demonstrate that they do not express a viral transcript. While the true method of their apparent reprogramming cannot be determined, the unreliability of saRNA-induced pluripotency will limit its application in iPSC generation.

Introduction

Regardless of the methodology used to generate them, putative induced pluripotent stem cell (iPSC) lines must be thoroughly characterized to assess their similarity to embryonic stem cells (ESCs). Briefly, they must maintain a stable karyotype and be expandable indefinitely *in vitro*, express genes and cell surface markers characteristic of ESCs, and be able to differentiate into all three germ layers[1]. This standard is particularly necessary for iPSCs due to the potentially genotoxic nature of factor-based reprogramming. Yamanaka's initial study demonstrating iPSC generation reported roughly 20 retroviral insertions per clone[2], allowing for the possibility of insertional mutagenesis. Many cancers have been shown to express pluripotency factors[3], so it is critical that gene expression analysis be complemented with a full karyotype to ensure the observed pluripotency is not associated with oncogenesis. Use of nonintegrating vectors does not negate the need for such analysis, as chromosomal aberrations can be acquired during reprogramming or after extended culture[4,5]. Finally, the ability of putative iPSCs to differentiate into cells representing each of the three embryonic germ layers—ectoderm, mesoderm, and endoderm—must be demonstrated for them to be considered pluripotent[6].

The above describes the bare minimum required for iPSC characterization. Many more assays are frequently performed to assess the degree of similarity between iPSC lines and standard ESC lines. In mice, where it is ethically appropriate, iPSCs can be tested for germline-competency[7], the highest standard for demonstrating pluripotency. Other common tests include comparison of DNA and histone methylation patterns[2,8,9], silencing of retroviral transgenes[7], and more recently, screening for immunogenicity of iPSC-derived cells[10–12]. In this study, we demonstrate the ability of saRNA to induce pluripotency when used in combination with a siRNA targeting the histone methyltransferase SUV39H1. These iPSCs express genes

characteristic of pluripotent cells and demonstrate an ability to differentiate into all three embryonic germ layers, but a teratoma assay is necessary to confirm this. However, the negative control oligonucleotide transfected cells also generated iPSCs, and these results were unable to be reproduced. Until saRNA can reliably induce pluripotency, their use in iPSC research will be limited.

Results

Induction of pluripotency with saRNA. We previously demonstrated a dramatic increase in downstream NANOG mRNA activation when fibroblasts were pretreated with a SUV39H1 siRNA before KLF4, OCT4, SOX2, and MYC saRNA transfection. When this protocol was repeated, numerous colonies were visible by day 10 after initial saRNA transfection, including in the negative control oligo transfected and positive control STEMCCA virus conditions (Figure 3-1A). To determine if these were cell aggregates or early reprogrammed colonies, immunostaining was performed for the stem cell markers NANOG, TRA-1-60, and TRA-1-81, with each condition showing positive staining (Figure 3-1B, C, D). NANOG mRNA expression was confirmed by RT-qPCR (Figure 3-1E). To rule out lentiviral contamination, RT-PCR for the STEMCCA transcript revealed only viral expression in the STEMCCA samples (Figure 3-1F). On day 20, colonies from each condition were picked and re-seeded on fresh MEF plates, from which 11 putative iPSC lines were established: 3 of negative control, 5 of saRNA, and 3 of STEMCCA lentivirus (Figure 3-1G).

Characterization of putative iPSC lines. These 11 lines were maintained for 10 passages in feeder-dependent culture conditions. All lines stained positive for the stem cell markers SSEA3, SSEA4, TRA-1-60, and TRA-1-81 (Figure 3-2). Most lines displayed unusual morphology after passaging, with colonies having holes in the middle that would gradually fill in over time (Figure 3-2). RT-qPCR for OCT4 and NANOG mRNA confirmed each line overexpressed these transcription factors at least an order of magnitude higher than the parental BJ fibroblasts, but on average an order of magnitude lower than H9 ESCs (Figure 3-3). A representative line with the highest NANOG mRNA expression from each of the negative control, saRNA, and STEMCCA lines was chosen for further characterization (negative 2-1,

saRNA 4-2, and STEMCCA 2-1). After adapting these three lines to feeder-free culture conditions, pluripotency was demonstrated by *in vitro* directed differentiation and immunostaining for Otx2, a marker of ectoderm[13], Brachyury, a marker of mesoderm[14], and Sox17[15], a marker of endoderm (Figure 3-4A, B, C). Karyotype results are pending.

Subsequent repeats of saRNA-induced pluripotency. This reprogramming protocol was repeated two additional times, for a total of four times including the previous study. No further iPSC colonies were generated in either the negative control or saRNA transfected samples.

Discussion

The appearance of iPSC-like colonies during the repeat of the SUV39H1 knockdown experiment was striking, especially considering the negative control samples also had colonies. The first possible explanation for this is contamination. However, these experiments were started in a tissue culture room where no iPSC or ESCs are cultured, ruling out the possibility of cross-contamination from established lines. Further, the very large number of colonies that were present in all wells of several multi-well plates in every condition makes it unlikely that contamination on such a large scale would occur accidentally. The possibility of STEMCCA lentiviral cross-contamination was also ruled out, as only the STEMCCA samples expressed the viral transcript. Unfortunately, in the absence of an obvious viral or plasmid integration, it is very difficult to determine how an iPSC became reprogrammed after the fact. Regardless, we have demonstrated that these are bona-fide iPSC lines. They express the markers characteristic of pluripotency and are able to differentiate into cells expressing markers of each of the three embryonic germ layers, but this must be confirmed by a teratoma assay. While the level of OCT4 and NANOG mRNA expression is an order of magnitude lower than H9 ESCs, it is within the range of overexpression compared to fibroblasts frequently reported in iPSC generation[16–18]. While karyotype results are pending, a normal karyotype would ensure that these are not transformed cells that have acquired a pluripotent phenotype in culture. If the karyotype is determined to be abnormal, and all three lines show the same aneuploidly, it is possible that all of the lines came from the same spontaneously transformed cell in culture. Alternatively, if the control is abnormal and others are normal, then we can conclude that saRNA is likely responsible for this induction of pluripotency. However, it remains that these results are not reproducible.

