Cell Reports
Report



# 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, Mail Code 0762, La Jolla, CA 92093, USA
*Correspondence: trana@ucsd.edu
http://dx.doi.org/10.1016/j.celrep.2014.07.002


## SUMMARY

**Although transcriptome analysis can uncover the molecular changes that occur during induced reprogramming, the functional requirements for a given factor during stepwise cell-fate transitions are left unclear. Here, we used a 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,* and *Pdia3*) constitute the major regulatory networks before cells acquire pluripotency. Activation of a specific gene set (e.g., *Utf1* or *Tdgf1*) is important for mature induced pluripotent stem cell formation. Strikingly, a major proportion of RNAi targets (~53% to 70%) includes genes whose expression levels are unchanged during reprogramming. Among these non-differentially expressed genes, *Dmbx1, Hnf4g, Nobox*, and *Asb4* are important, whereas *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, and c-Myc (OSKM); Nanog; and Lin28 (Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007). Induced pluripotent stem cells (iPSCs) strongly resemble embryonic stem cells (ESCs) and hold promise for 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 microRNAs (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 a few markers, including Thy1, alkaline phosphatase, and SSEA1, can be used to identify transformed cells through the process of induced reprogramming, whereas ESC-specific genes (*Nanog*, *Oct4*, and *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; Marión 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 a factor critical to efficient reprogramming (Carey et al., 2011; Soufi et al., 2012; Sridharan et al., 2009). Many ESC-specific genes (e.g., *Esrrb*, *Sall4*, and *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 this with a 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

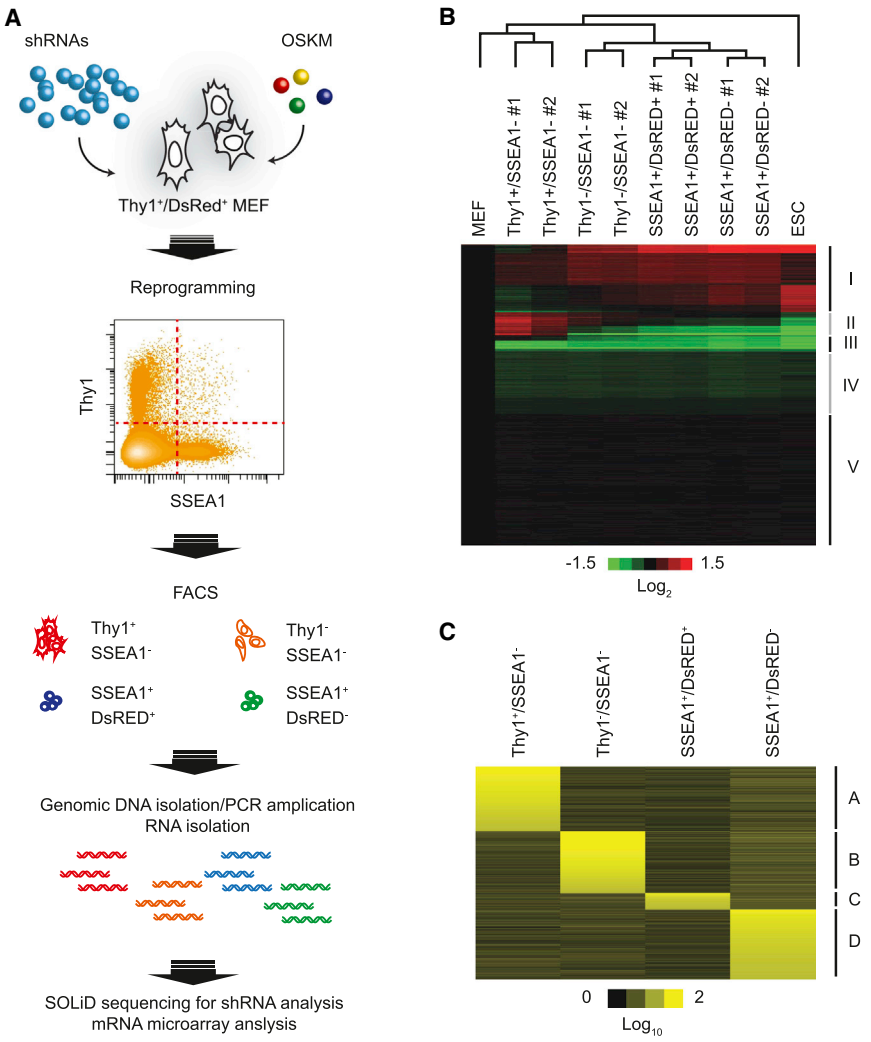



**Figure 1. RNAi Screen Identifies Key Modulators of Induced Reprogramming.**

(A) RNAi screening strategy. Thy1$^+$/DsRed$^+$ MEFs were transduced with a library of ~57,000 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 was analyzed by microarray. Data were processed and visualized using Cluster and Java TreeView, respectively. Gene expression patterns are clustered into groups defined as I–V. Duplicate samples are designated as nos. 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– D mark four distinct clusters. GO analysis was performed using IPA. Reads of shRNA-identified targets are shown in $\log_{10}$ scale.

See also Figures S1 and S2.

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 in modulating cell-fate transitions during reprogramming.

