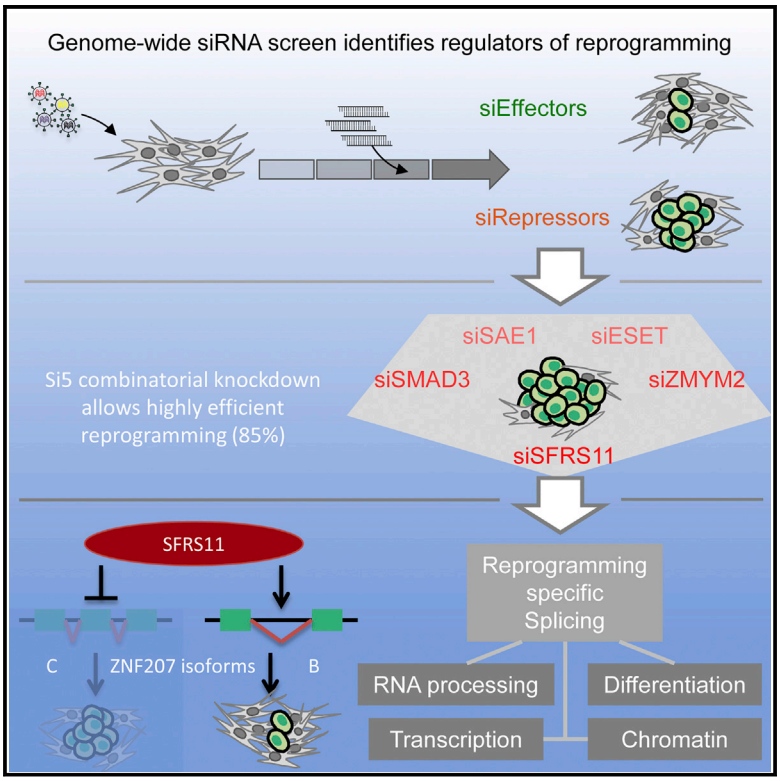
Report

Cell Reports

RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming

Graphical Abstract



Highlights

- Key functional regulators of human somatic cell reprogramming are identified
- Combinatorial knockdown of five repressors allows highly efficient reprogramming (85%)
- SFRS11 regulates splicing of genes that are critical for reprogramming
- Splicing of downstream *ZNF207* isoform B to A and C affects reprogramming efficiency

Authors

Cheng-Xu Delon Toh, Jun-Wei Chan, Zheng-Shan Chong, ..., Yoon-Pin Lim, Frederic A. Bard, Yuin-Han Loh

Correspondence

li.hu@mayo.edu (H.L.), bchlyp@nus.edu.sg (Y.-P.L.), yhloh@imcb.a-star.edu.sg (Y.-H.L.)

In Brief

Toh et al. identified genome-wide functional regulators of the early stages of human somatic cell reprogramming in a phase-specific RNAi screen. Combinatorial depletion of the top repressors allows reprogramming to proceed unhindered at a near-deterministic efficiency. SFRS11, an mRNA splicer, blocks reprogramming partially through the splicing of ZNF207 isoforms.

Accession Numbers

GSE68234

Toh et al., 2016, Cell Reports *15*, 2597–2607
June 21, 2016
http://dx.doi.org/10.1016/j.celrep.2016.05.049

Cell Reports
Report

RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming

Cheng-Xu Delon Toh,[1,2,24] Jun-Wei Chan,[1,24] Zheng-Shan Chong,[1,24] Hao Fei Wang,[1,3] Hong Chao Guo,[1,4] Sandeep Satapathy,[1] Dongrui Ma,[5] Germaine Yen Lin Goh,[6] Ekta Khattar,[7] Lin Yang,[8,9,10] Vinay Tergaonkar,[7,11] Young-Tae Chang,[12,13] James J. Collins,[10,14,15,16] George Q. Daley,[10,17,18,19,20] Keng Boon Wee,[21,22] Chadi A. EL Farran,[1,3] Hu Li,[23,*] Yoon-Pin Lim,[2,11,22,*] Frederic A. Bard,[6,11] and Yuin-Han Loh[1,2,3,*]

[1]Epigenetics and Cell Fates Laboratory, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore
[2]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 28 Medical Drive, Singapore 117456, Singapore
[3]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore
[4]College of Life Sciences, Nankai University, Tianjin 300071, China
[5]Research and Development Unit (RDU), National Heart Centre Singapore, 5th Hospital Drive, Singapore 169609, Singapore
[6]Membrane Traffic Laboratory, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore
[7]Division of Cancer Genetics and Therapeutics, Laboratory of NF-κB Signaling, A*STAR Institute of Molecular and Cell Biology, 61 Biopolis Drive Proteos, Singapore 138673, Singapore
[8]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[9]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA
[10]Howard Hughes Medical Institute, Boston, MA 02114, USA
[11]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117545, Singapore
[12]Lab of Bioimaging Probe Development, A*STAR Singapore Bioimaging Consortium (SBIC), 11 Biopolis Way Helios, Singapore 138673, Singapore
[13]Department of Chemistry and MedChem Program, Life Sciences Institute, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore
[14]Institute for Medical Engineering and Science Department of Biological Engineering, and Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[15]Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA
[16]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA
[17]Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Boston, MA 02115, USA
[18]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[19]Harvard Stem Cell Institute, Boston, MA 02115, USA
[20]Manton Center for Orphan Disease Research, Boston, MA 02115, USA
[21]A*STAR Institute of High Performance Computing (IHPC), Connexis, Singapore 138632, Singapore
[22]A*STAR Bioinformatics Institute, Singapore 138671, Singapore
[23]Center for Individualized Medicine, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA
[24]Co-first author
*Correspondence: li.hu@mayo.edu (H.L.), bchlyp@nus.edu.sg (Y.-P.L.), yhloh@imcb.a-star.edu.sg (Y.-H.L.)
http://dx.doi.org/10.1016/j.celrep.2016.05.049