While single-factor reprogramming has been reported[19,20], it has required cells that endogenously express all of the Yamanaka factors at a high level except OCT4. The observation here that cells transfected with only a SUV39H1 siRNA and a negative control oligonucleotide can be reprogrammed is controversial and not supported by the literature. Given the 60-fold NANOG mRNA activation seen when SUV39H1 siRNA is combined with 4-factor saRNA, generation of iPSCs is not entirely unexpected. However, given that neither the generation of iPSCs by SUV39H1 alone nor in combination with saRNA was able to be reproduced, it is ultimately unclear whether saRNA truly had a role in the reprogramming seen here. The failure of saRNA to reliably induce pluripotency may lie in its relative lack of off-target effects that have been shown to be beneficial in other reprogramming methodologies[21]. Even with non-integrating protocols such as modified mRNA and episomal plasmids, there is an associated innate immune response to exogenous RNA or DNA[21–23]. The gene expression changes induced by this innate immune response provide the Yamanaka factors with the necessary epigenetic changes to efficiently induce pluripotency. This leaves the role of saRNA in future reprogramming studies unclear. There is still a possibility for saRNA to complement current reprogramming methodologies by substituting for the oncogene c-Myc in integrating protocols. While viral reprogramming can be optimized or selected for to transduce nearly 100% of a cell population, a much lower percentage of cells ultimately become reprogrammed. The specific cooperation of chromatin modifiers has been shown to dramatically increase this percentage, and indeed they can improve saRNA activity as well. There may yet be another factor identified that can further augment saRNA activity, but the increasing complexity and relative difficulty of this methodology may prevent it from being adopted over standard protocols. Any future iPSCs reproducibly generated with saRNA must be shown to be more similar to ESCs in their

transcriptional epigenetic state before this method can deliver on its potential for a truly safe and clinically relevant source for iPSCs.

Materials and Methods

Cell culture

Neonatal human skin-derived fibroblasts ("BJ", ATCC, Manassas, Virginia) were cultured in fibroblast medium containing DMEM supplemented with 10% fetal bovine serum and 1X penicillin/streptomycin. For reprogramming and pluripotent cell culture, human embryonic stem cell medium contained DMEM/F12 supplemented with 20% Knockout Serum Replacement (Invitrogen, Carlsbad, California), 1X non-essential amino acids, penicillin/streptomycin, β-mercaptoethanol (Millipore, Billerica, Massachusetts), and 4ng/mL bFGF (Invitrogen). Mitomycin C-inactivated CF-1 mouse embryonic fibroblasts (Globalstem, Rockville, Maryland) were used as a feeder layer. Colonies were transferred to plates coated with reduced growth factor Matrigel (BD Biosciences, San Jose, California) and grown in MEF-conditioned hESC medium for karyotyping and *in vitro* differentiation. Karyotyping was done by the UCLA Cytogenetics Laboratory. *In vitro* differentiation was done using the Human Pluripotent Stem Cell Functional Identification Kit (R&D Systems, Minneapolis, Minnesota) following the manufacturer's protocol. RNA oligonucleotides were transfected using INTERFERin reagent (Polyplus-transfection, New York, New York) following the manufacturer's protocol with a concentration of 1nM per oligo. Transfections were performed every other day unless otherwise indicated. The BLOCK-iT Alexa Fluor Red Fluorescent Control (Invitrogen), which has no homology to any known gene, was used as a negative control.

Plasmids and Lentivirus Vector Production

The stem cell cassette (STEMCCA)[24] lentiviral reprogramming vector was used to make VSV-G pseudotyped second-generation lentivirus preparations produced using standard protocols;

briefly, packaging plasmids pMD2.G, psPAX2, and transfer vector were cotransfected into 293T cells with jetPRIME reagent (Polyplus-transfection), and 48 hours later virus-containing supernatant was collected and filtered. Vector titers were determined by p24 ELISA, performed by the UCLA Virology Core, and immunocytochemistry for Oct4 protein expression (sc-5279, Santa Cruz Biotechnology, Dallas, Texas).

iPSC Generation

Fibroblasts were seeded at a density of $1x10^4$ cells/cm^2 and were transfected with 1nM siRNA targeting SUV39H1 (SASI_Hs02_00335192, Sigma-Aldrich), and then an additional time three days later. Cells were then transfected with saRNA or negative control oligo starting the following day and every other day for the duration of the experiment. STEMCCA lentivirus was used for a positive control and these cells were transduced the day after the final siRNA transfection in medium containing polybrene at 4µg/mL. On the 5^{th} day following the initial saRNA transfection or transduction, cells were replated at a density of $1.7x10^3$ cells/cm^2 on a MEF feeder layer. On the day following re-seeding, medium was changed to hESC medium with transfections continuing every other day.

Quantitative RT-PCR

Total RNA was isolated using the Aurum Total RNA Mini Kit (Bio-Rad, Hercules, California), which includes a DNase I digestion step to remove genomic DNA. RNA was reverse-transcribed to cDNA using the High Capacity cDNA Kit (Applied Biosystems). Quantitative real-time PCR was performed using SsoFast Probes Supermix (Bio-Rad) on a MyiQ2 thermal cycler (Bio-Rad) according to the manufacturer's standard protocols. The Taqman primer sets used were as

follows: KLF4, Hs00358836_m1; POU5F1 (OCT4A isoform), Hs01895061_u1; SOX2, Hs00602736_s1; NANOG, Hs02387400_g1; MYC, Hs00153408_m1; ACTB, Hs00357333_g1 (Applied Biosystems). Experiments were performed in triplicate wells with at least 3 replicate reactions per PCR. Expression of β-actin mRNA was used as an internal control and samples were normalized to the scrambled sequence control oligonucleotide-treated samples.

