UC San Diego

UC San Diego Previously Published Works

Title

Genome-wide Functional Analysis Reveals Factors Needed at the Transition Steps of Induced Reprogramming

Permalink

https://escholarship.org/uc/item/49k4c0fw

Journal

Cell Reports, 8(2)

ISSN

2639-1856

Authors

Yang, Chao-Shun

Chang, Kung-Yen

Rana, Tariq M

Publication Date

2014-07-01

DOI

10.1016/j.celrep.2014.07.002

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at https://creativecommons.org/licenses/by-nc-nd/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

NIH Public Access
Author Manuscript
Cell Rep. Author manuscript; available in PMC 2015 July 24.

Published in final edited form as:
Cell Rep. 2014 July 24; 8(2): 327–337. doi:10.1016/j.celrep.2014.07.002.

Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming

Chao-Shun Yang[1,2,3], **Kung-Yen Chang**[1,3], and **Tariq M. Rana**[1,2,3,*]

[1]Program for RNA Biology, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

[2]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

[3]Department of Pediatrics, University of California San Diego School of Medicine, 9500 Gilman Drive MC 0762, La Jolla, California 92093, USA

SUMMARY

While transcriptome analysis can uncover the molecular changes that occur during induced reprogramming, the functional requirements for a given factor during step-wise cell-fate transitions are left unclear. Here, we useda genome-wide RNAi screen and performed integrated transcriptome analysis to identify key genes and cellular events required at the transition steps in reprogramming. Genes associated with cell signaling pathways (e.g., *Itpr1, Itpr2, Pdia3*) constitute the major regulatory networks before cells acquire pluripotency. Activation of a specific gene set (e.g., *Utf1, Tdgf1*) is important for mature iPSC formation. Strikingly, a major proportion of RNAi targets (~53% to 70%) includes genes whose expression levels are unchanged during reprogramming. Among these nondifferentially expressedgenes, *Dmbx1, Hnf4g, Nobox* and *Asb4*, are important, while *Nfe2, Cdkn2aip, Msx3, Dbx1, Lzts1, Gtf2i*, and *Ankrd22* are roadblocks, to reprogramming. Together, our results provide a wealth of information about gene functions required at transition steps during reprogramming.

INTRODUCTION

Somatic reprogramming to pluripotent status can be achieved by introducing a limited number of transcription factors, including Oct4, Sox2, Klf4, c-Myc (OSKM), Nanog, and

© 2014 The Authors. Published by Elsevier Inc. All rights reserved.

* Correspondence: trana@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

AUTHOR CONTRIBUTIONS

C-SY: concept and design, data collection, assembly, analysis and interpretation, and manuscript writing; K-YC: data analysis; TMR: concept and design, data analysis and interpretation, manuscript writing, and financial support. All authors approved the final version of this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Lin28 (Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007). Those induced pluripotent stem cells (iPSCs) highly resemble embryonic stem cells (ESCs) and hold promise to customized regenerative medicine (Grskovic et al., 2011; Jopling et al., 2011; Robinton and Daley, 2012; Tiscornia et al., 2011; Wu and Hochedlinger, 2011).

One of the primary obstacles to the successful application of iPSCs for medical purposes is their low reprogramming efficiency. Significant effort has been devoted to enhancing induced reprogramming efficiency, including approaches focusing on the use of mRNA (Warren et al., 2010); small molecules (Ichida et al., 2009; Li and Rana, 2012; Maherali and Hochedlinger, 2009; Nichols et al., 2009; Silva et al., 2008; Yang et al., 2011b; Ying et al., 2008; Zhu et al., 2011); and miRNAs(Choi et al., 2011; Judson et al., 2009; Kim et al., 2011; Li and He, 2012; Li et al., 2011; Liao et al., 2011; Lipchina et al., 2011; Melton et al., 2010; Pfaff et al., 2011; Subramanyam et al., 2011; Yang and Rana, 2013; Yang et al., 2011a). However, detailed functional insight into the molecular basis of reprogramming is still lacking.

It has been shown that few markers, including Thy1, alkaline phosphatase (AP), and SSEA1, can beused to identify transformed cells through the process of induced reprogramming, while embryonic stem cell-specific genes (*Nanog, Oct4, Tert*) are activated at later stages (Brambrink et al., 2008; Stadtfeld et al., 2008). More recent research further suggests that induced reprogramming is a stepwise event, comprising initial, mature, and stabilization stages (Samavarchi-Tehrani et al., 2010). Several key cellular events have been observed during reprogramming, such as mesenchymal-to- epithelial transition (Li et al., 2010; Samavarchi-Tehrani et al., 2010) and cell-cycle modulation (Banito et al., 2009; Hong et al., 2009; Kawamura et al., 2009; Li et al., 2009; Marion et al., 2009; Utikal et al., 2009). Furthermore, the epigenome is reset upon induced reprogramming (Koche et al., 2011; Maherali et al., 2007), and epigenetic regulators play important roles in the reprogramming process (Onder et al., 2012). The cooperation of OSKM has also been considered as a critical factor to efficient reprogramming (Carey et al., 2011; Soufi et al., 2012; Sridharan et al., 2009). Many ESC-specific genes (e.g., *Esrrb*, *Sall4*, *Nanog*) are shown to be markers for defining reprogramming stages (Brambrink et al., 2008; Stadtfeld et al., 2008). However, functional molecular networks required for cell-fate transitions are not clear during the reprogramming process.

Here, by isolating pure populations of cells during various stages of reprogramming and combining with genome-wide RNAi screen and transcriptome analysis, we were able to discover key genes and cellular events involved in the transitions associated with the reprogramming process. Moreover, we functionally identified the critical genes required to modulate the reprogramming process. We further validated a series of genes that either block or enhance the reprogramming process. We found that non-differentially expressed genes play important roles modulating cell-fate transitions during reprogramming.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

RESULTS

Experimental Strategy for Genome-Wide RNAi Screen in Induced Reprogramming

To elucidate the molecular requirementsof induced reprogramming, we conducted loss-of-function assaysduring thereprogramming process.We used a genome-wide RNAi screen and transcriptome analysis upon induced reprogramming to functionally validate their roles in a step-wise manner (Figure 1A). This method allowed us to identify the cell-fate determinants in reprogramming without making any assumptions about function based on gene expression (Figures 1A and S1A-S1F).

First, we established a set of markers to isolate desired cell populations from a heterogeneous pool of transformed/reprogrammed cells by OSKM reprogramming factors. Thy1 is highly expressed in mouse embryonic fibroblasts (MEFs) and subsequently diminishes during the progression of reprogramming (~day 3 to 5 post induced reprogramming), while SSEA1 is absent in MEFs, but gradually increases at ~day 7 upon induced reprogramming (Brambrink et al., 2008; Stadtfeld et al., 2008). Therefore, Thy1 can serve as an early-stage marker and SSEA1 can serve as a middle- to late-stage marker for assessing reprogramming progress. In addition, it has been shown that retroviral sequences are repressed in ES cells (Macfarlan et al., 2011; Wolf and Goff, 2007); thus we used the *DsRed* gene driven by retroviral LTRs (pMX-*DsRed*) as a marker to differentiate incomplete reprogrammed cells from mature ones. We used these three markers todefinefour differentcell-fate stages in reprogramming:Thy1^{+}/SSEA1^{-} for the initial stage, Thy1^{-}/SSEA1^{-} for the transition stage, SSEA1^{+}/DsRed^{+} for the pre-determined (early-reprogrammed) stage, and SSEA1^{+}/DsRed^{-} for the mature reprogrammed stage (Figure 1A). As starting material, we isolated high-purity (~98%) MEFsexpressing Thy1 and DsRed(Figure S1B and S1E) by fluorescence-activated cell sorting (FACS).Reprogramming was initiatedby transducing these cells with retroviruses expressing OSKM plus lentiviruses containing a whole-genome shRNA library (Figures 1A, S1C, and S1D).