SUMMARY

Incomplete knowledge of the mechanisms at work continues to hamper efforts to maximize reprogramming efficiency. Here, we present a systematic genome-wide RNAi screen to determine the global regulators during the early stages of human reprogramming. Our screen identifies functional repressors and effectors that act to impede or promote the reprogramming process. Repressors and effectors form close interacting networks in pathways, including RNA processing, G protein signaling, protein ubiquitination, and chromatin modification. Combinatorial knockdown of five repressors (*SMAD3*, *ZMYM2*, *SFRS11*, *SAE1*, and *ESET*) synergistically resulted in ~85% TRA-1-60-positive cells. Removal of the novel splicing factor SFRS11 during reprogramming is accompanied by rapid acquisition of pluripotency-specific spliced forms. Mechanistically, SFRS11 regulates exon skipping and mutually exclusive splicing of transcripts in genes involved in cell differentiation, mRNA splicing, and chromatin modification. Our study provides insights into the reprogramming process, which comprises comprehensive and multi-layered transcriptional, splicing, and epigenetic machineries.

INTRODUCTION

The overexpression of the transcription factors (TFs) OCT4, SOX2, KLF4, and c-MYC (OSKM) has been shown to convert

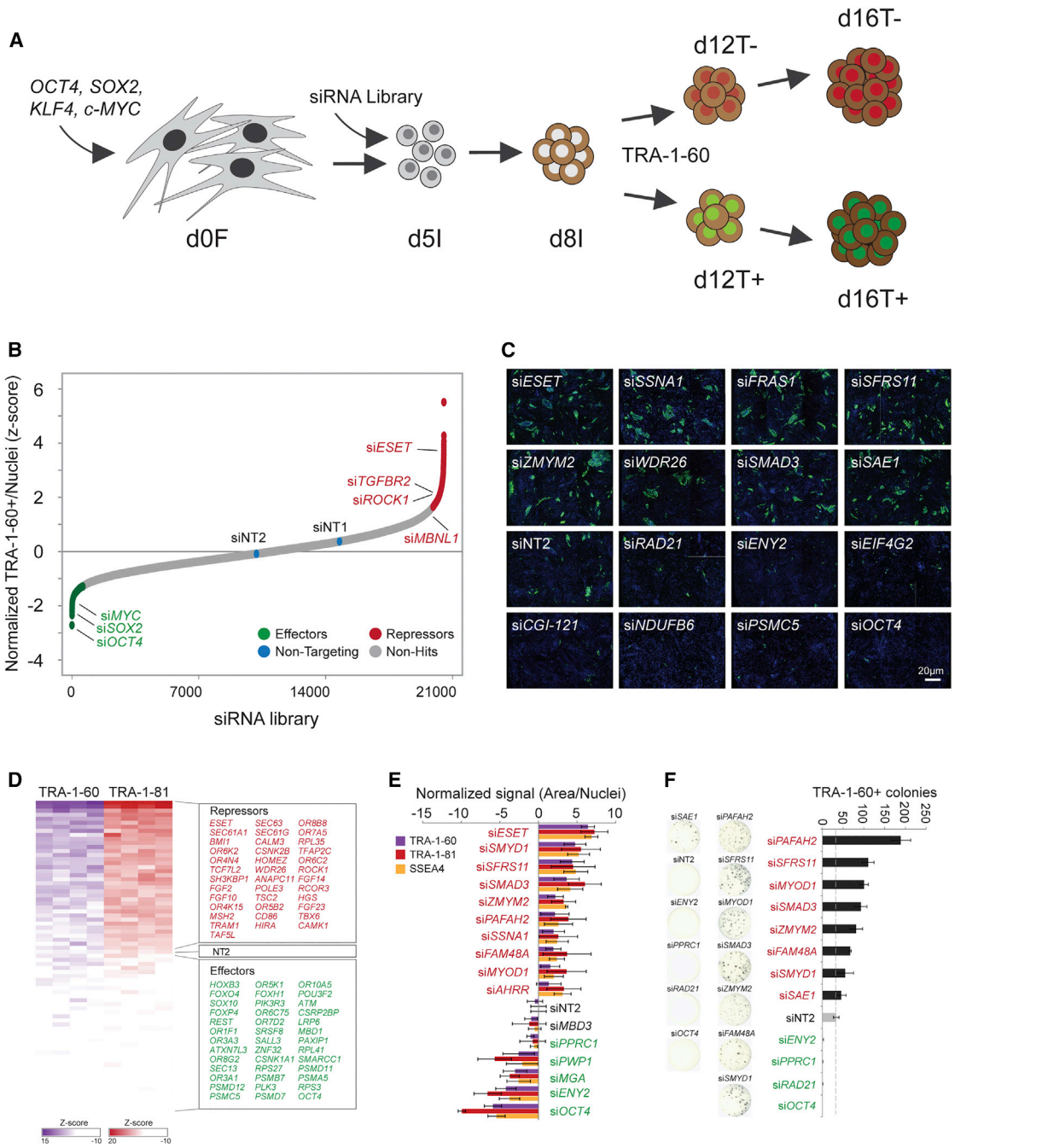


Figure 1. Genome-wide siRNA Screen Identifies Regulators of Human Reprogramming

(A) Schematic of reprogramming timeline and temporal cell samples collected for analysis. Reprogramming intermediates on 5 d.p.i. (d5I), 8 d.p.i. (d8I), 12 d.p.i., and 16 d.p.i. were harvested for mRNA-seq. 12 d.p.i. and 16 d.p.i. cells were sorted by the TRA-1-60 marker into positive (d12T+, d16T+) and negative (d12T−, d16T−) populations.

(B) Ranked distribution plot for the genome-wide siRNA screen. 599 genes with *Z* scores of >1.65 are considered potential candidates that are repressor of reprogramming (red dots). 557 genes with *Z* scores of < −1.3 are considered potential candidate effectors of reprogramming (green dots). The rest of the genes are indicated as gray dots.