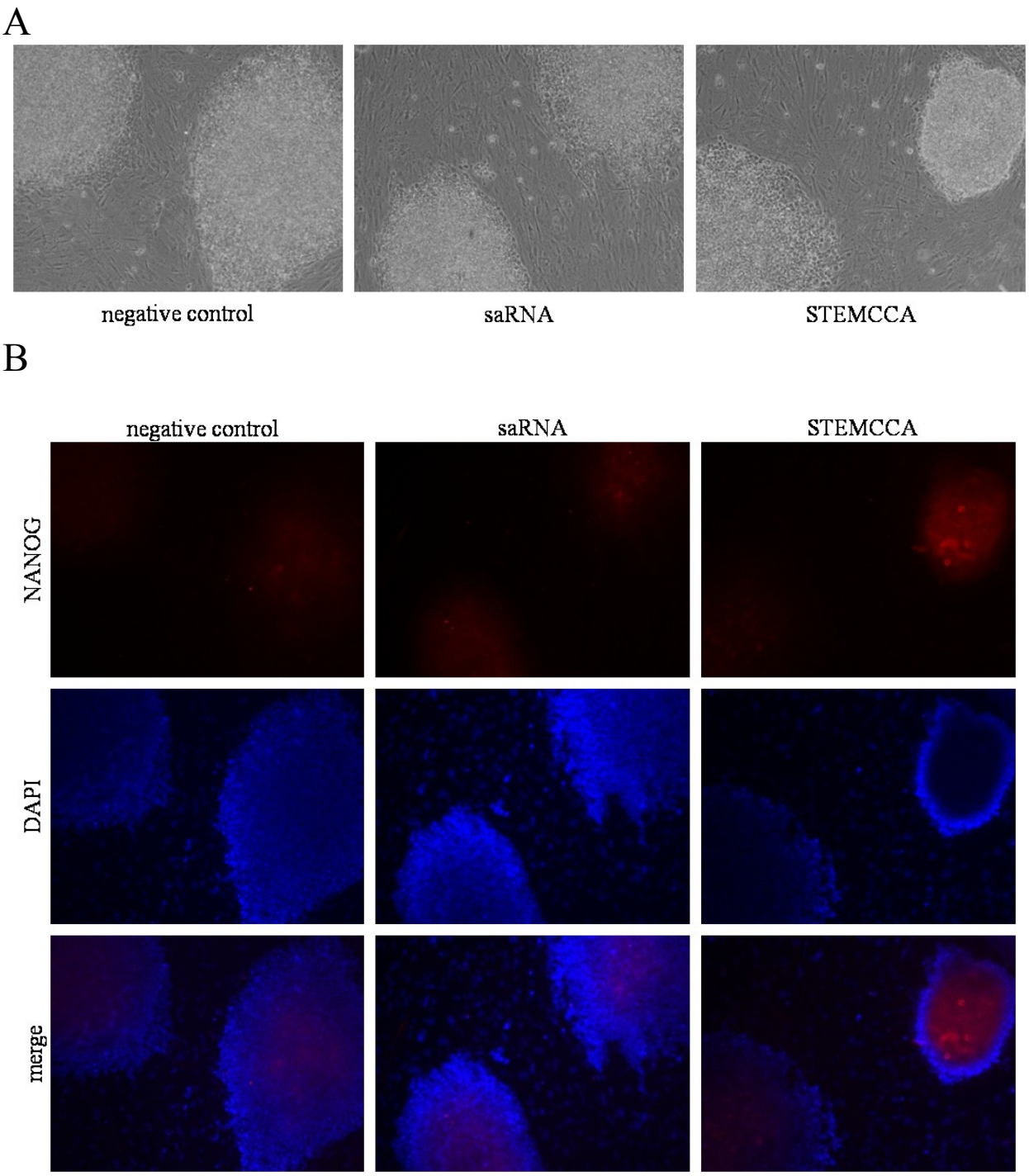


Figure 3-1. Characterization of cells during reprogramming. (A) Phase contrast microscopy showing iPSC-like colonies in every test condition. Images shown are from day 16 after initial saRNA transfection or lentiviral transduction. **(B)** Immunostaining showing colonies are positive for NANOG expression. Images shown are from day 16 after initial saRNA transfection or lentiviral transduction.

C

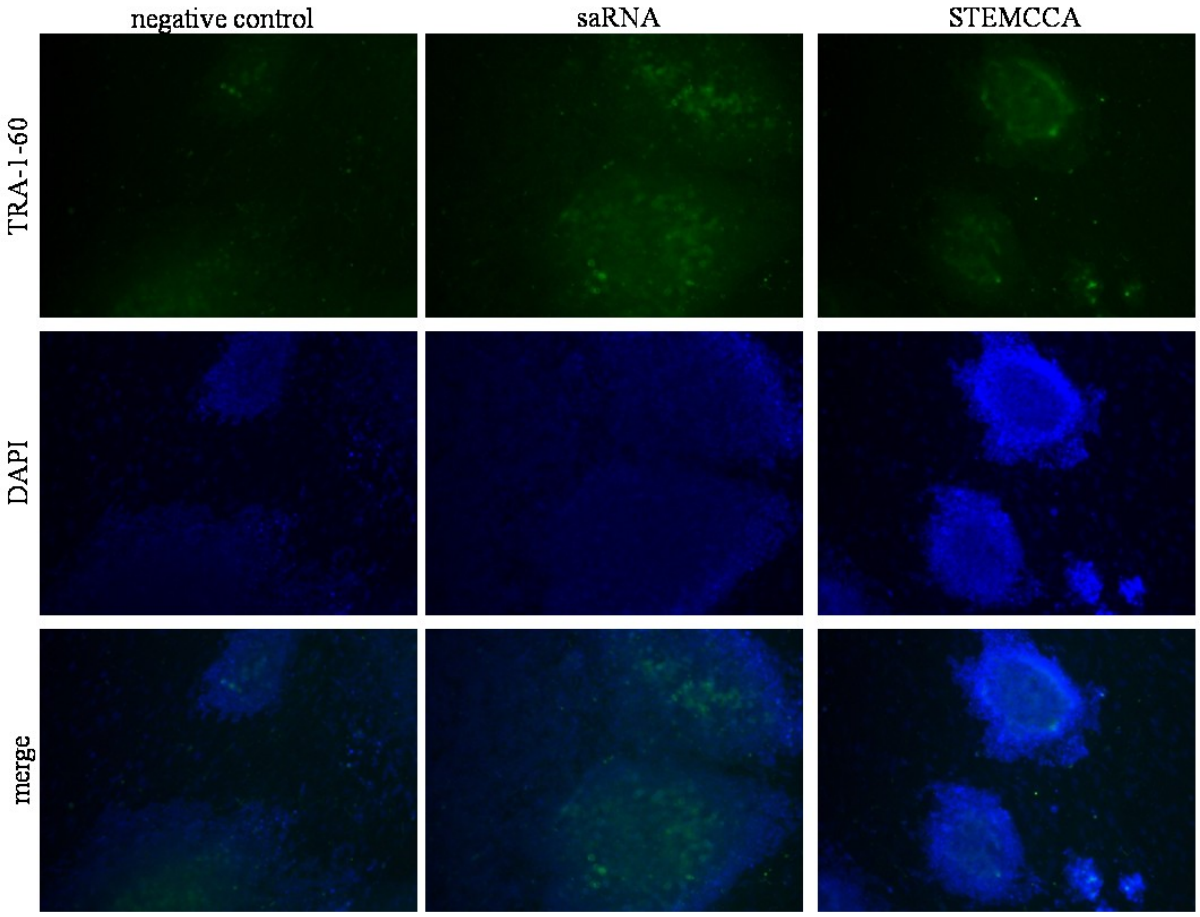


Figure 3-1, continued. (C) Immunostaining showing colonies are positive for TRA-1-60 expression. Images shown are from day 16 after initial saRNA transfection or lentiviral transduction.