We sorted cells 14 days later (Figure S1C), when reprogramming is reportedlycomplete in MEFs and the transcriptome relatively defined(Hanna et al., 2009; Yamanaka, 2009). Four high-purity cell populations (95–99 % purity) were isolated (Figure S1F) representingstages defined above (Figure 1A).Surprisingly, most (>80%) cells were at the transition (Thy1^{-}/SSEA1^{-}) stage, while only1%–2.5% reached SSEA1^{+} stages (Figure S1G and File S1),suggesting that re-establishing pluripotency networks is the rate-limiting step in reprogramming.Four sorted cell populations are confirmed to properly represent normal reprogramming process by examining embryonic stem cell (ESC)-specific regulators (e.g.,*Esrrb*, *Nanog*, *Lin28a*, and *Sall4*) and mesenchymal-to-epithelial transition regulators (e.g., *Cdh1*, *Ocln*, *Krt8*, *Snai1*, *Zeb1/2*, *Ncam1*) (Figure S1H).

To define thetranscriptomein sorted populations, we used k-means clustering to profile gene expression patternsand identified five groups (I to V)of mRNAs (Figure 1B and File S1). We defined groups I, II, and IIIas“differentially-expressed” genes and groups IV and V asnon-differentially expressed or unchanged genes during the reprogramming process. Gene ontology (GO) analysis showed thatgenes associated with embryonic development, cell cycle, and cell deathwere significantly over-represented in group I to III, and that genes

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

associated with cellular function and maintenance, molecular transport, and metabolism are significantly enriched in group IV and V(Figures S1I, J, and File S1). As expected, this finding demonstrates that differential-expressed genes (group I to III) are highly related to ES function and that non-differentially expressedgenes (group IV and V) are related to basal cellular functions.

Identifying key transcriptomehallmarks in each cell-fate transition during reprogramming

It remains poorly understood which molecular hurdles are critical to overcome for cells to make a transition from initial to mature stages of reprogramming. To address this, we examined transcriptome differences in each cell-fate transition. A majority of the transcriptome changes occurred at the MEF-to-Thy1^{+}/SSEA1^{-} (1373 genes) and Thy1^{+}/SSEA1^{-}-to-Thy1^{-}/SSEA1^{-} (1387 genes) transitions (Figures S2A and S2B), while fewer occurred in later Thy1^{-}/SSEA1^{-}-to-SSEA1^{+}/DsRed^{+} (312 genes) and SSEA1^{+}/DsRed^{+}-to-SSEA1^{+}/DsRed^{-} (283 genes) transitions. These results showed that a massive transcriptome reconstruction primarily occurs in the early stages before cells obtain an SSEA1^{+} marker, which pushescommitted cell populations toward pluripotency (Figure S2F). Our data suggest that the first two transitions may be the cell-fate-reorganizing phases, comprising the respond-to-reprogramming stress step and the de-constructing-of-somatic-networks step. Following these steps, the next two transitions are cell-fate-committing phases where ES cell-specific regulatory networks are acquired to attain pluripotent status, in the context ofdominant OSKM expression (Figures S2D, E, F).

Stage-specific genes are identified in each transition (e.g., *Lyz*, Lyzs, Mrc1, Slc38a5, Laptm5, Ms4a6d, Nanog, Sall4, Esrrb, Dppa4, Dppa5a, Dnmt3b, Dnmt3l) (Figures S2B). Cellular functions critical for transition steps of reprogramming were identified by GO analysis (Figure S2C), showing that modulating somatic-cell functions are required in the initial stages and that genes associated with ESC pluripotency are highly regulated in the subsequentstages.

GO analysis of the differentially-expressed genes at each transition suggests that a number of canonical pathways including hepatic fibrosis/stellate cell adhesion, matrix metalloproteases, and adhesion and diapedesis, are important to modulate the fibroblast property (Figure S2C) before cells reach the next two SSEA1^{+} stages.Consistently, key molecules associated with fibrotic properties, *Lyz* and *Lyzs*, are among the top 20 differentiated genes at the first two transitions (Figure S2B). This is consistent with previous findings that an early step in reprogramming is to destroysomatic regulatory networks(Brambrink et al., 2008; Stadtfeld et al., 2008). Genes involved in embryonic stem cell pluripotency are activated starting at the Thy1^{-}/SSEA1^{-}-to-SSEA1^{+}/DsRed^{+} transition (Figure S2C). Additional ES cell-specific networks are activated in the final transition from the SSEA1^{+}/DsRed^{+} to SSEA1^{+}/DsRed^{-} stages (Figure S2C). See File S1 for the detailed information of differentially expressed genes between transitions in reprogramming.

Our data indicate that to reach the “early reprogrammed” SSEA1^{+}/DsRed^{+} stage, it is important to activate many of the key players involved in the embryonic stem cell core circuitry, including Nanog, Sall4, Esrrb, Dppa4, Dppa5a, Dnmt3b, and Dnmt3l (Figure S2D) (Buganim et al., 2012; Hansson et al., 2012; Polo et al., 2012). Our cell sorting data (Figure

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

S1G) also suggested that the transition of Thy1^{-}-to-SSEA1^{+} is the rate limiting step, because a majority (~80%) of transformed cells were "trapped" in Thy1-/SSEA1- stage and the initial induction of several ESC-specific factors (e.g., Nanog, Sall4, Esrrb) is required to overcome this threshold.For pre-determined cells (SSEA1^{+}/DsRed^{+}), in order to progress to a mature reprogrammed status (SSEA1^{+}/DsRed^{-}), those molecules are further induced to a higher expression level(Figure 2E), possibly to acquire a complete pluripotent state. Furthermore, when cells proceed from the SSEA1^{+}/DsRed^{+} to SSEA1^{+}/DsRed^{-} stages (Figure 2F), more extensive interactions of ESC core regulators are established, including Utf1, Tdgf1, Gsc, Fgf10, T, Chrd, Dppa3, Fgf17, Eomes, Foxa2; indicating that the final step of reprogramming is to reinforce the regulatory pathways in ESC core circuitry.