(C) Representative images of TRA-1-60 (green) and DNA (blue) for selected hits identified from the genome-wide screen.

(D) Secondary screen of 76 genes with a second marker for pluripotency (TRA-1-81). Each row represents the knockdown of a single repressor, effector or NT control. Each column represents individual replicate experiments.

(legend continued on next page)

somatic cells into induced pluripotent stem cells (iPSCs) (Park et al., 2008; Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). Biochemical and genetic dissection of the molecular mechanisms underlying TF-induced reprogramming has been hampered by its low efficiency, slow kinetics, and the fact that it proceeds via heterogeneous intermediate cell populations (Polo et al., 2012; Stadtfeld and Hochedlinger, 2010). Nonetheless, recent studies have indicated that reprogramming is a stepwise process with at least three distinct phases: initiation, maturation, and stabilization (Samavarchi-Tehrani et al., 2010), with each phase marked by unique epigenomic and transcriptomic changes (Buganim et al., 2012; Polo et al., 2012). Several key cellular events, such as cell-cycle modulations (Kawamura et al., 2009; Qin et al., 2012; Zhao et al., 2008) and mesenchymal-to-epithelial transitions (Li et al., 2010; Samavarchi-Tehrani et al., 2010), were reported to influence reprogramming efficiency. In addition, the activity of epigenetic regulators, such as H3K9 methyltransferases (Chen et al., 2013) (EHMT1, SETDB1) and a H3K79 methyltransferase (Onder et al., 2012) (DOT1L), were found to limit the efficiency of reprogramming.

Recent attempts to attain a comprehensive overview of players involved in reprogramming utilized a pooled small hairpin RNA (shRNA) approach to conduct functional genome-wide screens in both mouse (Yang et al., 2014) and human (Qin et al., 2014) fibroblast lines. However, the constitutive knockdown of targeted genes by shRNA renders such screens unsuitable in the identification of key regulators affecting only a specific stage of the multiphasic reprogramming process (Buganim et al., 2012; Polo et al., 2012; Samavarchi-Tehrani et al., 2010).

To systematically identify genetic factors involved in early-phase reprogramming, we developed a high-throughput reprogramming assay adapted to a 384-well format for genome-wide small interfering RNA (siRNA) screening. This screening strategy allowed for transient knockdowns within a specific reprogramming phase, allowing us to identify phase-specific factors whose effects might be otherwise masked due to constitutive expression of the integrated shRNA.

RESULTS

Systematic Genome-wide siRNA Screen Identifies Functional Effectors and Repressors of Reprogramming

To identify the functional switches of early reprogramming, siRNA-mediated gene knockdown was performed on day 5 postinfection (5 d.p.i.) with lentiviral OSKM (Figure 1A). ESET (Chen et al., 2013) and OCT4 were selected as controls to optimize an unbiased genome-wide siRNA screen (Figure S1A). A preliminary screen for 972 human kinases and phosphatases was performed to demonstrate the robustness and reproducibility of the assay (Figures S1B–S1E; Table S1). We then proceeded with the genome-wide screen, targeting 21,121 human genes (Figures 1B, 1C, and S1F). The average Z' factor from two biological replicates of the genome-wide screen was 0.35 for si*ESET* and 0.37 for si*OCT4*. It was reassuring that our screen detected *OCT4* and *ESET* as the top effector and repressor of reprogramming, respectively (Figure 1B). Additionally, we identified several previously reported regulators, including *TGFBR2* (Ichida et al., 2009; Lin et al., 2009; Shi et al., 2008), *ROCK1* (Lai et al., 2010; Subramanyam et al., 2011), *MBNL1* (Han et al., 2013), *SALL4* (Tsubooka et al., 2009), c-*MYC* (Samavarchi-Tehrani et al., 2010), and *SOX2* (Samavarchi-Tehrani et al., 2010) (Figure 1B). Based on a targeted error rate of 0.05, a cutoff of +1.65 and −1.3 was used to identify 599 putative repressors and 557 putative effectors (Figures 1B and 1C; Table S1) (Birmingham et al., 2009). Gene Ontology (GO) analysis of these candidate repressors and effectors showed enrichment of multiple biological processes and pathways (Figures S1G and S1H). Next, we randomly selected 76 putative regulators for secondary screens. Selected regulators consistently reproduced their reprogramming phenotype when an independent pluripotent marker (TRA-1-81) was utilized (Figures 1D and S1I). Based on the *Z* score obtained from the genome-wide screen, we validated the top ten repressors and top five effectors in a different retroviral reprogramming system (Figure 1E). These regulators consistently affected reprogramming efficiencies when the assay was extended to 21 d.p.i. (Figure 1F). Pooled siRNA for the top repressors were then deconvoluted into the individual siRNAs and each recapitulated the effect on reprogramming (Figure S1J). To identify bona fide regulators of reprogramming, the *Z* score was normalized to nuclei number to remove the effect of proliferation on assay readout. Unlike the knockdown of *P53*, siRNA of *SAE1*, *SFRS11*, *ZMYM2, SMAD3*, and *ESET* did not significantly affect cell proliferation (Figures S1K and S1L).