D

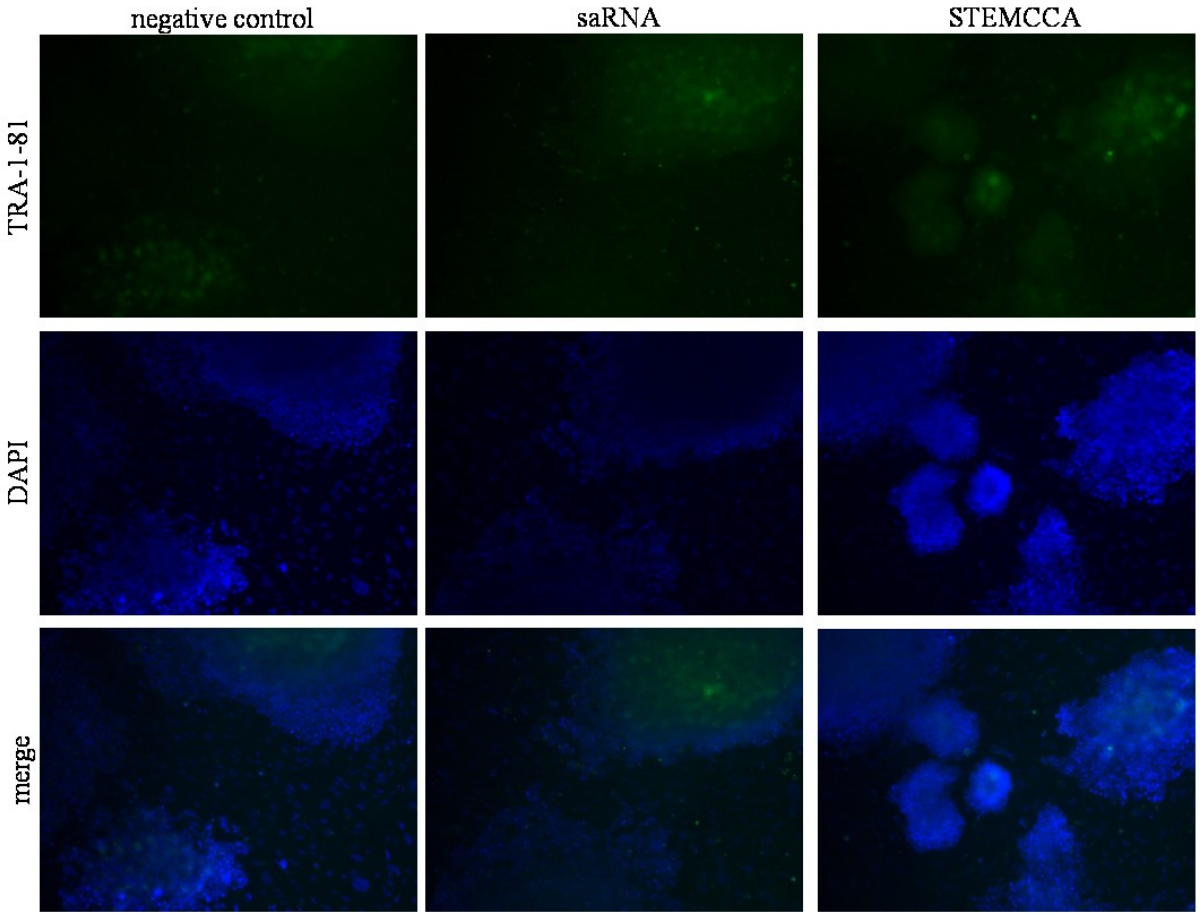
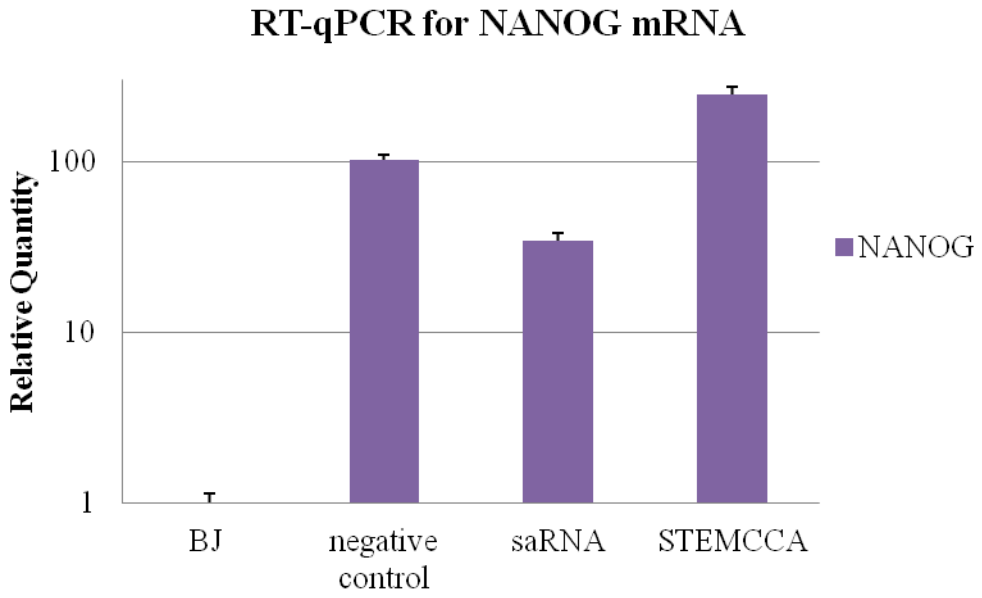


Figure 3-1, continued. (D) Immunostaining showing colonies are positive for TRA-1-81 expression. Images shown are from day 16 after initial saRNA transfection or lentiviral transduction.

E



F

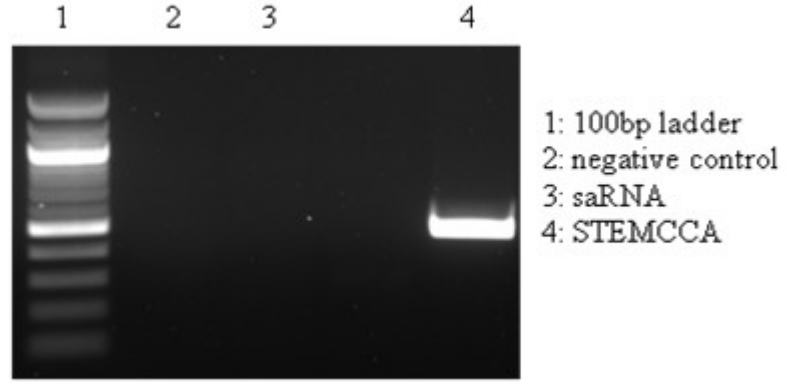


Figure 3-1, continued. (E) RT-qPCR for NANOG mRNA on day12 of the reprogramming experiment. Values are shown on a logarithmic scale relative to normal BJ fibroblasts. **(F)** RT-PCR using primers specific for the STEMCCA transcript showing viral expression only in the STEMCCA cells.