## RESULTS

### Experimental Strategy for Genome-wide RNAi Screen in Induced Reprogramming

To elucidate the molecular requirements of induced reprogramming, we conducted loss-of-function assays during the reprogramming process. We used a genome-wide RNAi screen and transcriptome analysis upon induced reprogramming to functionally validate the roles of key regulators in a stepwise 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 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– 5 post induced reprogramming), whereas 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 ESCs (Macfarlan et al., 2011; Wolf and Goff, 2007); thus, we used the *DsRed* gene driven by retroviral long terminal repeats (pMX-*DsRed*) as a marker to differentiate incomplete reprogrammed cells from mature ones. We used these three markers to define four different cell-fate stages in reprogramming: Thy1$^+$/SSEA1$^-$ for the initial stage, Thy1$^-$/SSEA1$^-$ for the transition stage, SSEA1$^+$/DsRed$^+$ for the predetermined (early reprogrammed) stage, and SSEA1$^+$/DsRed$^-$ for the mature reprogrammed stage (Figure 1A). As starting material, we isolated high-purity (~98%) MEFs expressing Thy1 and DsRed (Figures S1B and S1E) by using fluorescence-activated cell sorting (FACS). Reprogramming was initiated by transducing these cells with retroviruses expressing OSKM plus lentiviruses containing a whole-genome small hairpin RNA (shRNA) library (Figures 1A, S1C, and S1D).

We sorted cells 14 days later (Figure S1C), when reprogramming is reportedly complete 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) representing stages defined above (Figure 1A). Surprisingly, most cells (>80%) were at the transition (Thy1$^-$/SSEA1$^-$) stage, whereas only 1%–2.5% reached SSEA1$^+$ stages (Figure S1G and Table S1), suggesting that re-establishing pluripotency networks is the rate-limiting step in reprogramming. Four sorted cell populations were confirmed to properly represent the normal reprogramming process by examining ESC-specific regulators (e.g., *Esrrb*, *Nanog*, *Lin28a*, and *Sall4*) and mesenchymal-to-epithelial transition regulators (e.g., *Cdh1*, *Ocln*, *Krt8*, *Snai1*, *Zeb1/2*, and *Ncam1*) (Figure S1H).

To define the transcriptome in sorted populations, we used k-means clustering to profile gene expression patterns and identified five groups (I–V) of mRNAs (Figure 1B and Table S1). We defined groups I, II, and III as "differentially expressed" genes and groups IV and V as non-differentially expressed or unchanged genes during the reprogramming process. Gene ontology (GO) analysis showed that genes associated with embryonic development, cell cycle, and cell death were significantly overrepresented in groups I– III, and that genes associated with cellular function and maintenance, molecular transport, and metabolism were significantly enriched in groups IV and V (Figures S1I and S1J and Table S1). As expected, this finding demonstrates that differentially expressed genes (groups I– III) are highly related to ESC function and that non-differentially expressed genes (groups IV and V) are related to basal cellular functions.

## Identifying Key Transcriptome Hallmarks 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$^-$ (1,373 genes) and Thy1$^+$/SSEA1$^-$-to-Thy1$^-$/SSEA1$^-$ (1,387 genes) transitions (Figures S2A and S2B), whereas 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 pushes committed 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 deconstructing-of-somatic-networks step. Following these steps, the next two transitions are cell-fate-committing phases, wherein ESC-specific regulatory networks are acquired for attaining pluripotent status in the context of dominant OSKM expression (Figures S2D–S2F).

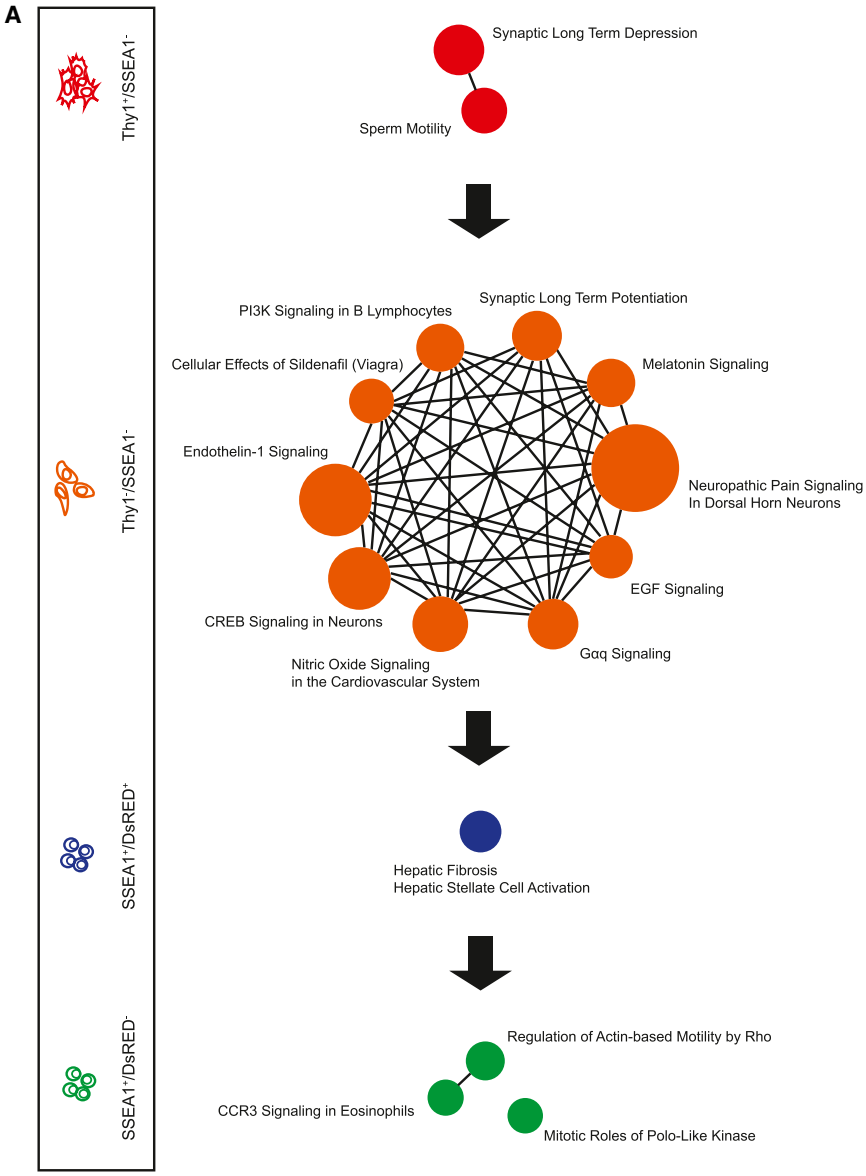
Stage-specific genes are identified in each transition (e.g., *Lyz*, *Lyzs*, *Mrc1*, *Slc38a5*, *Laptm5*, *Ms4a6d*, *Nanog*, *Sall4*, *Esrrb*, *Dppa4*, *Dppa5a*, *Dnmt3b*, and *Dnmt3l*) (Figures S2B). Cellular functions critical for transition steps of reprogramming were identified by GO analysis (Figure 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 subsequent stages.