Combinatorial Removal of Repressors Allows for High Reprogramming Efficiencies

The strongest repressors—namely *SMAD3* (transforming growth factor β [TGF-β] signal transducer), *ZMYM2* (epigenetic modifier), *SFRS11* (putative splicing factor), *SAE1* (Sumo-activating enzyme), and *ESET* (H3K9 methyltransferase) —function in distinct pathways. Therefore, we hypothesize that each factor is an independent barrier against reprogramming. As expected, the pairwise knockdown of two repressors had an additive effect on reprogramming enhancement (Figure 2A). We then pooled the siRNAs against repressors in groups of five (*ESET*, *ZMYM2*, *SAE1*, *SFRS11*, and *SMAD3*) or eight (*si5*, *PAFAH2*, *FRAS1*, and *SSNA1*). The combinatorial knockdown of five (si5) or eight (si8) repressors enhanced reprogramming several fold as compared to the removal of individual repressors (Figures 2B and 2C). In addition, si5 depletion enhanced the formation of iPSCs in multiple reprogramming systems and parental cell lines (Figures S2A–S2C). As DOT1L and TGF-β signaling pathways

(E) Validation of candidate genes in a retroviral reprogramming system. The area of each marker stained was normalized to cell number, and the *Z* score was obtained with reference to a siNT2 negative control. n = 4; error bar indicates SD.
(F) Validation of selected candidates assayed at 21 d.p.i. Left: representative images of TRA-1-60 staining. n = 3; error bar indicates SD. Right: graph of the number of colonies stained with TRA-1-60. The gray dashed line indicates the reprogramming efficiency of negative control (siNT). n = 3; error bar indicates SD.

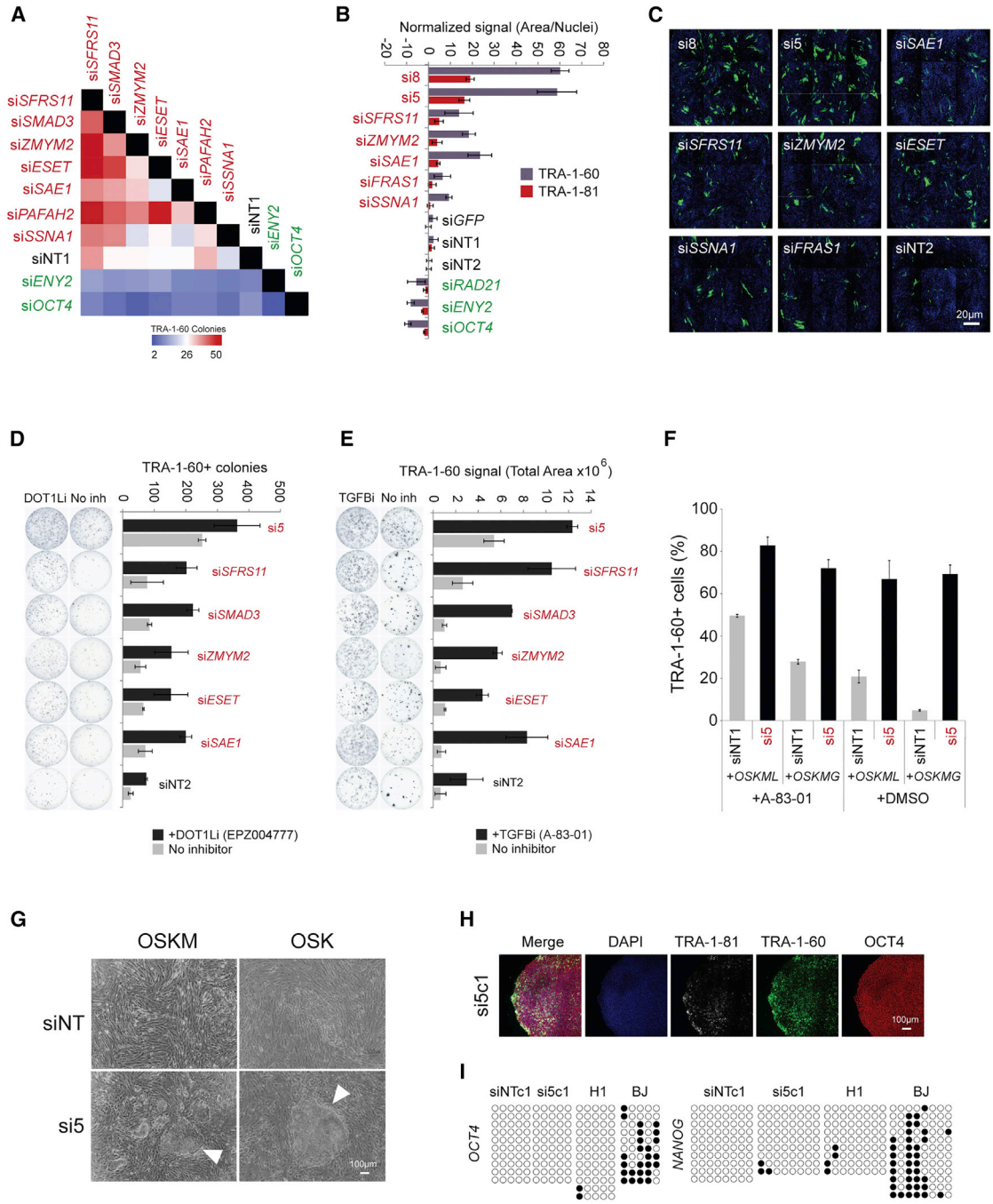


Figure 2. Combinatorial Knockdown of Repressors Synergistically Increases Reprogramming Efficiency

(A) Heatmap of the TRA-1-60 signal at 12 d.p.i., when two candidates were depleted simultaneously. n = 8.

(B) Removal of multiple repressors allows high reprogramming efficiency. Reprogramming was assayed at 16 d.p.i. All combinations of siRNA were reverse transfected with equimolar amounts of different siRNA. n = 8; error bars indicate SD.

(C) Representative images of TRA-1-60 (green) and DNA (blue) for the knockdown of repressors singly or in combination.

(D) Repressors function independently from the DOT1L pathway. Left: representative images of TRA-1-60+ colonies on 12-well plates after knockdown of the indicated genes. Right: bar graph indicating the number of TRA-1-60+ colonies obtained at 21 d.p.i. n = 3; error bar indicates SD.