Discovering a variety ofsources for induced reprogramming and cell-fate manipulation

The choice of somatic cells contributes largely to reprogramming efficiency (Gonzalez et al., 2011). Therefore, we reasoned that our sorted cells might resemblecertain tissue types, which could bebetter and alternative resources for induced reprogramming. To test that ideawe comparedtranscriptome profiles from the reprogramming process with those from various tissue types *in vivo* (Body Atlas analysis)(Kupershmidt et al., 2010). This algorithm was designed to find correlations between genes of interest (queries) and normalized gene expression across all available tissues, cell types, cell lines, and stem cells in a library; this is accomplished by calculating mRNA expression profiles with a positive or negative correlation. We found that the transcriptome of SSEA1^{+}/DsRed^{-} cells most resembled that of cells derived from the visual (choriocapillaris endothelium) (p value < 1X10^{-153}), urogenital (p value < 1X10^{-130}), and immune (p value < 1X10^{-40}) systems (Figure S2G). Interestingly, Thy1^{-}/SSEA1^{-} cells have low significant correlations (p value < 1X10^{-9} to 1X10^{-17}) with any tissue types (Figure S2G), suggesting a high degree of heterogeneity of cell contents in this status (Thy1^{-}/SSEA1^{-}). Thy1^{-}/SSEA1^{-} status might serve as the cell-fate-decisive stage prior to commitment of cell types, because of high heterogeneous tissue types with low mRNA expression correlations. Finally, we showed that cells from the visual system (choriocapillaris endothelium) and immune system might serve as alternative resources for efficient reprogramming due to high transcriptome correlation parameters.

Cell signaling pathways are determinative factors in the "prime" stage before cell-fate commitment

We reasoned that essential genes of cell-fate transitions should be identified in specificsorted cells in reprogramming by genome-wide RNAi screen (Figure 1A). To obtain enriched shRNAs integrated in specific cell stages, we isolated genomic DNA from sorted cells and sequenced by high throughput sequencing (HT seq). Next, to findshRNA targets enriched specifically in each cell population, we performed K-means clustering for identified reads from sorted populations based on the relative enrichment in different cell populations. We obtained four stage-specific gene clusters (A, B, C, D) enriched in each population (Figure 1C). 829 genes are specifically targeted in cluster A (Thy1^{+}/SSEA1^{-}); 784 genes are in cluster B (Thy1^{-}/SSEA1^{-}); 206 genes are in cluster C (SSEA1^{+}/DsRed^{+}); 898 genes are in cluster D (SSEA1^{+}/DsRed^{-}).1972 genes, which are not categorized, are grouped into cluster E (File S2). Surprisingly, we got the highest number of target genes (898 out of 2717 identified genes) from the least cell number population (SSEA1^{+}/DsRed^{-};

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

~0.2-0.4 % of transduced cells; please see File S1), suggesting that ourRNAi screen indeed identified genes with relevant functions to reprogramming, regardless of cell number in each sorted population.

To understand biological functions of shRNA-identified genes, we conducted meta-analysis of enriched-shRNAhits using IPA software (http://www.ingenuity.com/). We identified several canonicalpathways, which were significantly targeted to influence the transitions between each stage of reprogramming (Figure 2A). *Pla2g10*, *Pla2g12b*, *Npr1*, *Gucy1a3*, and *Plch2* (Sperm motility and Synaptic long term depression pathways) are required for thede-differentiation of fibroblasts, because cells were "stuck" in the initial stage these genes are depleted. Strikingly, various signaling pathways are highly over-representative in the second stage of reprogramming ($Thy1^{-}/SSEA1^{-}$). We found a number of known-reprogramming regulators including*PI3K* and *Akt*(CREB signaling pathway)(Yu et al., 2013). Additionally, we found that *Itpr1*, *Itpr2*, *Pdia3*, and *Camk4* are common components linkingseveral signaling pathways (Figure 2A and File S2), such as Nitric oxide, Neuropathic pain, CREB, EGF signaling pathways, and others. This significant enrichment of signaling pathways in this cell population indicates that this stage ($Thy1^{-}/SSEA1^{-}$) might be the "prime" stage, requiring a significant amount of sensing and signaling to define the specific cell fate in the next step of reprogramming.

In Figure 2A, *Egf*, *Flt1*, *Il1rl1*, and *Ly96* (Hepatic fibrosis pathway) are identified in pre-commitment stage ($SSEA1^{+}/DsRed^{+}$), suggesting that it is critical to modulate cell-to-cell signaling and interaction so that transformed cells are able to overcome the rate-limiting step from the "prime" stage. To reach the last stage of reprogramming ($SSEA1^{+}/DsRed^{-}$), depletion of *Cfl1*, *Mprip*, and *Ppp1r12* (Regulation of actin-based motility by Rho pathway) benefits the maturation process of reprogrammed cells (Figure 2A and File S2), indicating that transforming the cytoskeleton is an important step to build ESC-like cellular organization (Sakurai et al., 2014).

Notably, most genes associated with key networks were targeted by shRNAs in the same stage (Figure S3A, S3B, and File S2), including cell signaling, cellular assembly, gene expression control, development, protein synthesis, cell cycle, cell programmed death, and metabolism. Thesehighly targeted networks may serve ascentral hubs to determine the transition of cell-identities.

As previously reported,genes associated with cell cycle or cell death and survival (Banito et al., 2009; Hong et al., 2009; Kawamura et al., 2009; Li et al., 2009; Marion et al., 2009; Utikal et al., 2009) are also identified in our RNAi screen (Figure 2B and File S2) and act as checkpointsin the initial or laststage (Figure 2B). Surprisingly, we found that a significant proportion of essential networks/functions are also responsible for maintaining basal cellular functions, such as cell signaling, metabolism, cell morphology, cellular assembly and organization (Figure 2B and File S2). This finding prompted us to further investigate whether many important regulators in reprogramming are always tissue-specifically expressed in ESCs/iPSCs and whether some of those regulators might be genes whose expression does not change very significantly during the reprogramming process [AU: OK?,yes, OK]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Non-differentially expressedgenes play important roles in modulating cell-fate transitions during reprogramming

To examine whether thenon-differentially expressed genes play any roles in cell-fate decision during reprogramming, we performed in-depth analysis by integratingdata sets generated fromRNAi screen and transcriptome analysis (Figure S3C). First, we asked whether we could identify specific mRNA expression patterns from genes targeted by shRNAs. To do so we used target lists developed via RNAi screen (Figure 1C) as seeds (queries) to identify expression profiles from transcriptome analysis (Figure 1B). We defined mRNA profiles corresponding to five clusters of shRNA-identified targets. However, we did not observe enrichment of specific gene expression patterns among these groups (Figures 3A and 3B), indicating that genes with stage-specific functions to reprogramming may not show corresponding changes in mRNA expression level. Strikingly, a major proportion of identified shRNA targets (~53% to 70%) are genes whose expression did not change during reprogramming (as indicated by the yellow rectangles inFigure3A), showing that these nondifferentially expressedgenes are indeed important for reprogramming transitions.

We further asked whether mRNA expression profiles could predict cell-fate-specific functions in reprogramming. To do so, we took an approach described above but instead using gene lists from transcriptome analysis (Figure 1B) as queries to find specific patterns of enriched shRNA targets. Surprisingly, majority of the genes in group I show no specific enrichments in corresponding shRNA targets during reprogramming (as indicated by the orange rectangles in Figure 3C), suggesting that 50% (238 out of 476 genes) of ESC-enrichedgenes have little function in reprogramming process(Figure 3D). Importantly, group-I, II, and III genes with functional influence in reprogramming are present across various sorted cell populations, instead of only certain stage (Figure 3C and D), indicating that specific roles of tissue-enriched genes couldn't be comprehensively revealed by expression profiling in reprogramming.