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 for modulating 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 the destruction of somatic regulatory networks (Brambrink et al., 2008; Stadtfeld et al., 2008). Genes involved in ESC pluripotency are activated starting at the Thy1$^-$/SSEA1$^-$-to-SSEA1$^+$/DsRed$^+$ transition (Figure S2C). Additional ESC-specific networks are activated in the final transition from the SSEA1$^+$/DsRed$^+$ to SSEA1$^+$/DsRed$^-$ stages (Figure S2C). See Table S1 for the detailed information about 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 ESC 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 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 the Thy1$^-$/SSEA1$^-$ stage, and the initial induction of several ESC-specific factors (e.g., Nanog, Sall4, and Esrrb) is required to overcome this threshold. For predetermined 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*, and *Foxa2*, indicating that the final step of reprogramming is to reinforce the regulatory pathways in ESC core circuitry.

## Discovering a Variety of Sources for Induced Reprogramming and Cell-Fate Manipulation

The choice of somatic cells contributes significantly to reprogramming efficiency (González et al., 2011). Therefore, we reasoned that our sorted cells might resemble certain tissue types, which could be better and alternative resources for induced reprogramming. To test this idea, we compared transcriptome profiles from the reprogramming process with those from various tissue types in vivo (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 < $1 \times 10^{-153}$), urogenital (p value < $1 \times 10^{-130}$), and immune (p value < $1 \times 10^{-40}$) systems (Figure S2G). Interestingly, Thy1$^-$/SSEA1$^-$ cells have low significant correlations



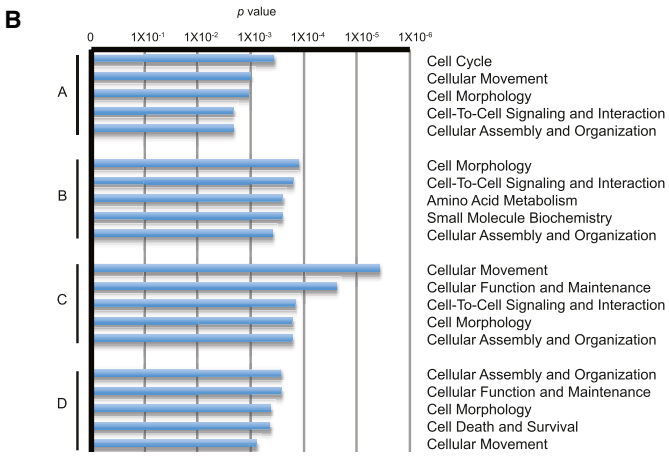




**Figure 2. Key Regulatory Hubs Are Identified in Each Stage during Induced Reprogramming**

(A) Overrepresentative 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.

See also Figure S3.

(p value < $1 \times 10^{-9}$ to $1 \times 10^{-17}$) with any tissue type (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 highly heterogeneous cell types with low mRNA expression correlations to well-defined tissue types. 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 specific sorted cells in reprogramming by a genome-wide RNAi screen (Figure 1A). To obtain enriched shRNAs integrated in specific cell stages, we isolated genomic DNA from sorted cells and sequenced it with high-throughput sequencing. Next, to find shRNA 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, and D) enriched in each population (Figure 1C). In cluster A, 829 genes are specifically targeted (Thy1$^+$/SSEA1$^-$); 784 genes are in cluster B (Thy1$^-$/SSEA1$^-$); 206 genes are in cluster C (SSEA1$^+$/DsRed$^+$); and 898 genes are in cluster D (SSEA1$^+$/DsRed$^-$). Grouped into cluster E are 1,972 genes that are not categorized (Table S2). Surprisingly, we got the highest number of target genes (898 out of 2,717 identified genes) from the population with the lowest cell number (SSEA1$^+$/DsRed$^-$; ~0.2%–0.4% of transduced cells; see Table S1), suggesting that our RNAi screen indeed identified genes with relevant functions to reprogramming regardless of the cell number in each sorted population.

To understand the biological functions of shRNA-identified genes, we conducted a meta-analysis of enriched-shRNA hits using Ingenuity Pathway Analysis (IPA) software (http://www.ingenuity.com/). We identified several canonical pathways, 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 the dedifferentiation of fibroblasts, because cells were "stuck" in the initial stage in which these genes are depleted. Strikingly, various signaling pathways are highly overrepresentative 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., 2014). Additionally, we found that *Itpr1*, *Itpr2*, *Pdia3*, and *Camk4* are common components linking several signaling pathways (Figure 2A and Table S2), such as nitric oxide, neuropathic pain, CREB, and EGF signaling pathways. 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 the precommitment 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 Table S2), indicating that transforming the cytoskeleton is an important step for building ESC-like cellular organization (Sakurai et al., 2014).