(E) Repressors function independently of the TGF-β pathway. Left: representative images of TRA-1-60+ colonies on 12-well plates after knockdown of the indicated genes. Right: bar graph indicating the area of TRA-1-60 staining obtained at 21 d.p.i. n = 3; error bar indicates SD.

(F) si5 in combination with OSKML (LIN28) induction and TGF-β inhibition increase reprogramming yield. Bar graph of the percentage of cells that stained positive for TRA-1-60 at 28 d.p.i. The highest reprogramming efficiency (82.75%) was observed with the inhibition of TGF-β and si5, with the simultaneous overexpression of LIN28. n = 3; error bar indicates SD.

(legend continued on next page)

were previously implicated as critical barriers of reprogramming, we examined the effect of inhibiting these pathways upon si5 depletion. The inhibition of DOT1L and TGF-β signaling further enhanced reprogramming, suggesting that the repressors' effect on reprogramming is at least partially independent of DOT1L and TGF-β signaling (Figures 2D and 2E). The combined inhibition of DOT1L or TGF-β with si5 allowed most cells to undergo homogenous morphological changes, resulting in rapid development of reprogrammed colonies (Figures 2D, 2E, and S2D). The ectopic expression of OSKM+L (LIN28), coupled with the depletion of si5 and TGF-β signaling, resulted in the generation of ~85% TRA-1-60+ cells at 28 d.p.i. (Figures 2F and S2E). Moreover, si5 depletion could replace c-*MYC* in the reprogramming cocktail (Figure 2G). iPSC lines obtained from si5 knockdown share similar characteristics with human embryonic stem cells (hESCs) and have normal chromosomal karyotypes (Figures 2H, 2I, and S2F–S2I). Taken together, these data demonstrate that the simultaneous removal of these repressors allows reprogramming to proceed in a more deterministic manner.

Knockdown of Regulators Affects Dynamics of Reprogramming

To understand the underlying mechanisms of the identified reprogramming regulators, we performed mRNA sequencing (mRNA-seq) on cells undergoing reprogramming (Figure 1A; Table S2). Differential GO analysis shows the shift of biological functions associated with fibroblasts toward processes that define pluripotency (Figure 3A). Additionally, transcriptome rewiring is accompanied by a progressive epigenetic shift toward bivalency on lineage-associated genes and the activation of pluripotency genes (Figures S3A and S3B). Gene set enrichment analysis (GSEA) and principal-component analysis (PCA) illustrate the progressive shift toward the iPSC fate (Figure S3C).

To determine if the knockdown of repressors accelerates reprogramming, we examined the kinetics of SSEA4 and TRA-1-60 expression. The knockdown of repressors (siRepressor) accelerated the detection of TRA-1-60 to as early as 8 d.p.i. (Figures 3B and S3D). Correspondingly, siRepressor accelerated the acquisition of a transcriptome profile closer to that of more advanced reprogrammed cells (Figures 3C–3E). siRepressor consistently resulted in higher expression of pluripotency-associated genes and lower expression of fibroblast-associated genes at earlier time points (Figures 3D–3F and S3E). The repression of epithelial-to-mesenchymal transition (EMT)-associated genes and activation of mesenchymal-to-epithelial transition (MET)-associated genes were accelerated in siRepressor samples (Figures 3E, 3F, and S3E). The differential expression analysis of the selected regulators resulted in transcriptome profiles that are unique to each knockdown condition (Figures S3F and S3G). This affirms the presence of distinctive mechanistic effects by which these regulators drive or deter reprogramming.

SFRS11 Regulates Alternative Splicing during Reprogramming

Within the si5 combination, si*SFRS11* contributed most of the increased reprogramming efficiency (Figures S4A and S4B). Moreover, SFRS11 is one of the strongest repressors (Figures 1E, 1F, and 2B–2E) in the majority of validation assays. We therefore focused on deciphering the mechanism through which SFRS11 represses reprogramming. siSFRS11-derived iPSC clones had characteristics similar to hESC lines (Figures S4C–S4E) and the capacity to form teratomas in immunodeficient mice (Figure S4F). Interestingly, *SFRS11* depletion had no effect on pluripotency (Figures S4G and S4H) (Chia et al., 2010). These data suggest that SFRS11 specifically regulates the acquisition, but not maintenance, of pluripotency. Additionally, clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-mediated knockout of *SFRS11* also enhanced reprogramming efficiency, further confirming the repressive effects of SFRS11 on reprogramming (Figure S4I).

To gain mechanistic insights into the role of SFRS11, we analyzed deep mRNA-seq of the *SFRS11*- and *SMAD3*-depleted (non-splicing repressor control) reprogramming cells at 8 d.p.i. Exon inclusion and de novo transcript assembly analyses showed that *SFRS11* knockdown resulted in drastic changes in differential isoform expression (Figures 4A, 4B, and S4J). Reads were mapped to exon-splice junction sites and analyzed to identify different splicing events. Depletion of *SFRS11* resulted in a change in 696 splicing events (489 genes), whose inclusion ratio indicated significant splicing (false discovery rate [FDR] < 0.01) when compared to the non-targeting siRNA (siNT) control (Table S3). In contrast, *SMAD3* knockdown resulted in a change of 43 splicing events (33 genes, FDR < 0.01). SFRS11 regulates the alternative splicing of genes that were enriched with several biological process, including cell differentiation, cell proliferation, programmed cell death, mRNA splicing, and chromatin modifications (Figure 4C).

Out of these splicing events, four splicing patterns can be observed (Figures 4D, S4K, and S4L). Exon skipping (62.4%) and mutually exclusive (16.4%) events account for the majority of events detected upon *SFRS11* depletion (Table S3). The depletion of *SFRS11* predominantly increased the inclusion of skipped exons (Figure 4E) and of downstream mutually exclusive exons (Figure 4F). To validate these alternatively spliced events, we performed PCR for regions flanking skipped exons or mutually exclusive exons. Gel electrophoresis and band intensity analysis enable us to obtain the relative proportion of each alternatively spliced exon that differed in size. We validated four exon-skipping and four mutually exclusive splicing events at different time points of the reprogramming process. Interestingly, *SFRS11* depletion resulted in splicing isoforms that are closer to the pluripotent state (Figures 4G and S4M). These

(G) Combinatorial depletion of si5 can replace *c-MYC* in the reprogramming cocktail. Bright-field images of reprogramming BJ fibroblasts at 28 d.p.i. are shown. White arrows indicate forming colonies.