G

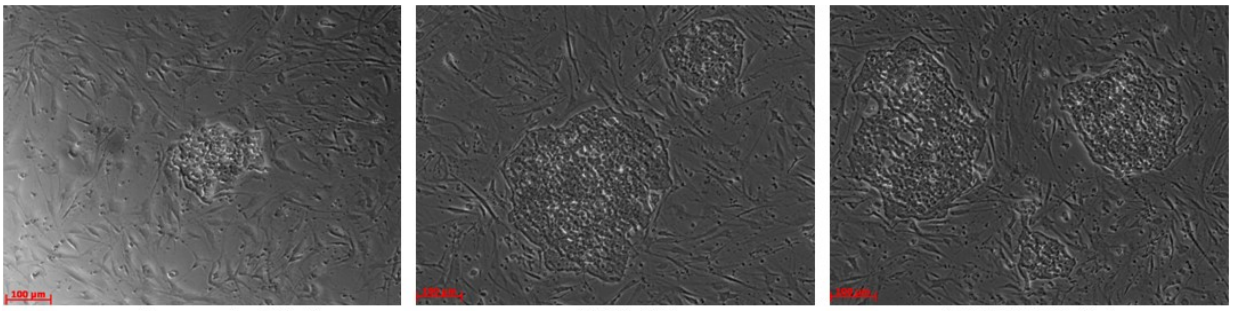


Figure 3-1, continued. (G) Representative phase contrast micrographs showing derived putative iPSCs at passage 2.

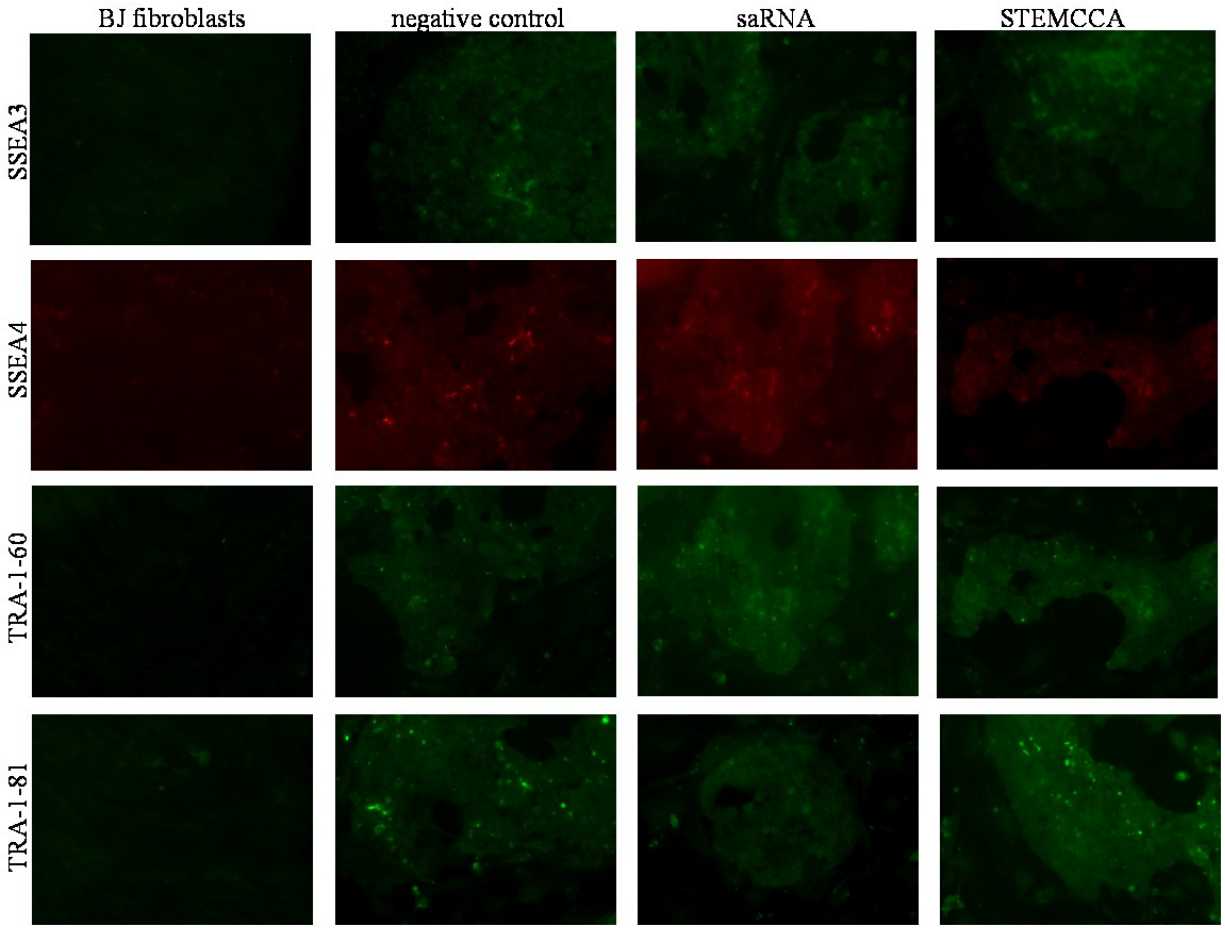


Figure 3-2. Cell surface marker characterization of iPSC lines. Displayed are representative images showing positive staining for the stem cell markers SSEA3, SSEA4, TRA-1-60, and TRA-1-81 at passage 7. Note the unusual gaps where no cells are present in the colonies.