We found that 362 out of 566 genes (~64%) and 668 out of 1365 genes (~49%) in group IV and V respectively (blue rectangles in Figure 3C), are specifically identified by RNAi screenduring various stages of reprogramming (Figures 3C and 3D), showing that non-differentially expressed genes in group IV and V significantly contribute to cell-fate transitions. As expected, agreat proportion of matched genes (~36% and 51% respectively) in mRNA groups IV and V are clustered in non-enriched shRNA group E (as indicated by the orange rectangle in Figure 3C and Figure 3D). For complete information ofRNAi-identified targets see Files S2. In summary, we found that non-differentially expressedgenesin groupsIV and Vindeed play important roles to modulate the reprogramming progress.

High discovery rate in identifyingpositive regulators or barrier genes for reprogramming

We next performed validation experiments on target genes identified from RNAi screen and bioinformatics analyses presented above. To assess shRNA-identified targets, we selectedstage-specifically enriched targets (reads with $\log_{10}$ value > 1.5; File S2) from the initial and mature-reprogrammed stages(Figure 1C, S3A, and S3B). To examine targets

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

from transcriptome analysis, we selected genes highly induced in the group I(Figure 1B and File S1). Most selected genes from both analysisencodedproteins involved in transcriptionalregulation.

To determine ifshRNA-identified genesin specific populationscan promote or comprise reprogramming, we performedsiRNA-mediated knockdown of specific genes upon induced reprogramming. We first picked genes selectively targeted by shRNAsin theThy1^{+}/SSEA1^{-} cell population (group A)(Figure S3A and File S2), reasoning that these genes might be positive regulators for reprogramming.To assess reprogramming efficiencyof cells with siRNA-mediated depletion,we quantifiedOct4-GFP positive colonies two weeks after virus transduction. Of six selected genes, depletion of five (~83%), *Dmbx1, Gsc, Med21, Hnf4g, Nobox,* and *Asb4*, significantly reduced (p value < 0.05) reprogramming efficiency(Figure 4A). The knockdown level of target genes was verified by quantitative RT polymerase chain reaction (qRT-PCR)(Figure S5A). To further independently validate the observed phenotype of these genes by an alternative approach, we employed shRNA-mediated knockdown of thesegenes. Stronger reduction of iPSC generation was observed with shRNA-mediated depletion of *Dmbx1*, *Gsc, Med21, Hnf4g, Nobox,* and *Asb4* (Figure S4A). Additional genes from shRNA group A (*Psmd9* and *Mef2c*) were also tested to show the same phenotype of iPSC reduction (Figure S4A). The knockdown level of target genes was verified by qRT-PCR(Figure S5B). Most importantly, 4 out of 8 tested positive regulators (*Dmbx1, Hnf4g, Nobox*and *Asb4*; Figure S4A) show no expression changesduring reprogramming (Figure 4D), supporting our hypothesis that nondifferentially expressed genes indeed contribute to cell-fate decision.Overexpression of nondifferentially expressed gene (*Nobox*) is sufficient to boost reprogramming efficiency by ~2-fold, compared with *DsRed* control (Figure S4G, H).

Using a similar approach, we examined the effect of genes (group D) selectively targeted by shRNAs in mature reprogrammed cells (SSEA1^{+}/DsRed^{-}), assumingthat they might represent reprogramming barriers (Figure S3B and File S2). Following knockdown of 16 candidates at early stages of reprogramming (including Tfdp1, Gtf2e1, Nfe2, Foxn3, Erf, Cdkn2aip, Msx3, Ssbp3, Dbx1, Hoxd4, Lzts1, Arx, Hoxd12, Gtf2i, Ankrd22, and *Hoxc10*), depletion of 12(75%)improved reprogramming efficiency by at least a two-fold (dotted line, Figure 4B), compared with controls.Several gene targets (*Tfdp1, Cdkn2aip, Msx3, Ssbp3, Dbx1, Ankrd22*) were further confirmed by shRNA KD for their barrier roles in reprogramming (Figure S4B). The mRNA level of targeted genes by siRNAs or shRNAs was verified by qRT-PCR(Figure S5C and S5D). Strikingly, 8 out of 16 tested barrier genes (*Nfe2, Cdkn2aip, Msx3, Dbx1, Lzts1, Arx, Gtf2i,* and *Ankrd22*) showed no expression changes during reprogramming(Figure 4E), again supporting our findings that many nondifferentially expressed genes act as important modulators for cell-fate transition.To further examine the roles of these non-differentially expressed genes, we picked several genes with no expression changes (mRNA group IV and V) from each shRNA-enriched group (A to D) for testing reprogramming efficiency (File S2). We found that genes(*Gja3, Olfr1271, Fkbp11, Mdm1, Myo15, Gucy2g*) identified in early or pre-committed cell populations (shRNA group A to C) are required to efficient reprogramming, while genes identified in shRNA group D (*Lasp1, Hspa8*) are obstacles for reprogramming. The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

knockdown efficiency of select genes targeted by shRNAs was verified by qRTPCR(Figure S5E). For detailed information of non-differentially expressed genes with cell-fate modulation functions, see File S2.

Next, we tested the function of barrier-genesby overexpressing themduring reprogramming with OSKM. Expression of these factors was confirmed by western blotting or immunofluorescence (Figure S3D and S3E). Overexpression of barrier genes compromised reprogramming efficiency by ~40% to 80%, compared with DsRed controls (Figure 4C), demonstrating that targets identified by our RNAi screen and bioinformatics analyses indeed function as barriers to reprogramming. We further examined the roles of target genes identified in shRNA group A and B in reprogramming. During reprogramming, we expressed Mef2c (shRNA-identified group A) and Pdia3 (key component in signaling pathway in shRNA-identified group B; Figure 2A) and observed that iPSC generation is greatly enhanced by 4~6-fold, compared with DsRed control (Figure S4G, H).

To test the functionof genes identified by the transcriptome analysis, we asked whether genes highly inducedduring reprogramming (group I)contributeto maintainingESC identity. Among groupI genes, we analyzed the effect of a panel of transcription factors with little-known function onESC self-renewal (Figure S3F). In addition, we determined the role of positive regulators (Figure 4A, S4A) in ESC identity. To do so, we treatedOct4-EGFP ESCswith specific siRNAsandfour days later assessedESC self-renewal using flow cytometry to detect anEGFP signal.In 16 of 64 tested genes (25%),the Oct4-EGFP signal was significantlyreduced (Z score >2)(Figure S3F). In addition to the known-regulatory factors to ESCs/iPSCs (Nanog and Oct4), we discovered several novel key players to maintain ESC identity, such as Asb4, Dmbx1, Gbx2, Gsc, Hnf4g, Klf5, L3mbtl2, Med21, Mef2c, Nobox, Pcgf6, Phox2a, Tcf15, and *Trim28*. In summary, our genome-wide RNAi screen with sorted-cell populations efficiently identifiedkey regulators, serving either positive roles (e.g., *Dmbx1, Gsc, Med21, Hnf4g, Mef2c*, and *Psmd9*) or barrier roles (e.g., *Nfe2, Cdkn2aip, Msx3, Dbx1, Lzts1, Arx, Gtf2i*, and *Ankrd22*) during reprogramming.We also identified severalnew regulators (e.g., *Asb4, Gbx2, Gsc, Hnf4g, Mef2c*) that play important functions to maintain ESC identities. Collectively, our genome-wide RNAi screen has identified numerous novel regulators of reprogramming (Figure S6), which lays the comprehensive foundation of molecular requirements and regulatory networks during reprogramming.