(H) Antibody staining of the si5c1 iPSC line shows that they express markers of pluripotent stem cells, namely TRA-1-81 (white), TRA-1-60 (green), OCT4 (red), and DNA (blue).

(I) Bisulfite DNA methylation analysis on the promoter of *OCT4* and *NANOG* in si5c1 iPSCs. siNTc1 iPSC cell line, parental BJ fibroblasts, and H1 hESCs were used as controls. The white and black circles represent unmethylated and methylated CpG islands in the promoter region, respectively.

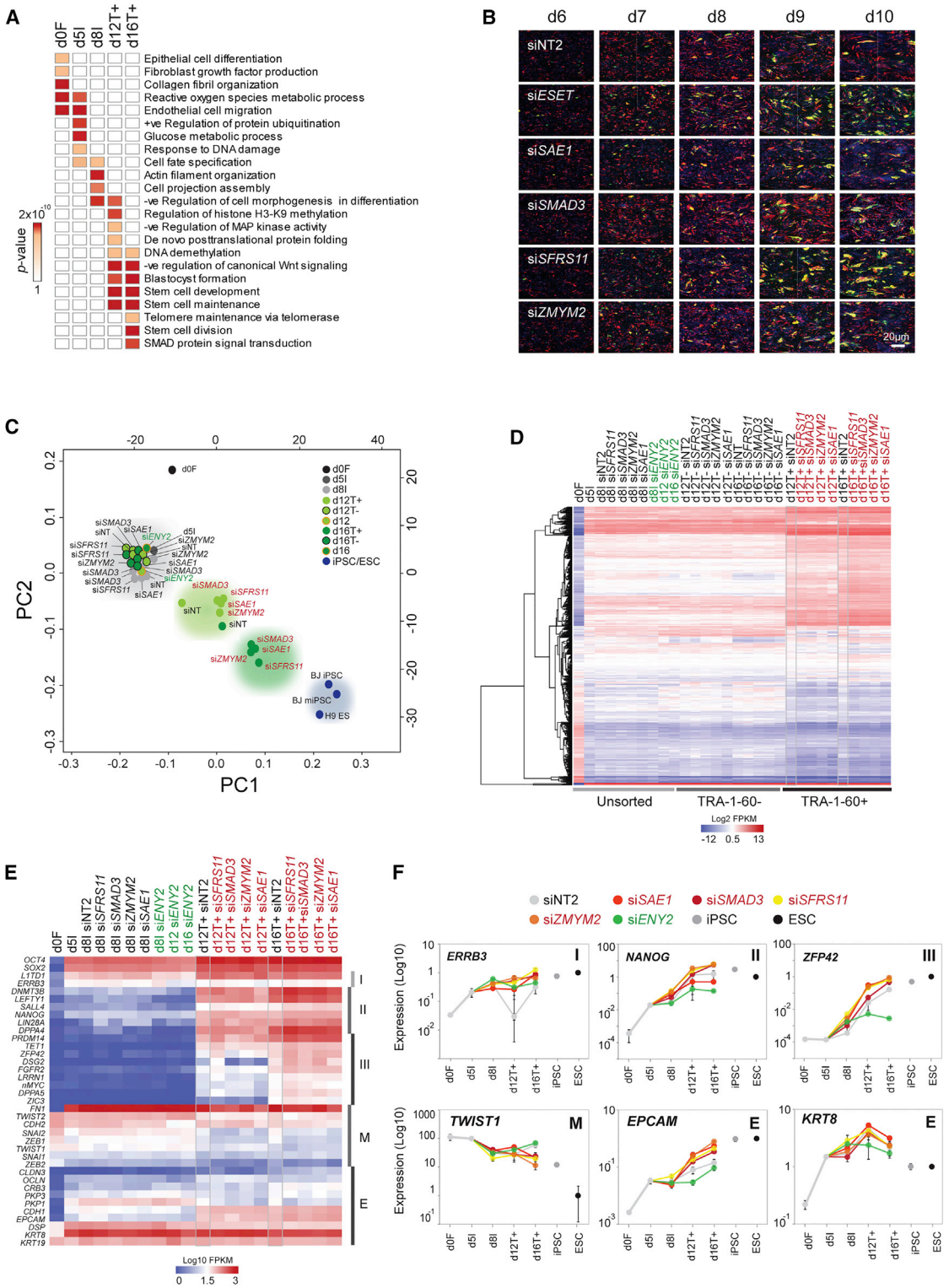


Figure 3. Perturbation of Repressors and Effectors Affects Reprogramming Kinetics

(A) Differential GO analysis for different stages of reprogramming. Darker red represents higher significance.

(B) siRepressor knockdown accelerate reprogramming kinetics. Immunostaining of TRA-1-60 (green), SSEA4 (red), and DNA (blue) at different days of reprogramming as indicated.

(legend continued on next page)

results suggest that specific splicing events could repress or promote reprogramming.

SFRS11 Regulates Splicing of Reprogramming Effector *ZNF207*

From exon-skipping events, we identified *ZNF207* to be a putative downstream splicing target of SFRS11 during the early phase of reprogramming (Figures 4E and 4G). The mouse homolog *zfp207* was reported to undergo alternative splicing during mouse reprogramming (Ohta et al., 2013). We first validated that the splicing of *ZNF207* was specific to *SFRS11* depletion and not due to the accelerated reprogramming kinetics (Figure 4H). Depletion of *SFRS11* increased levels of *ZNF207* isoform A and C transcripts while decreasing isoform B transcripts, corresponding to the increased retention of exon 9 (Figure 4H).