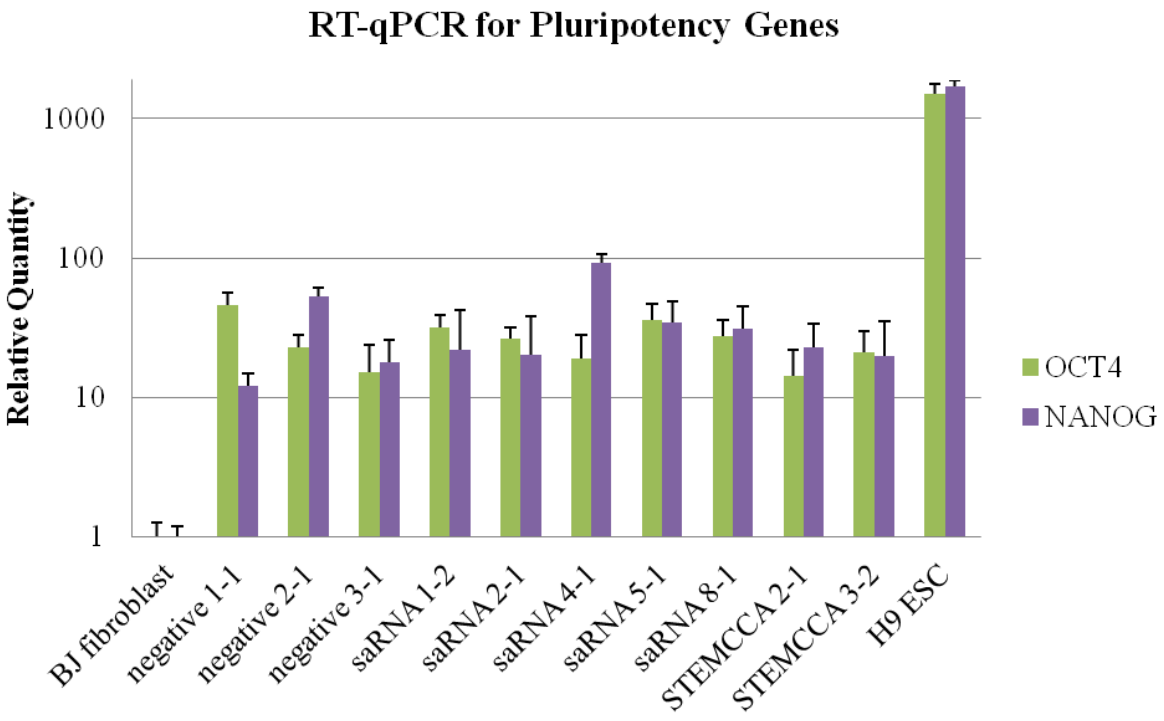


Figure 3-3. Gene expression characterization of iPSC lines. RT-qPCR for OCT4 and NANOG mRNA in the individual iPSC lines. Values are shown on a logarithmic scale relative to normal BJ fibroblasts.

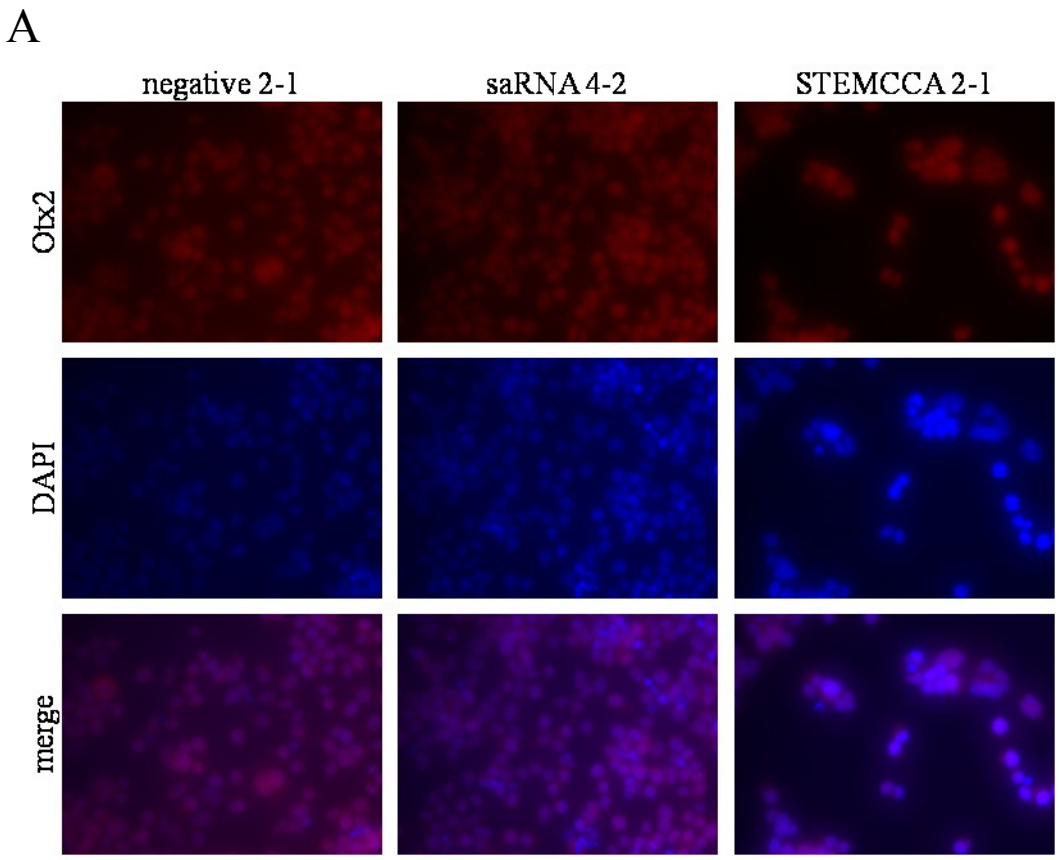


Figure 3-4. Trilineage differentiation of iPSCs. (A) Immunostaining for Otx2 after directed ectodermal differentiation.

B

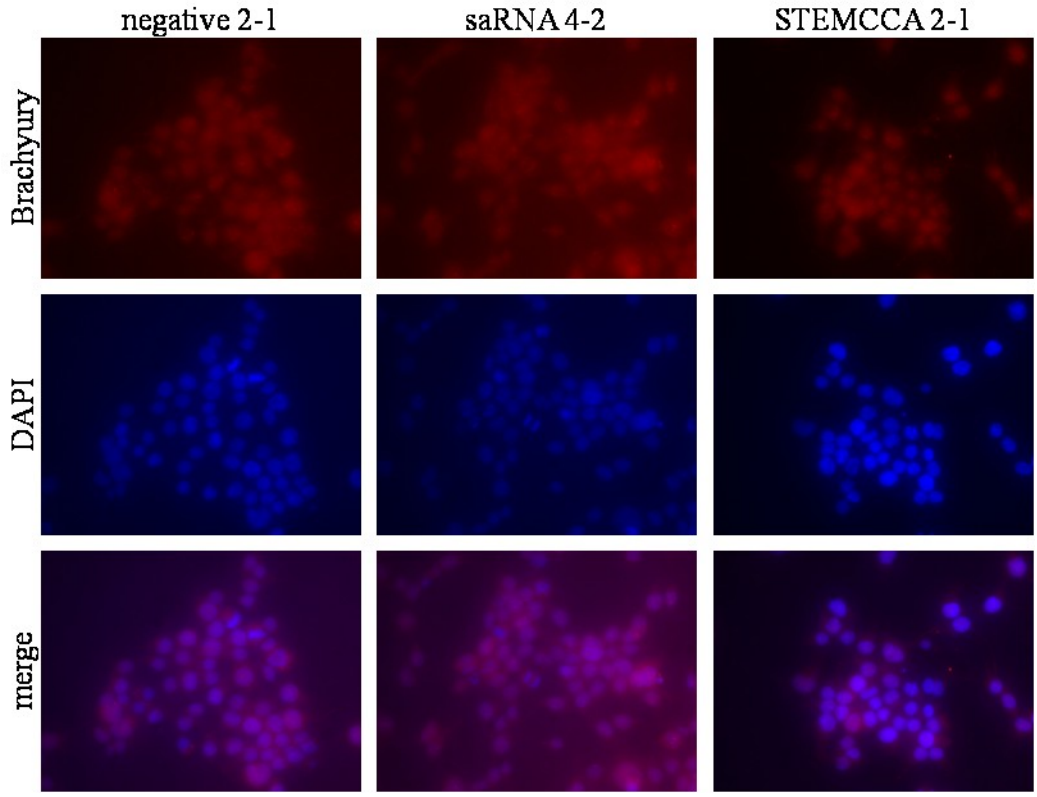


Figure 3-4, continued. (B) Immunostaining for Brachyury after directed mesodermal differentiation.