Notably, most genes associated with key networks were targeted by shRNAs in the same stage (Figures S3A and S3B and Table S2), including cell signaling, cellular assembly, gene expression control, development, protein synthesis, cell cycle, cell programmed death, and metabolism. These highly targeted networks may serve as central hubs for determining 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; Marión et al., 2009; Utikal et al., 2009) are also identified in our RNAi screen (Figure 2B and Table S2) and act as checkpoints in the initial or final stage (Figure 2B). Surprisingly, we found that a significant proportion of essential networks and functions are also responsible for maintaining basal cellular functions, such as cell signaling, metabolism, cell morphology, cellular assembly, and organization (Figure 2B and Table S2). This finding prompted us to further investigate whether many important regulators in reprogramming are always tissue-specifically expressed in ESCs or iPSCs and whether some of those regulators might be genes whose expression does not change very significantly during the reprogramming process.

## Non-Differentially Expressed Genes Play Important Roles in Modulating Cell-Fate Transitions during Reprogramming

To examine whether the non-differentially expressed genes play any roles in cell-fate decision during reprogramming, we performed an in-depth analysis by integrating data sets generated from RNAi 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) for identifying expression profiles from transcriptome analysis (Figure 1B). We defined mRNA profiles corresponding to four 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 in reprogramming may not show corresponding changes in mRNA expression levels. Strikingly, a major proportion of identified shRNA targets (~53%– 70%) are genes whose expression did not change during reprogramming (as indicated by the yellow rectangles in Figure 3A), showing that these non-differentially expressed genes are indeed important for reprogramming transitions.

Second, we asked whether mRNA expression profiles could predict cell-fate-specific functions in reprogramming. To do so, we took the approach described above but instead used gene lists from the transcriptome analysis (Figure 1B) as queries to

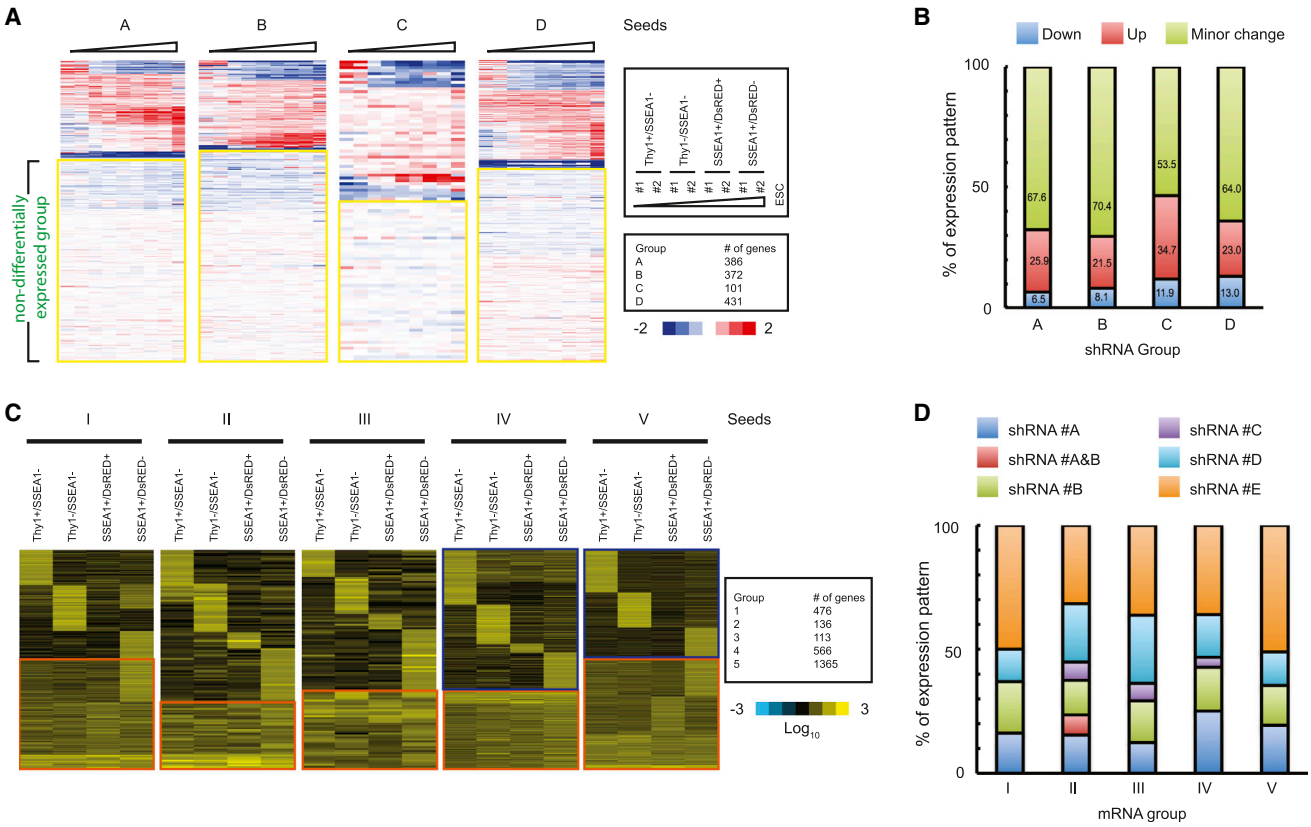



**Figure 3. Revealing Non-Differentially Expressed Genes Regulating Cell-Fate Transitions in Reprogramming by Integrated shRNA Screening and Transcriptome Profiling**

In-depth analysis performed by integrating data sets generated from RNAi screen and transcriptome analysis. A scheme showing the strategy for performing 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 a read cutoff >1.5 ($log_{10}$ scale) are analyzed. shRNA target genes were used as seeds for identifying 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 the number of qualified targets is listed in the lower right rectangle. Non-differentially expressed genes are highlighted within the 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 shRNAs (groups A–D) in Figure 1C. Blue indicates downregulated genes, red indicates upregulated 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 the 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–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 Figure 3D and in Table S2). Non-differentially expressed genes with cell-fate modulation functions are highlighted within the blue rectangle in groups IV and V. Reads of the shRNA library are shown in $log_{10}$ scale.