To understand the interaction between SFRS11 and *ZNF207*, we deleted the two functional domains of SFRS11, namely the RNA-recognition motif (RRM) that confers the ability to bind to single-stranded RNA and an arginine-serine rich domain (RS) that facilitates protein-protein interactions (Lunde et al., 2007) (Figures S4N and S4O). Consistent with the repressive phenotype observed from siRNA-mediated knockdown, overexpression of wild-type (WT) SFRS11 dramatically repressed reprogramming ($p < 0.05$). In contrast, mutation of either the RRM (ΔRRM) or the RS (ΔRS) domain partially rescued the repressive effects of SFRS11 (Figure 4I). We then asked if *ZNF207* transcripts were directly regulated by SFRS11 by performing ultraviolet crosslinking coupled to immunoprecipitation and qPCR (CLIP-qPCR) in cells undergoing reprogramming (8 d.p.i.). We confirmed that SFRS11 binds directly to endogenous *ZNF207* transcripts via its RRM domain (Figures 4J, 4K, and S4P). Detection of *ZNF207* transcripts was only possible when antibody against WT SFRS11 was used in the immunoprecipitation (Figures 4J and 4K).

We next asked if the overexpression of individual *ZNF207* isoforms affects reprogramming efficiency. Interestingly, the overexpression of *ZNF207* isoforms A and C consistently enhanced reprogramming efficiency by 1.12- and 1.5-fold, respectively ($p < 0.05$), while overexpression of the *ZNF207* isoform B decreased reprogramming efficiency to 0.7-fold ($p < 0.05$) (Figure 4L). The knockdown of all three *ZNF207* isoforms using siRNA decreased overall reprogramming efficiency dramatically (Figure S4Q). We therefore designed steric hindrance antisense oligonucleotides (AONs) to induce specific skipping of either *ZNF207* exons 6 or 9 (Figure 4H), resulting in the specific switching of *ZNF207* isoform C to A (AON 1 and 2) or switching of *ZNF207* isoform A and C to B (AON 3–5) (Figures S4R and S4S; Table S4). The introduction of these AONs reduced *ZNF207* isoform C (Figures S4R and S4S) during reprogramming, resulting in a decrease in reprogramming efficiency ($p < 0.001$). As AON 3–5 promotes the switching of *ZNF207* isoform C to isoform B, the increased of *ZNF207* isoform B expression further reduces reprogramming efficiency as compared to AON 1 and 2 ($p < 0.001$) (Figures S4R and S4S). These data support the results of the overexpression studies, showing that an increase in *ZNF207* isoforms A and C is crucial for reprogramming whereas an increase in *ZNF207* isoform B hinders reprogramming.

DISCUSSION

Somatic cell reprogramming proceeds through ordered and distinct phases of initiation, maturation, and stabilization (Golipour et al., 2012; Samavarchi-Tehrani et al., 2010). We have developed a sensitive high-throughput screen assay, which enabled us to systematically unravel 599 repressors and 557 effectors of the early phase of reprogramming, including established and previously undescribed regulators (Onder et al., 2012; Qin et al., 2014; Rais et al., 2013; Samavarchi-Tehrani et al., 2010; Yang et al., 2014). Expectedly, the importance of OSM during early reprogramming was highlighted by our screen, consistent with a previous focused screen performed in the mouse system (Samavarchi-Tehrani et al., 2010). Interestingly, analysis of mRNA expression indicates that a significant proportion of reprogramming regulators remain unchanged during reprogramming (Table S2). This supports a previous study reporting that ~53% to 70% of the identified regulators remain transcriptionally stable during mouse reprogramming (Yang et al., 2014).

Among the repressors identified, the effect on reprogramming was strongest when *SMAD3*, *SFRS11*, *ZMYM2*, *ESET*, or *SAE1* were depleted. These repressors represent functionally distinct regulators of unique pathways. The simultaneous removal of these repressors had a cumulative and maximal effect on improving reprogramming efficiency and kinetics. In combination with the overexpression of LIN28 and small-molecule inhibition of TGF-β, up to ~85% TRA-1-60+ cells can be detected at 28 d.p.i. Thus, we established that the simultaneous elimination of these functionally distinct repressors allows reprogramming to proceed synchronously in a near-deterministic manner with accelerated kinetics.

Temporal proteomic profiling of cells undergoing reprogramming indicated the robust upregulation of RNA processing factors during the early phase of reprogramming (Hansson et al., 2012). Our genome-wide screen indicates that RNA processing factors could functionally repress reprogramming (Figure S1H). Among these RNA-binding proteins, SFRS11 was the strongest repressor. Analysis of deep mRNA-seq samples at 8 d.p.i.

(C) Principal-component analysis (PCA) of RNA-sequencing libraries (refer to Table S2). The knockdown of repressor genes accelerated the switch to a pluripotent state. Day 12 and 16 samples are sorted with TRA-1-60.

(D) Dynamic expression of genes that are differentially expressed between fibroblasts and H1 hESCs during reprogramming. d12T+ and d16T+ siRepressor samples demonstrate a higher rate of downregulation of fibroblast-associated genes. siNT2 samples are boxed (light gray).

(E) Heatmap demonstrating dynamic expression of the indicated genes clustered into categories consisting of genes expressed at the early (I), intermediate (II), or late stages (III) of reprogramming. Genes belonging to EMT (M) and MET (E) are also indicated. siNT2 samples are boxed (light gray).

(F) qRT-PCR quantification of pluripotency (*ERRB3*, *NANOG*, *ZFP42*), mesenchymal (*TWIST1*), and epithelial (*EPCAM*, *KRT8*) markers. Data shown are normalized to H1 hESCs. n = 2; error bars indicate SD.