DISCUSSION

In this study, we sought to define the molecular signatures of step-wise induced reprogramming by functional genomics. We dissected the regulatory networks employing a pooled genome-wide shRNA library in a step-wise manner by applying FACS to isolate groups of distinct cell populations, representing four critical steps from initiation to maturation of induced reprogramming. Results of our RNAi screen provided unbiasedfunctional insight into essential factors during each step of the reprogramming progress. The high validation rate of identified genes in this study suggests that our new strategy is highly valuable to discover key regulatory molecules/networks in the reprogramming process.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We found that the majority of transformed cells are “trapped” in the transition stage (Thy1^{-}/SSEA1-), with divergent transcriptome showing correlations to various tissue types.This finding implies that cells are reset at this “prime” phase where cells might have the potential to adopt distinct cell fates until the “right” molecular networks are rebuilt. This notion is supported by recent studies (Hansson et al., 2012; Polo et al., 2012; Shu et al., 2013) showing that re-administration of OSKM or lineage specifiers into those transitioning cells drove more cells into pluripotent or other desired states. The potential diversity of cell fates at the Thy1^{-}/SSEA1^{-} stage is usually ignored, probably because the only desired cell type here is the pluripotent stem cells. But, these “transitioning” cells with high plasticity may provide a good starting point for various cell-fate inter-conversions.

A recent study usinga similar approach (Polo et al., 2012) has shown that a majority of cells even with prolonged culture after sorting did not greatly change their identities. Our transcriptome analysis is consistent with previous studies (Brambrink et al., 2008; Buganim et al., 2012; Polo et al., 2012) to delineate the step-wise marker genes during the reprogramming. Despite our efforts to obtain a relatively “terminated” cell-fate of transformed cells in each population at the end of 2weeks reprogramming course, it is possible that there are cells present at different levels and stages of latencies in reprogramming making is difficult to completely rule out the heterogeneity issues in our analysis.

In this study, we used DsRed expression as anindicator of “mature reprogrammed” cells. Although Nanog, Sall4, and Essrrb genes have been slightly activated in SSEA1^{+}/DsRed^{+} cells, these genes are further induced to higher expression level (File S1) and ESC-like epigenetic regulation is resotred to silence the retroviral gene (pMX-*DsRed*) in only SSEA1^{+}/DsRed^{-} cells. Therefore, we reasonthat pMX-*DsRed* silencing may provide a better definition of “mature reprogrammed” cells, instead of activation of Oct4 gene, which has been surpringly shownto represent heterogeneously reprogrammed cells (Polo et al., 2012).

Our RNAi screen may not have captured all possible modulators of reprogramming, probably owing to several factors including heterogeneity of virus transduction of shRNAs and OSKM, insufficient knockdown of target genes, and other potential technical issues in this multiple-step screening process. These limitationscould be overcome by using newer algorithms to design efficient shRNA libraries, CRISPR-Cas9 technologies, and homogeneous reprogramming system (polycistronic expression or somatic cells harboring inducible reprogramming factors). Despite these caveats, we provide aproof of principle that unbiased pooled RNAi screen can be usedto dissect functional requirement in multistep complex biological pathways.

Recently, it has been suggested that the non-differentially expressed genes or conserved pathways might play complex roles contributing to tissue-specific functions or oncogenesis(Locasale, 2013). Herewe vigorously tested the hypothesis that the non-differentially expressed genes play important roles in directing cell fate decisions. Our functional genomics approach show that in addition to tissue-specific genes, many non-differentially expressed genes actually play important roles in cell-fate transitionduring reprogramming (Figure S6B). Thus we suggest that studies such as ours that use genome-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

wideRNAi screening to define reprogramming mechanismswill havenumerous applications in this field, such as providing novel approaches to small molecule targeting, cell-fate manipulation, and progenitor derivation.More importantly, ourwork not only uncovered the landscape of reprogramming, but also defines the cell-fate determinants at each transition step of induced reprogramming.In summary, our results provide a wealth of information about the functional genetic requirements at various transition steps during reprogramming and may lead to a paradigm shift in viewing the functional significance of genomic infrastructure in biology.

EXPERIMENTAL PROCEDURES

Oct4-EGFP Mouse Embryonic Fibroblast Derivation

Oct4-EGFP MEFs were derived from the mouse strain B6;129S4-Pou5f1$^{tm2(EGFP)Jae}$/J (Jackson Laboratory; stock #008214) using the protocol provided on the WiCell Research Institute website (http://www.wicell.org/). In brief, E13.5 embryos were collected from time-mated pregnant female mice. Cells isolated from embryos then were tested for microbial contamination. All animal work was approved by the Institutional Review Board and was performedfollowing Institutional Animal Care and Use Committee guidelines. Oct4-EGFP MEFs were maintained in MEF complete medium (DMEM with 10% FBS, nonessential amino acids, L-glutamine, and no sodium pyruvate). Robustly growing cells (usually < 4 passages) were used for induced reprogramming.

FACS and Whole-genome RNAi Screening

Cells were transduced with retroviruses containing pMXs-DsRed plasmids, harvested three days later and then stained with phycoerythrin-Cy7 (PE-Cy7)-conjugated antibodies targeting Thy1 (25-0902, eBioscience). Thy1/DsRed double-positive cells (Thy1$^+$/DsRed$^+$) were isolated by FACS and allowed to recover three days before introduction of the shRNA library and OSKM. Pseudo viruses expressing a pGIPz-shRNA library and pMXs-OSKM were generated in 293FT and Plate-E cells, respectively. Pseudo viruses were administered at day 0 and day 1 during reprogramming to maximize transduction efficiency. ESC medium was used for culturing transformed cells at day 3 post induction. Two weeks later, cells were harvested and dissociated with trypsin/EDTA. PE-Cy7-conjugated antibodies targeting Thy1 (25-0902, eBioscience) and Alexa Fluor®647-conjugated SSEA1 antibodies (51-8813, eBioscience) were used to detect Thy1 and SSEA1 surface markers. Before isolating cells with FACS, SSEA1$^+$ cells were enriched using Anti-SSEA-1 (CD15) MicroBeads (130-094-530, MiltenyiBiotecGmbH). SSEA1-enriched cells were used for sorting SSEA1$^+$/DsRed$^+$ and SSEA1$^+$/DsRed$^-$ cell populations. SSEA1-depleted cells were used for sorting Thy1$^+$/SSEA1$^-$ and Thy1$^-$/SSEA1$^-$ cell populations. shRNA-library screening in reprogramming was conducted independently three times. Total RNAs and genomic DNAs were extracted from sorted populations for mRNA microarray analysis and SOLiD sequencing analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Acknowledgments

We thank Dr.Tingting Duand Rana Lab members for their advice, helpful discussions, and support. We are grateful for the use of the following Sanford-Burnham Instituteshared resource facilities: genomics and informatics and data management core facilities for array experiments and data analysis; the animal facility for chimera generation and mouse teratomaassays; and the histology and molecular pathology core for characterization of tissues in teratoma tumors.This work was supported in part by grants from the National Institutes of Health.