C

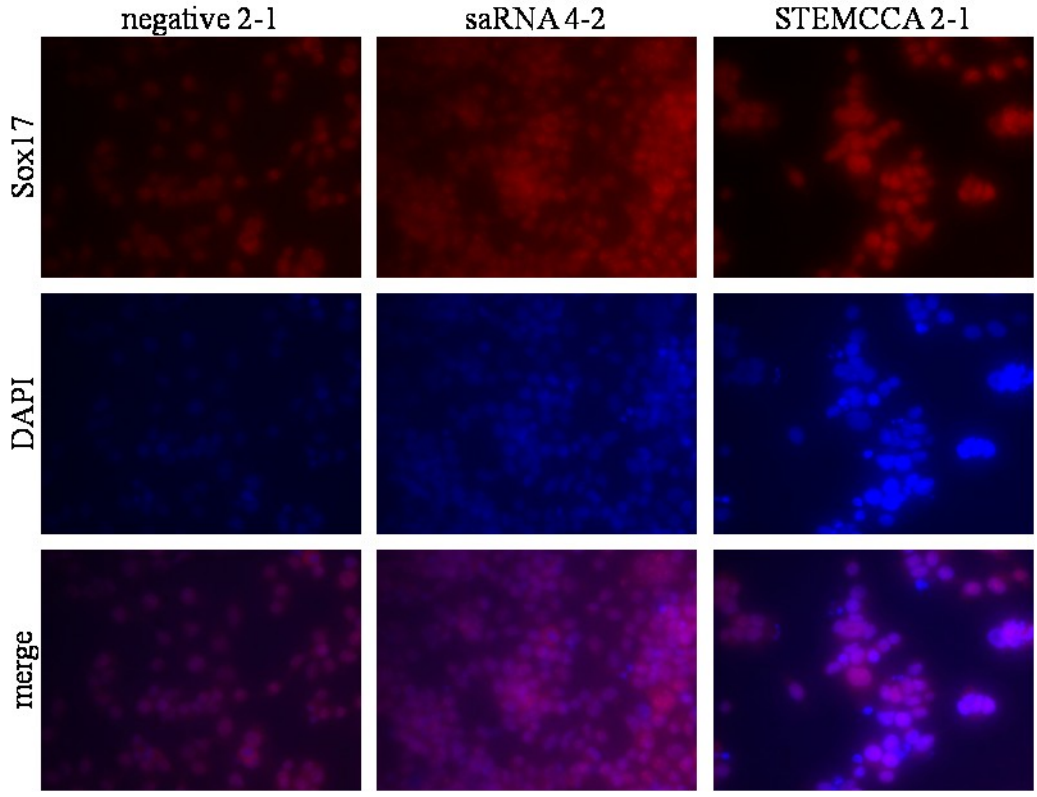


Figure 3-4, continued. (C) Immunostaining for Sox17 after directed endodermal differentiation.

References

1. Hoffman, L. M. & Carpenter, M. K. Characterization and culture of human embryonic stem cells. *Nat. Biotechnol.* **23,** 699–708 (2005).

2. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).

3. Schoenhals, M. *et al.* Embryonic stem cell markers expression in cancers. *Biochem. Biophys. Res. Commun.* **383,** 157–162 (2009).

4. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7,** 521–531 (2010).

5. Baker, D. E. C. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat. Biotechnol.* **25,** 207–215 (2007).

6. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282,** 1145–1147 (1998).

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

8. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41,** 1350–1353 (2009).

9. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).

10. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. Immunogenicity of induced pluripotent stem cells. *Nature* **474,** 212–215 (2011).

11. Araki, R. *et al.* Negligible immunogenicity of terminally differentiated cells derived from induced pluripotent or embryonic stem cells. *Nature* **494,** 100–104 (2013).

12. Guha, P., Morgan, J. W., Mostoslavsky, G., Rodrigues, N. P. & Boyd, A. S. Lack of immune response to differentiated cells derived from syngeneic induced pluripotent stem cells. *Cell Stem Cell* **12,** 407–412 (2013).

13. Larsen, K. B., Lutterodt, M. C., Møllgård, K. & Møller, M. Expression of the homeobox genes OTX2 and OTX1 in the early developing human brain. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **58,** 669–678 (2010).

14. Showell, C., Binder, O. & Conlon, F. L. T-box genes in early embryogenesis. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **229,** 201–218 (2004).

15. Niakan, K. K. *et al.* Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev.* **24,** 312–326 (2010).

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

17. Esteban, M. A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6,** 71–79 (2010).

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

19. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136,** 411–419 (2009).

20. Tsai, S.-Y. *et al.* Single transcription factor reprogramming of hair follicle dermal papilla cells to induced pluripotent stem cells. *Stem Cells Dayt. Ohio* **29,** 964–971 (2011).

21. Lee, J. *et al.* Activation of innate immunity is required for efficient nuclear reprogramming. *Cell* **151,** 547–558 (2012).

22. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

23. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448,** 501–505 (2007).

24. Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells Dayt. Ohio* **27,** 543–549 (2009).

Conclusions and Future Directions

Here we have demonstrated the application of an algorithm that can consistently identify target sequences for saRNA-induced activation of endogenous gene expression. In a comprehensive study of KLF4 and MYC saRNAs, we showed that saRNA not only upregulates target gene expression, but this increase in mRNA is due to transcriptional activation. These 19nt RNA duplexes are small enough to not induce an innate immune response[1], and we show that no interferon response to dsRNA is seen with saRNA transfection. This was the first published study to show global gene expression changes after saRNA transfection[2]. These gene expression changes were similar to overexpression by lentiviral transduction, and many of the differentially expressed pathways were related to viral transduction. After an exhaustive search for cryptic transcripts arising from the promoter regions of KLF4 and MYC, we were unable to identify any antisense RNAs involved in the regulation of these genes. This does not completely discount the possibility that other antisense RNAs may be regulating KLF4 or MYC in *trans*, as has been reported with OCT4[3]. However, no novel transcripts arising outside of the promoter region at all were identified using saRNA-specific primers for RACE, indicating that antisense RNA regulation of these genes is unlikely. Future mechanistic work should seek to confirm the presence of Argonaute-2 at the promoters of saRNA target genes and investigate whether they are physically binding to RNA or DNA.