(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 the different colors listed above.

See also Figure S3.

find specific patterns of enriched shRNA targets. Surprisingly, the 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-enriched genes have little function in the reprogramming process (Figure 3D). Importantly, group I, II, and III genes with functional influence in reprogramming are present across various sorted cell populations, rather than only a certain stage (Figures 3C and 3D), 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 1,365 genes (~49%) in groups IV and V, respectively (blue rectangles in Figure 3C), are specifically identified by RNAi screen during various stages of reprogramming (Figures 3C and 3D), showing that non-differentially expressed genes in groups IV

and V significantly contribute to cell-fate transitions. As expected, a large 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 Figures 3C and 3D). For complete information about RNAi-identified targets, see Table S2. In summary, we found that non-differentially expressed genes in groups IV and V indeed play important roles in modulating the reprogramming progress.

## High Discovery Rate in Identifying Positive Regulators or Barrier Genes for Reprogramming

Next, we performed validation experiments on target genes identified in the RNAi screen and bioinformatics analyses presented above. To assess shRNA-identified targets, we selected stage-specifically enriched targets (reads with $\log_{10}$ value > 1.5; Table S2) from the initial and mature-reprogrammed stages (Figures 1C, S3A, and S3B). To examine targets from the transcriptome analysis, we selected genes highly induced in group I (Figure 1B and Table S1). Most selected genes from both analyses encoded proteins involved in transcriptional regulation.

To determine whether shRNA-identified genes in specific populations can promote or comprise reprogramming, we performed siRNA-mediated knockdown of specific genes upon induced reprogramming. We first picked genes selectively targeted by shRNAs in the Thy1$^+$/SSEA1$^-$ cell population (group A) (Figure S3A and Table S2), reasoning that these genes might be positive regulators for reprogramming. To assess reprogramming efficiency of cells with small interfering RNA (siRNA)-mediated depletion, we quantified Oct4-GFP-positive colonies 2 weeks after virus transduction. Of six selected genes, depletion of five (~83%; *Dmbx1*, *Gsc*, *Med21*, *Hnf4g*, and *Nobox*) significantly reduced reprogramming efficiency (p value < 0.05) (Figure 4A). The knockdown level of target genes was verified by quantitative RT-PCR (qRT-PCR) (Figure S5A). To further independently validate the observed phenotype of these genes through an alternative approach, we employed shRNA-mediated knockdown of these genes. 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 changes during reprogramming (Figure 4D), supporting our hypothesis that non-differentially expressed genes indeed contribute to cell-fate decision. Overexpression of a non-differentially expressed gene (*Nobox*) is sufficient to boost reprogramming efficiency by ~2-fold, compared with *DsRed* control (Figures S4G and S4H).

Using a similar approach, we examined the effect of genes (group D) selectively targeted by shRNAs in mature reprogrammed cells (SSEA1$^+$/DsRed$^-$), assuming that they might represent reprogramming barriers (Figure S3B and Table 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 2-fold (dotted line, Figure 4B) compared with controls. The barrier roles of several gene targets (*Tfdp1*, *Cdkn2aip*, *Msx3*, *Ssbp3*, *Dbx1*, and *Ankrd22*) in reprogramming were further confirmed by shRNA knockdown (Figure S4B). The mRNA levels of genes targeted by siRNAs or shRNAs was verified by qRT-PCR (Figures 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 non-differentially 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 groups IV and V) from each shRNA-enriched group (A to D) for testing reprogramming efficiency (Table S2). We found that genes (*Gja3*, *Olfr1271*, *Fkbp11, Mdm1, Myo15*, and *Gucy2g*) identified in early or pre-committed cell populations (shRNA groups A– C) are required for efficient reprogramming, whereas genes identified in shRNA group D (*Lasp1* and *Hspa8*) are obstacles for reprogramming. The knockdown efficiency of select genes targeted by shRNAs was verified by qRT-PCR (Figure S5E). For detailed information about non-differentially expressed genes with cell-fate modulation functions, see Table S2.

Next, we tested the function of barrier genes by overexpressing them during reprogramming with OSKM. Expression of these factors was confirmed by western blotting or immunofluorescence (Figures S3D and S3E). Overexpression of barrier genes compromised reprogramming efficiency by ~40%– 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 groups A and B in reprogramming. During reprogramming, we expressed Mef2c (shRNA-identified group A) and Pdia3 (key component in the signaling pathway in shRNA-identified group B; Figure 2A) and observed that iPSC generation is greatly enhanced by ~4- to 6-fold compared with DsRed control (Figures S4G and S4H).