REFERENCES

Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V, Walsh M, et al. Senescence impairs successful reprogramming to pluripotent stem cells. Genes & development. 2009; 23:2134–2139. [PubMed: 19696146]

Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell stem cell. 2008; 2:151–159. [PubMed: 18371436]

Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 2012; 150:1209–1222. [PubMed: 22980981]

Carey BW, Markoulaki S, Hanna JH, Faddah DA, Buganim Y, Kim J, Ganz K, Steine EJ, Cassady JP, Creyghton MP, et al. Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell stem cell. 2011; 9:588–598. [PubMed: 22136932]

Choi YJ, Lin CP, Ho JJ, He X, Okada N, Bu P, Zhong Y, Kim SY, Bennett MJ, Chen C, et al. miR-34 miRNAs provide a barrier for somatic cell reprogramming. Nature cell biology. 2011; 13:1353–1360.

Grskovic M, Javaherian A, Strulovici B, Daley GQ. Induced pluripotent stem cells--opportunities for disease modelling and drug discovery. Nature reviews Drug discovery. 2011; 10:915–929.

Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature. 2009; 462:595–601. [PubMed: 19898493]

Hansson J, Rafiee MR, Reiland S, Polo JM, Gehring J, Okawa S, Huber W, Hochedlinger K, Krijgsveld J. Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. Cell reports. 2012; 2:1579–1592. [PubMed: 23260666]

Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature. 2009; 460:1132–1135. [PubMed: 19668191]

Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, Di Giorgio FP, Koszka K, et al. A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell stem cell. 2009; 5:491–503. [PubMed: 19818703]

Jopling C, Boue S, Izpisua Belmonte JC. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nature reviews Molecular cell biology. 2011; 12:79–89.

Judson RL, Babiarz JE, Venere M, Blelloch R. Embryonic stem cell-specific microRNAs promote induced pluripotency. Nature biotechnology. 2009; 27:459–461.

Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Izpisua Belmonte JC. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature. 2009; 460:1140–1144. [PubMed: 19668186]

Kim NH, Kim HS, Li XY, Lee I, Choi HS, Kang SE, Cha SY, Ryu JK, Yoon D, Fearon ER, et al. A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial mesenchymal transition. The Journal of cell biology. 2011; 195:417–433. [PubMed: 22024162]

Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A. Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell stem cell. 2011; 8:96–105. [PubMed: 21211784]

Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, Blasco MA, Serrano M. The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature. 2009; 460:1136–1139. [PubMed: 19668188]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Li MA, He L. microRNAs as novel regulators of stem cell pluripotency and somatic cell reprogramming. BioEssays : news and reviews in molecular, cellular and developmental biology. 2012; 34:670–680.

Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q, et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell stem cell. 2010; 7:51–63. [PubMed: 20621050]

Li Z, Rana TM. A kinase inhibitor screen identifies small-molecule enhancers of reprogramming and iPS cell generation. Nature communications. 2012; 3:1085.

Li Z, Yang CS, Nakashima K, Rana TM. Small RNA-mediated regulation of iPS cell generation. The EMBO journal. 2011; 30:823–834. [PubMed: 21285944]

Liao B, Bao X, Liu L, Feng S, Zovoilis A, Liu W, Xue Y, Cai J, Guo X, Qin B, et al. MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. The Journal of biological chemistry. 2011; 286:17359–17364. [PubMed: 21454525]

Lipchina I, Elkabetz Y, Hafner M, Sheridan R, Mihailovic A, Tuschl T, Sander C, Studer L, Betel D. Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. Genes & development. 2011; 25:2173–2186. [PubMed: 22012620]

Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. Nature reviews Cancer. 2013; 13:572–583.

Maherali N, Hochedlinger K. Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Current biology : CB. 2009; 19:1718–1723. [PubMed: 19765992]

Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R, et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell stem cell. 2007; 1:55–70. [PubMed: 18371336]

Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, Fernandez-Capetillo O, Serrano M, Blasco MA. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature. 2009; 460:1149–1153. [PubMed: 19668189]

Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature. 2010; 463:621–626. [PubMed: 20054295]

Nichols J, Silva J, Roode M, Smith A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development. 2009; 136:3215–3222. [PubMed: 19710168]

Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, et al. Chromatin-modifying enzymes as modulators of reprogramming. Nature. 2012; 483:598–602. [PubMed: 22388813]

Pfaff N, Fiedler J, Holzmann A, Schambach A, Moritz T, Cantz T, Thum T. miRNA screening reveals a new miRNA family stimulating iPS cell generation via regulation of Meox2. EMBO reports. 2011; 12:1153–1159. [PubMed: 21941297]

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012; 151:1617–1632. [PubMed: 23260147]

Robinton DA, Daley GQ. The promise of induced pluripotent stem cells in research and therapy. Nature. 2012; 481:295–305. [PubMed: 22258608]

Sakurai K, Talukdar I, Patil VS, Dang J, Li Z, Chang KY, Lu CC, Delorme-Walker V, Dermardirossian C, Anderson K, et al. Kinome-wide functional analysis highlights the role of cytoskeletal remodeling in somatic cell reprogramming. Cell stem cell. 2014; 14:523–534. [PubMed: 24702998]

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell stem cell. 2010; 7:64–77. [PubMed: 20621051]

Shu J, Wu C, Wu Y, Li Z, Shao S, Zhao W, Tang X, Yang H, Shen L, Zuo X, et al. Induction of Pluripotency in Mouse Somatic Cells with Lineage Specifiers. Cell. 2013; 153:963–975. [PubMed: 23706735]

Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A. Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS biology. 2008; 6:e253. [PubMed: 18942890]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell. 2012; 151:994–1004. [PubMed: 23159369]

Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K. Role of the murine reprogramming factors in the induction of pluripotency. Cell. 2009; 136:364–377. [PubMed: 19167336]

Stadtfeld M, Maherali N, Breault DT, Hochedlinger K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell stem cell. 2008; 2:230–240. [PubMed: 18371448]

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

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007; 131:861–872. [PubMed: 18035408]

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]

Tiscornia G, Vivas EL, Izpisua Belmonte JC. Diseases in a dish: modeling human genetic disorders using induced pluripotent cells. Nature medicine. 2011; 17:1570–1576.

Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K. Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature. 2009; 460:1145–1148. [PubMed: 19668190]

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A, et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell stem cell. 2010; 7:618–630. [PubMed: 20888316]

Wu SM, Hochedlinger K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nature cell biology. 2011; 13:497–505.