Many factors that potentially regulate saRNA activity also remain unclear. Each of the genes targeted in this work has a basal level of gene expression, although for OCT4 and SOX2 at an extremely low level in fibroblasts[4]. It would be helpful to determine if some level of gene expression is necessary for saRNA activity. If there is a correlation between basal gene expression level and fold activation across multiple cell types, it could serve as a predictor for

potential saRNA activity in untested cells. It is also possible that cells must be dividing for saRNA to work. For a hypothetical gene that is not expressed at all in a certain cell type, chromatin access may be so restricted that the Ago2-saRNA complex may only be able to access the promoter during DNA replication. Investigating these possibilities will greatly improve the understanding of what factors influence saRNA activity. As most published reports on RNAa have studied saRNA activity in cancer cell lines[5], it is critical that these experiments be validated in normal cells with properly functioning gene regulatory networks.

By targeting the Yamanaka factors, we sought to develop a novel approach to iPSC generation that would require no genomic alteration or introduction of exogenous transgenes. As saRNA uses endogenous cellular processes, we hypothesized that this would provide a more "natural" way to induce pluripotency for more ideal and safe clinical applications. However, the advances in the iPSC field have shown that the cells generated by the traditional 4 factor approach are epigenetically flawed and ultimately inferior to cells reprogrammed by SCNT or ESCs in their gene expression and differentiation potential[6–12]. Our results combining a histone methyltransferase inhibitor with saRNA agree with the growing consensus that altering the epigenetic state of the cell is critical for improving pluripotency factor gene expression and reprogramming efficiency[13–16]. However, it was only inhibition of SUV39H1, not DOT1L, that provided this benefit with saRNA. If this observation is generalizable to saRNA activation of other genes and in other cell types, this could be an important method to enhance the utility of saRNA. The selectivity of this effect may also provide a key to further insights regarding the mechanisms involved in saRNA-mediated transcriptional activation. Whether or not these chromatin-modifying approaches improve iPSC efficacy in regard to their similarity to ESCs remains to be proven. Ultimately, we believe that the saRNA-alone approach here has proven

unreliable due to the lack of off target effects and beneficial gene expression changes associated with transgene delivery in other reprogramming protocols[17]. However, there may still be utility in substituting exogenous transgenes for saRNA, such as the oncogene c-Myc. Determining which specific activated or repressed genes cooperate with the Yamanaka factors during successful reprogramming will guide future optimizations of a saRNA approach to iPSC generation. Given that the proteins present in the oocyte are more effective at establishing an ESC-like epigenetic state[8], it is likely that a focus on oocyte-specific factors[18,19] will provide the greatest promise for a role for saRNA in iPSC generation.

Fortunately, there are many applications for saRNA outside of nuclear reprogramming, and this work provides a proof-of-concept for the utility of our saRNA design algorithm. Use of KLF4 saRNA has already been shown to inhibit prostate cancer growth, consistent with KLF4's role in inducing cell cycle arrest[20]. Our collaborators have also demonstrated anti-cancer activity of C/EBPα saRNA in an *in vivo* model of hepatocellular carcinoma[21]. Activating endogenous cellular pathways with saRNA may also provide a more specific method for directed differentiation of stem cells or transdifferentiation, as demonstrated by the production of insulin-secreting cells from human CD34+ progenitors with saRNA transfection[22]. The use of saRNA may provide a more specific method to induce differentiation in pluripotent cells, potentially replacing expensive recombinant proteins and culture additives. While unreliable without further optimization in its present state for iPSC generation, the potential for saRNA technology is only limited by our knowledge of gene function in development and disease. With siRNA, biomedical research was given an efficient method to knockdown the expression of any gene. With saRNA, we believe we will now be able to activate any gene, and we hope its application will be just as transformative.

References

1. Reynolds, A. *et al.* Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* **12,** 988–93 (2006).

2. Voutila, J. *et al.* Gene Expression Profile Changes After Short-activating RNA-mediated Induction of Endogenous Pluripotency Factors in Human Mesenchymal Stem Cells. *Mol. Ther. Nucleic Acids* **1,** e35 (2012).

3. Hawkins, P. & Morris, K. V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* **1,** 1–11 (2010).

4. Page, R. L. *et al.* Induction of Stem Cell Gene Expression in Adult Human Fibroblasts without Transgenes. *Cloning Stem Cells* **11,** 417–426 (2009).

5. Portnoy, V., Huang, V., Place, R. F. & Li, L.-C. Small RNA and transcriptional upregulation. *Wiley Interdiscip. Rev. RNA* **2,** 748–760 (2011).

6. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).

7. Wang, T. *et al.* Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. *Nat. Cell Biol.* **15,** 700–711 (2013).

8. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290 (2010).

9. Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat. Biotechnol.* **29,** 1117–1119 (2011).

10. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. Immunogenicity of induced pluripotent stem cells. *Nature* **474,** 212–215 (2011).

11. Araki, R. *et al.* Negligible immunogenicity of terminally differentiated cells derived from induced pluripotent or embryonic stem cells. *Nature* **494,** 100–104 (2013).

12. Guha, P., Morgan, J. W., Mostoslavsky, G., Rodrigues, N. P. & Boyd, A. S. Lack of immune response to differentiated cells derived from syngeneic induced pluripotent stem cells. *Cell Stem Cell* **12,** 407–412 (2013).

13. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).

14. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151,** 994–1004 (2012).

15. Luo, M. *et al.* NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells Dayt. Ohio* **31,** 1278–1286 (2013).

16. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).

17. Lee, J. *et al.* Activation of innate immunity is required for efficient nuclear reprogramming. *Cell* **151,** 547–558 (2012).

18. Awe, J. P. & Byrne, J. A. Identifying candidate oocyte reprogramming factors using cross-species global transcriptional analysis. *Cell. Reprogramming* **15,** 126–133 (2013).

19. Shinagawa, T. *et al.* Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. *Cell Stem Cell* **14,** 217–227 (2014).

20. Wang, J. *et al.* Prognostic Value and Function of KLF4 in Prostate Cancer: RNAa and Vector-Mediated Overexpression Identify KLF4 as an Inhibitor of Tumor Cell Growth and Migration. *Cancer Res.* **70,** 10182–10191 (2010).

21. Reebye, V. *et al.* Novel RNA oligonucleotide improves liver function and inhibits liver carcinogenesis in vivo. *Hepatol. Baltim. Md* **59,** 216–227 (2014).

22. Reebye, V. *et al.* A Short-activating RNA Oligonucleotide Targeting the Islet β-cell Transcriptional Factor MafA in CD34(+) Cells. *Mol. Ther. Nucleic Acids* **2,** e97 (2013).