To test the function of genes identified by the transcriptome analysis, we asked whether genes highly induced during reprogramming (group I) contribute to maintaining ESC identity. Among group I genes, we analyzed the effect of a panel of transcription factors with little-known function on ESC self-renewal (Figure S3F). In addition, we determined the role of positive regulators (Figures 4A and S4A) in ESC identity. To do so, we treated Oct4-EGFP ESCs with specific siRNAs and assessed ESC self-renewal 4 days later using flow cytometry to detect an enhanced GFP (EGFP) signal. In 16 of 64 tested genes (25%), the Oct4-EGFP signal was significantly reduced ($Z$ score > 2) (Figure S3F). In addition to the known regulatory factors of ESCs or iPSCs (Nanog and Oct4), we discovered several key players that 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 identified key 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 several additional

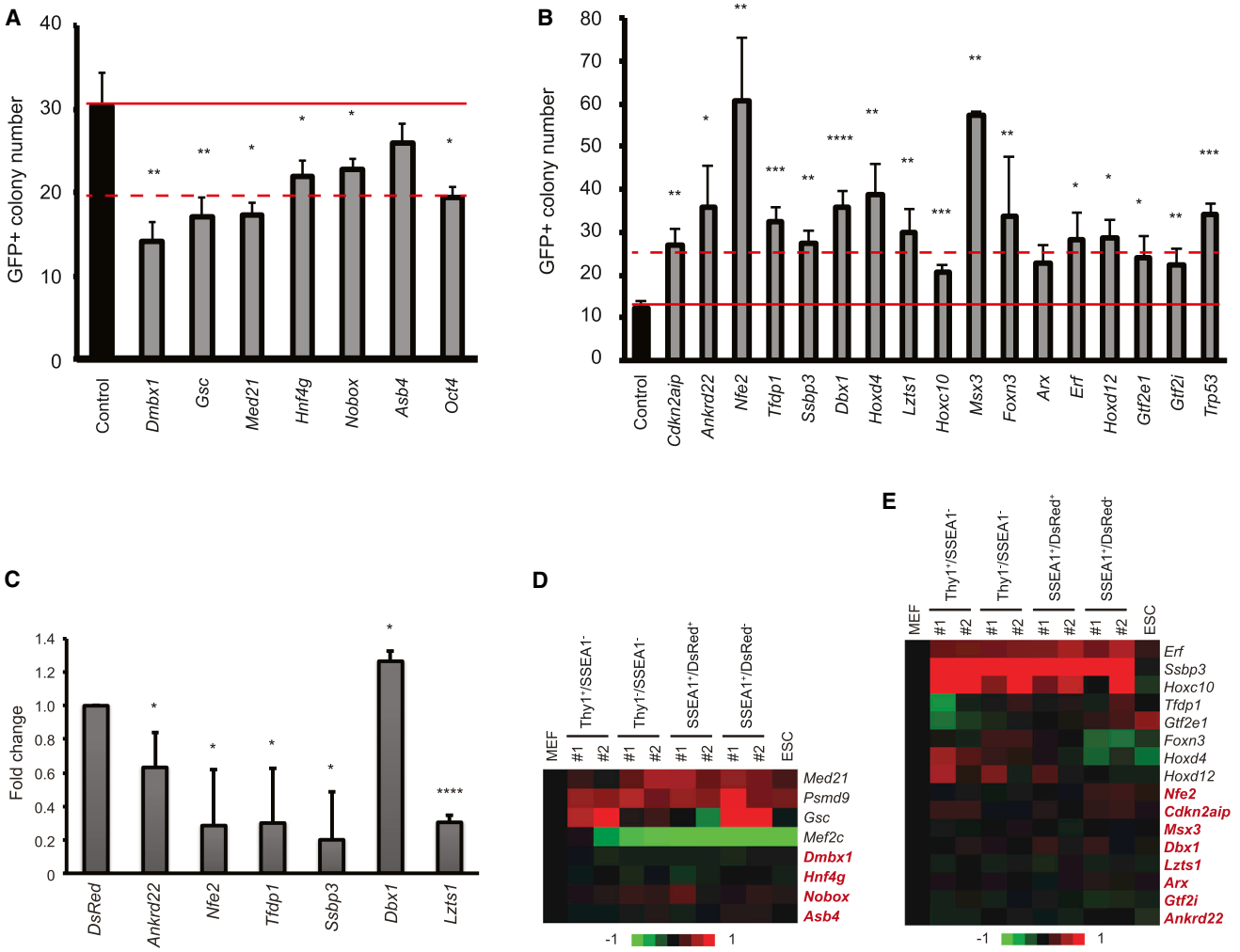



**Figure 4. Functional Validation Shows a High Discovery Rate in Identifying Positive Regulators or Barrier Genes for Reprogramming**

(A) Bar graph showing reprogramming efficiency following siRNA knockdown of positive regulators. Indicated siRNAs plus OSKM were introduced into 4 × $10^4$ cells of Oct4-EGFP MEFs, and colonies were scored for EGFP positivity. *Oct4* knockdown served as positive control. Nontargeting siRNA served as negative control (Control). Error bars represent SEM, $n \geqq 3$. The solid line marks the control value, and the dashed line shows the cutoff value based on *Oct4* knockdown. Student's t test, *p < 0.05; **p < 0.005.

(B) Bar graph showing MEF reprogramming efficiency following barrier gene depletion. Reprogramming efficiency was assayed as in (A). *Trp53* (*p53*) knockdown served as positive control. Nontargeting siRNA served as negative control (Control). Error bars represent SEM, $n \geqq 3$. The solid line marks the control value, and the dashed line marks the cutoff value of 2-fold changes. Student's t test, *p < 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 4 × $10^4$ cells of Oct4-EGFP MEFs and assayed as described above. Reprogramming efficiency was calculated following normalization to DsRed control. Error bars represent SEM, $n \geqq 3$. Student's t test, *p < 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 nos. 1 and 2. Non-differentially expressed genes are highlighted in red boldface text. Fold-change values are presented on a $\log_2$ 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 boldface text.