Yamanaka S. Elite and stochastic models for induced pluripotent stem cell generation. Nature. 2009; 460:49–52. [PubMed: 19571877]

Yang C-S, Rana TM. Learning the molecular mechanisms of the reprogramming factors: let's start from microRNAs. Molecular bioSystems. 2013; 9:10–17. [PubMed: 23037570]

Yang CS, Li Z, Rana TM. microRNAs modulate iPS cell generation. RNA. 2011a; 17:1451–1460. [PubMed: 21693621]

Yang CS, Lopez CG, Rana TM. Discovery of nonsteroidal anti-inflammatory drug and anticancer drug enhancing reprogramming and induced pluripotent stem cell generation. Stem cells. 2011b; 29:1528–1536. [PubMed: 21898684]

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453:519–523. [PubMed: 18497825]

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007; 318:1917–1920. [PubMed: 18029452]

Yu Y, Liang D, Tian Q, Chen X, Jiang B, Chou BK, Hu P, Cheng L, Gao P, Li J, et al. Stimulation of Somatic Cell Reprogramming by Eras-Akt-Foxo1 Signaling Axis. Stem Cells. 2013

Zhu S, Wei W, Ding S. Chemical strategies for stem cell biology and regenerative medicine. Annual review of biomedical engineering. 2011; 13:73–90.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

HIGHLIGHTS

1. Genome-wide RNAi screen identified genes required for various steps of reprogramming
2. Non-differentially expressed genes play important roles in cell-fate-transitions
3. *Dmbx1*, *Hnf4g*, *Nobox* and *Asb4* are essential for efficient reprogramming
4. *Nfe2*, *Cdkn2aip*, *Msx3*, *Dbx1*, *Lzts1*, *Arx*, *Gtf2i*, and *Ankrd22* compromise reprogramming

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

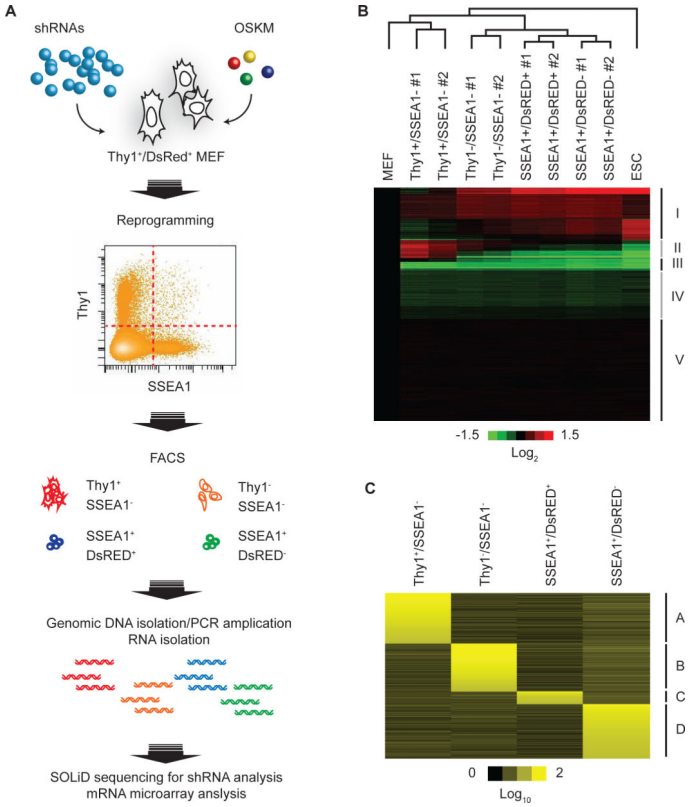

Figure 1. RNAiScreen IdentifiesKey Modulators of Induced Reprogramming

(A)RNAi screening strategy. Thy1^{+}/DsRed^{+} MEFs were transduced with a library of ~57000 shRNAs and OSKM and sorted into four populations based on Thy1, SSEA-1, and DsRed marker combinations. Integrated shRNAs were amplified from genomic DNA isolated from those populations and sequenced.

(B) Heat map showing mRNA expression profiles during reprogramming. RNA extracted from cell populations in was analyzed by microarray. Data was processed and visualized using Cluster and Java TreeView, respectively. Gene expression patterns are clustered defined as I to V. Duplicate samples are designated #1 and #2. Fold changes in mRNA level relative to MEFs in five expression groups are represented in $\log_2$ scale.

(C) Heat map showing enriched shRNA targets in sorted populations along the reprogramming process. Targets identified by shRNA reads were clustered by using Cluster 3.0 and visualized with Java TreeView. Letters A to D mark five distinct clusters. Gene ontology analysis was performed using IPA. Reads of shRNA-identified targets are shown in $\log_{10}$ scale.

Please also see Figures S1 and S2.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

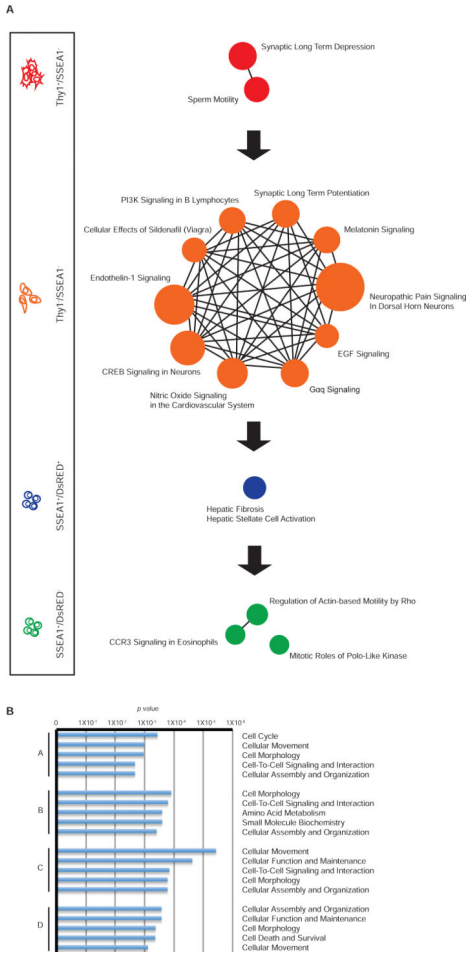

Figure 2. Key regulatory hubs are identified in each stage during induced reprogramming

(A) Over-representative canonical pathways identified in reprogramming.

Qualified hits (shRNA reads > 1.5 in $\log_{10}$ scale) were analyzed using IPA. Only the most significant pathways (p value < 0.01) are shown here. The size of each circle is proportional to the p value to represent the significance. The cell stages are shown inside the box on the left.

(B) Key molecular and cellular functions identified by shRNA screening.

Qualified hits (shRNA reads > 1.5 in $\log_{10}$ scale) were analyzed using IPA. Cluster identifications are shown at left. p values are based on Fisher's Exact test.

Please also see Figure S3.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

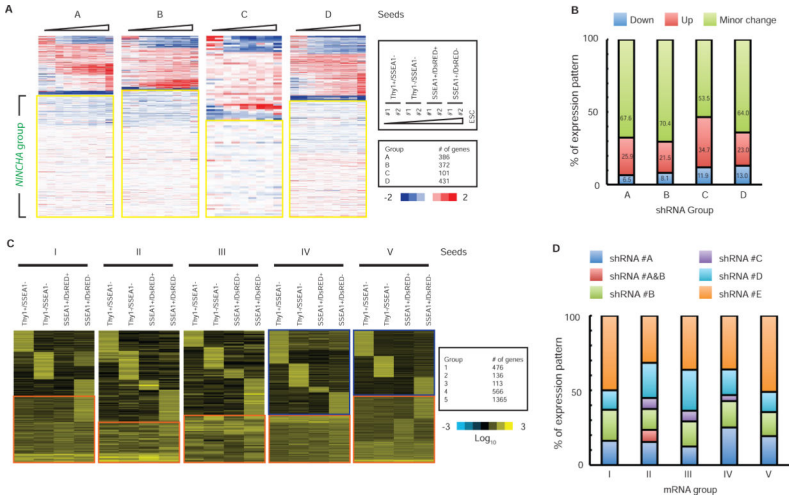

Figure 3. Revealing Non-Differentially Expressed GenesRegulating Cell-fate Transitions in Reprogramming by Integrated shRNAScreening and TranscriptomeProfiling

In-depth analysis performed by integrating data sets generated from RNAi screen and transcriptome analysis. Scheme showing strategy to perform integrative analysis of shRNA screening and transcriptome data is shown in **Figure S3C**. **(A)** Heat map illustrating mRNA expression profile of genes identified from shRNA(groups A-D) library screening in Figure 1C. Only shRNA target genes having read cutoff > 1.5 (log_{10} scale) are analyzed. shRNA target genes were used as seeds to identify mRNA expression profiles from transcriptome analysis presented in Figure 1B. mRNA profiles corresponding to four clusters of shRNA-identified targets were defined. From left to right, mRNA expression trends are shown from initiation maturation stages of reprogramming (upper right rectangle), and number of qualified targets is listed in lower right rectangle. Non-differentially expressedgenes are highlighted within yellow rectangle in each group. Fold-changes in expression relative to MEFs are shown in log_2 scale.**(B)**Proportion of three different expression patterns corresponding to shRNA-identified genes shown in Figure 3A. Genes identified in Figure 1B were shown as a percentage corresponding to each group of shRNAsgroups A-D in Figure 1C. Blue indicates down-regulated genes, red indicates up-regulated genes, and green indicates minor expression changes. The relative proportion of each expression pattern is indicated in three different colors.

(C) Heat map illustrating shRNA reads of unique expression pattern groups shown in Figure 1B. Reads from shRNA screen were analyzed using gene lists for each group from Figure 1B (group I-V). Only qualified targets (shRNA reads > 1.5 in log_{10} scale) are analyzed. Genes from each cluster (I to V) were used as seed to identify shRNA enrichment. From left to right, shRNA reads were presented from initiation stage to maturation stages as indicated above. The number of qualified genes from mRNA microarray analysis and shRNA library screening is listed on right. Genes identified as non-enriched shRNA targets are highlighted with orange rectangles (described as group E in Fig 3D below, in text, and in **Supplementary File 2**). Non-differentially expressed genes with cell-fate modulation functions are highlighted within blue rectangle in group IV and V. Reads of the shRNA library are shown in log_{10} scale.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(D) Proportion of enriched shRNA-identified targets corresponding to clusters in transcriptome analysis shown in Figure 3C. shRNA-identified genes in Figure 1C were shown as a percentage corresponding to expression groups clustered in Figure 1B. Enriched shRNA targets are indicated in different colors above.
Please also see Figure S3.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

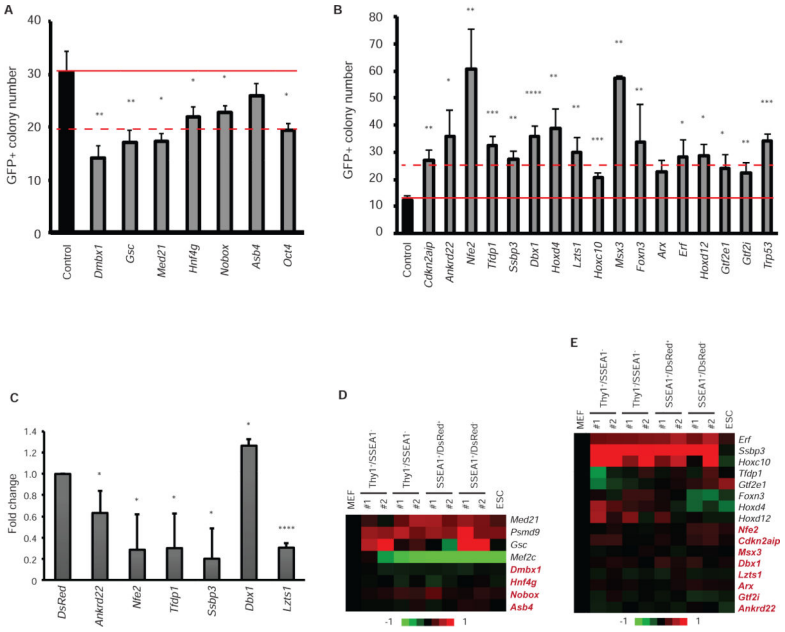

Figure 4. Functional Validation shows a High discovery rate in identifying positive regulators or barrier genes for reprogramming

(A) Bar graph showing reprogramming efficiency following siRNA knockdown of positive regulators. Indicated siRNAs plus OSKM were introduced into 4X10^4 cells of Oct4-EGFP MEFs, and colonies were scored for EGFP positivity. Oct4 knockdown served as positive control.

NontargetingsiRNAserved as negative control (Control). Error bars represent S.E.M, n 3. Solid line marks Control value, and dashed line shows cutoff value based on Oct4 knockdown. Student's *t*-test, * p value < 0.05; ** p< 0.005.

(B) Bar graph showing MEF reprogramming efficiency following barrier gene depletion.Reprogramming efficiency was assayedas in (A). *Trp53*(*p53*) knockdown served as positive control. NontargetingsiRNA served as negative control (Control). Error bars represent S.E.M, $n \geq 3$. Solid line marks Control value, and dashed line marks the cutoff value of two-fold changes. Student's *t*-test, * p value < 0.05; ** p< 0.005; *** p< 0.0005; **** p< 0.00005.

(C) Fold-changes in MEF reprogramming efficiency following barrier gene overexpression.Transgenes plus OSKM were introduced into 4X10^4 cells of Oct4-EGFP MEFs and assayed as described above. Reprogramming efficiency was calculated following normalization to DsRed control.Student's *t*-test, * p value < 0.05 and **** p< 0.00005.

(D) Expression profiling of genes potentially essential to reprogramming. Expression of specific genes was examined during reprogramming. MEFs and ESCs serve as controls for two determined cell types. Replicates are designated #1 and #2. Non-differentially expressed genes are highlighted in red-bold text form.Fold-change values are presented ona $\log_2$ scale.

(E) Expression profiling for putative barrier genes. Expression of specific genes was analyzed as in (D). Non-differentially expressed genes are highlighted in red-bold text form. Please also see Figure S3, S4, and S5.