See also Figures S3, S4, and S5.

regulators (e.g., *Asb4, Gbx2, Gsc, Hnf4g,* and *Mef2c*) that play important functions in maintaining ESC identities. Collectively, our genome-wide RNAi screen has identified numerous 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 stepwise induced reprogramming by functional genomics. We dissected the regulatory networks, employing a pooled genome-wide shRNA library in a stepwise 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 unbiased functional insight into essential factors during each step of the reprogramming progress. The high validation rate of identified genes in this study suggests that our strategy is highly valuable for the discovery of key regulatory molecules and networks in the reprogramming process.

We found that the majority of transformed cells are "trapped" in the transition stage (Thy1$^-$/SSEA1$^-$), with divergent transcriptomes 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 readministration 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 pluripotent stem cells. However, these "transitioning" cells with high plasticity may provide a good starting point for various cell-fate interconversions.

A recent study using a similar approach (Polo et al., 2012) has shown that a majority of cells that even had a 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) delineating the stepwise 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 a 2-week reprogramming course, it is possible that there are cells present at different levels and stages of latencies in reprogramming, making it difficult to completely rule out the heterogeneity issues in our analysis.

In this study, we used DsRed expression as an indicator of "mature reprogrammed" cells. Although *Nanog*, *Sall4*, and *Essrrb* genes have been slightly activated in SSEA1$^+$/DsRed$^+$ cells, these genes are further induced to a higher expression level (Table S1), and ESC-like epigenetic regulation is resotred to silence the retroviral gene (pMX-*DsRed*) only in SSEA1$^+$/DsRed$^-$ cells. Therefore, we reason that pMX-*DsRed* silencing may provide a better definition of "mature reprogrammed" cells than activation of the *Oct4* gene, which has surprisingly been shown to 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 limitations could be overcome by using newer algorithms to design efficient shRNA libraries, CRISPR-Cas9 technologies, and a homogeneous reprogramming system (polycistronic expression or somatic cells harboring inducible reprogramming factors). Despite these caveats, we provide a proof of principle that an unbiased pooled RNAi screen can be used to 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). Here, we vigorously tested the hypothesis that the non-differentially expressed genes play important roles in directing cell-fate decisions. Our functional genomics approach shows that in addition to tissue-specific genes, many non-differentially expressed genes actually play important roles in cell-fate transition during reprogramming (Figure S6B). Thus, we suggest that studies such as ours that use genome-wide RNAi screening to define reprogramming mechanisms will have numerous applications in this field, such as providing novel approaches to small-molecule targeting, cell-fate manipulation, and progenitor derivation. More importantly, our work not only uncovers 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 MEF Derivation

Oct4-EGFP MEFs were derived from the mouse strain B6;129S4-Pou5f1$^{tm2(EGFP)Jae}$/J (The Jackson Laboratory, stock no. 008214) with the use of the protocol provided on the WiCell Research Institute website (http://www.wicell.org/). In brief, embryonic day 13.5 embryos were collected from time-mated pregnant female mice. Cells isolated from embryos were then tested for microbial contamination. All animal work was approved by the Sanford-Burnham institutional review board and was performed following Institutional Animal Care and Use Committee guidelines. Oct4-EGFP MEFs were maintained in MEF complete medium (Dulbecco's modified Eagle's medium with 10% fetal bovine serum, nonessential amino acids, and L-glutamine, without sodium pyruvate). Robustly growing cells (usually more than four passages) were used for induced reprogramming.

### FACS and Whole-Genome RNAi Screening

Cells were transduced with retroviruses containing pMXs-DsRed plasmids, harvested 3 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 3 days before introduction of the shRNA library and OSKM. Pseudoviruses expressing a pGIPz-shRNA library and pMXs-OSKM were generated in 293FT and Plate-E cells, respectively. Pseudoviruses were administered at days 0 and 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, Miltenyi Biotec GmbH). 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.

## ACCESSION NUMBERS

The GEO accession number for the mRNA microarray data reported in this paper is GSE59175.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.07.002.

## AUTHOR CONTRIBUTIONS

C-.S.Y. contributed to the concept and design; data collection, assembly, analysis, and interpretation; and manuscript writing. K-.Y.C. performed data analysis. T.M.R. contributed to the concept and design, data analysis and interpretation, manuscript writing, and financial support. All authors approved the final version of this manuscript.

## ACKNOWLEDGMENTS

We thank Dr. Tingting Du and members of the T.M.R. lab for their advice, helpful discussions, and support. We are grateful for the use of the following Sanford-Burnham Institute shared resource facilities: flow cytometry core facility for cell sorting experiments, genomics and informatics and data-management core facilities for array experiments and data analysis, the animal facility for chimera generation and mouse teratoma assays, 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 (AI41404, AI43198, and DA30199).

Received: January 2, 2014
Revised: June 29, 2014
Accepted: July 8, 2014
Published: July 17, 2014

## REFERENCES

Banito, A., Rashid, S.T., Acosta, J.C., Li, S., Pereira, C.F., Geti, I., Pinho, S., Silva, J.C., Azuara, V., Walsh, M., et al. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev. *23*, 2134–2139.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151–159.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Carey, B.W., Markoulaki, S., Hanna, J.H., Faddah, D.A., Buganim, Y., Kim, J., Ganz, K., Steine, E.J., Cassady, J.P., Creyghton, M.P., et al. (2011). Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell Stem Cell *9*, 588–598.

Choi, Y.J., Lin, C.P., Ho, J.J., He, X., Okada, N., Bu, P., Zhong, Y., Kim, S.Y., Bennett, M.J., Chen, C., et al. (2011). miR-34 miRNAs provide a barrier for somatic cell reprogramming. Nat. Cell Biol. *13*, 1353–1360.

González, F., Boué, S., and Izpisúa Belmonte, J.C. (2011). Methods for making induced pluripotent stem cells: reprogramming à la carte. Nat. Rev. Genet. *12*, 231–242.

Grskovic, M., Javaherian, A., Strulovici, B., and Daley, G.Q. (2011). Induced pluripotent stem cells—opportunities for disease modelling and drug discovery. Nat. Rev. Drug Discov. *10*, 915–929.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Hansson, J., Rafiee, M.R., Reiland, S., Polo, J.M., Gehring, J., Okawa, S., Huber, W., Hochedlinger, K., and Krijgsveld, J. (2012). Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. Cell Rep. *2*, 1579–1592.

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

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., et al. (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell *5*, 491–503.

Jopling, C., Boue, S., and Izpisua Belmonte, J.C. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nat. Rev. Mol. Cell Biol. *12*, 79–89.

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

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Izpisúa Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140–1144.

Kim, N.H., Kim, H.S., Li, X.Y., Lee, I., Choi, H.S., Kang, S.E., Cha, S.Y., Ryu, J.K., Yoon, D., Fearon, E.R., et al. (2011). A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. J. Cell Biol. *195*, 417–433.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96–105.

Kupershmidt, I., Su, Q.J., Grewal, A., Sundaresh, S., Halperin, I., Flynn, J., Shekar, M., Wang, H., Park, J., Cui, W., et al. (2010). Ontology-based meta-analysis of global collections of high-throughput public data. PLoS ONE *5*, e13066.

Li, M.A., and He, L. (2012). microRNAs as novel regulators of stem cell pluripotency and somatic cell reprogramming. BioEssays *34*, 670–680.

Li, Z., and Rana, T.M. (2012). A kinase inhibitor screen identifies small-molecule enhancers of reprogramming and iPS cell generation. Nat. Comm. *3*, 1085.

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

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51–63.

Li, Z., Yang, C.S., Nakashima, K., and Rana, T.M. (2011). Small RNA-mediated regulation of iPS cell generation. EMBO J. *30*, 823–834.

Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y., Cai, J., Guo, X., Qin, B., et al. (2011). MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. J. Biol. Chem. *286*, 17359–17364.

Lipchina, I., Elkabetz, Y., Hafner, M., Sheridan, R., Mihailovic, A., Tuschl, T., Sander, C., Studer, L., and Betel, D. (2011). Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. Genes Dev. *25*, 2173–2186.

Locasale, J.W. (2013). Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat. Rev. Cancer *13*, 572–583.

Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., and Pfaff, S.L. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. *25*, 594–607.

Maherali, N., and Hochedlinger, K. (2009). Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Curr. Biol. *19*, 1718–1723.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Marión, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA

damage response limits reprogramming to ensure iPS cell genomic integrity. Nature *460*, 1149–1153.

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

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215–3222.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602.

Pfaff, N., Fiedler, J., Holzmann, A., Schambach, A., Moritz, T., Cantz, T., and Thum, T. (2011). miRNA screening reveals a new miRNA family stimulating iPS cell generation via regulation of Meox2. EMBO Rep. *12*, 1153–1159.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Robinton, D.A., and Daley, G.Q. (2012). The promise of induced pluripotent stem cells in research and therapy. Nature *481*, 295–305.

Sakurai, K., Talukdar, I., Patil, V.S., Dang, J., Li, Z., Chang, K.Y., Lu, C.C., Delorme-Walker, V., Dermardirossian, C., Anderson, K., et al. (2014). Kinome-wide functional analysis highlights the role of cytoskeletal remodeling in somatic cell reprogramming. Cell Stem Cell *14*, 523–534.

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

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell *153*, 963–975.

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

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

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

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

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

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

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

Tiscornia, G., Vivas, E.L., and Izpisúa Belmonte, J.C. (2011). Diseases in a dish: modeling human genetic disorders using induced pluripotent cells. Nat. Med. *17*, 1570–1576.

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145–1148.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., et al. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618–630.

Wolf, D., and Goff, S.P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell *131*, 46–57.

Wu, S.M., and Hochedlinger, K. (2011). Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat. Cell Biol. *13*, 497–505.

Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. Nature *460*, 49–52.

Yang, C.-S., and Rana, T.M. (2013). Learning the molecular mechanisms of the reprogramming factors: let's start from microRNAs. Mol. Biosyst. *9*, 10–17.

Yang, C.S., Li, Z., and Rana, T.M. (2011a). microRNAs modulate iPS cell generation. RNA *17*, 1451–1460.

Yang, C.S., Lopez, C.G., and Rana, T.M. (2011b). Discovery of nonsteroidal anti-inflammatory drug and anticancer drug enhancing reprogramming and induced pluripotent stem cell generation. Stem Cells *29*, 1528–1536.

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

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

Yu, Y., Liang, D., Tian, Q., Chen, X., Jiang, B., Chou, B.K., Hu, P., Cheng, L., Gao, P., Li, J., et al. (2014). Stimulation of somatic cell reprogramming by Eras-Akt-Foxo1 signaling axis. Stem Cells *32*, 349–363.

Zhu, S., Wei, W., and Ding, S. (2011). Chemical strategies for stem cell biology and regenerative medicine. Annu. Rev. Biomed. Eng. *13*, 73–90.