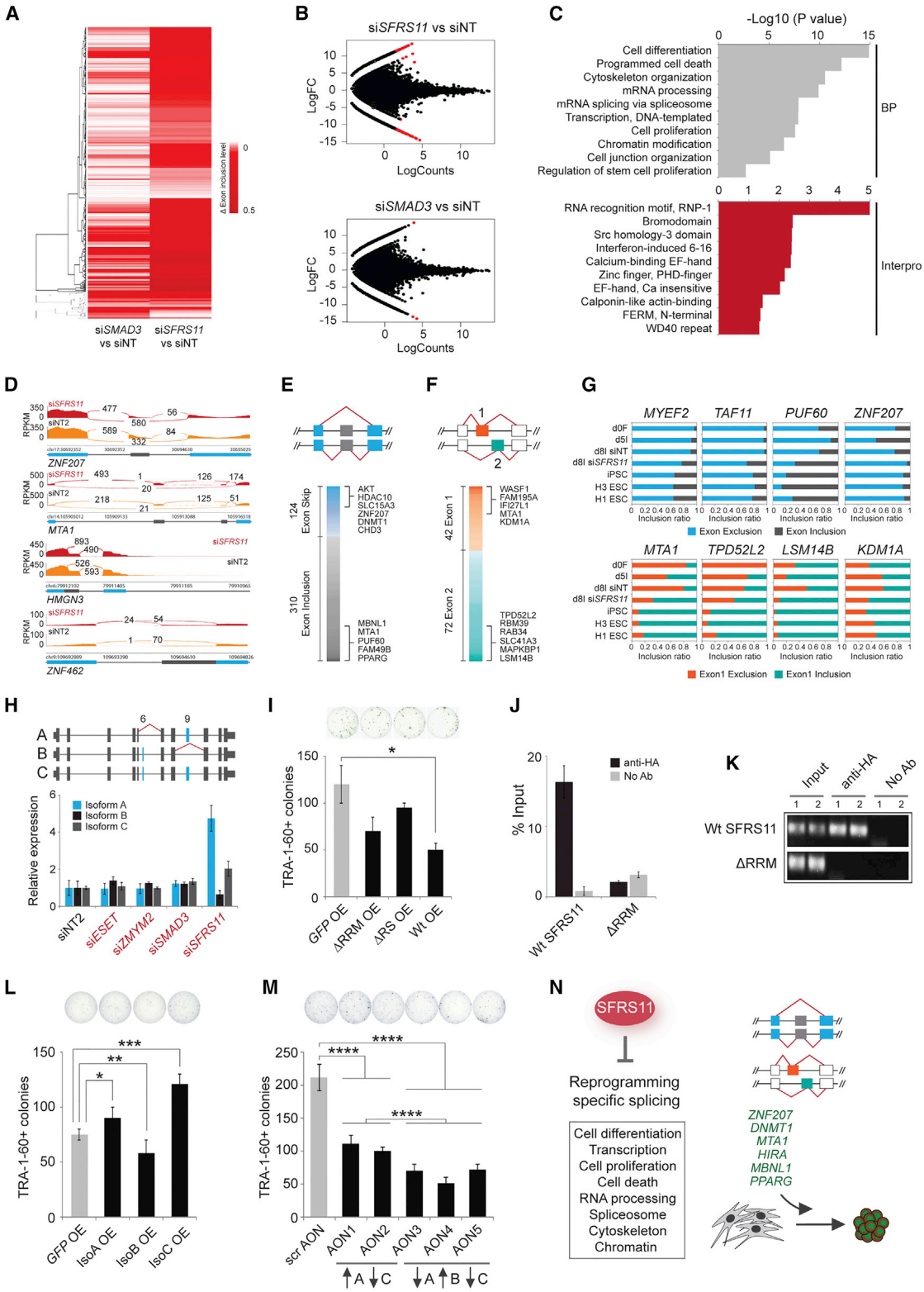


Figure 4. SFRS11 Regulates the Splicing of the Transcripts for Reprogramming Effector *ZNF207*

(A) Difference in exon inclusion levels of significant differential splicing events between si*SFRS11* versus siNT and si*SMAD3* versus siNT. Heatmap color scale ranges from white (no change in inclusion level) to red (large difference in inclusion level).

(legend continued on next page)

revealed that 489 genes are alternatively spliced when *SFRS11* was transiently depleted (Table S3). Interestingly, these genes were enriched for biological functions associated with mRNA splicing (*MBNL1*, *SFRS3*, *U2AF1*, *SFRS2*, and *PUF60*). MBNL1 was previously shown to regulate the splicing of a pluripotency switch (*FOXP1*) in hESCs (Han et al., 2013) and was also identified as repressor from our genome-wide screen (Figure 1B). The mouse homologs for SFRS3 and U2AF1 were previously identified as regulators of mouse reprogramming (Ohta et al., 2013). This strongly suggests that SFRS11 could regulate reprogramming through the splicing of downstream pluripotent-specific splicing factors (Figure 4N).

Our study identified *ZNF207* as one of the effectors of human reprogramming. *SFRS11* depletion during reprogramming induces the isoform switch of *ZNF207* B to A and C. The identification of *ZNF207* and other SFRS11-regulated splicing events highlights an extensive role for the alternative splicing network during reprogramming (Figure 4N). While SFRS11 repression enhances reprogramming by at least 5-fold (Figure 1F), overexpression of ZNF207 isoform C enhances reprogramming by 1.5-fold. It is highly possible that other SFRS11-regulated splicing events are contributing to its role as a reprogramming repressor. The functional significance of other SFRS11-regulated splicing variants remains to be validated in the reprogramming context.

Our study demonstrates that the regulation of a multiphasic process, such as reprogramming, involve distinct regulators that are involved in multiple pathways. Together, our systematic and transient siRNA screen provides a wealth of information about the functional regulators required specifically at the early phase of reprogramming. These regulators may not have been uncovered with the complete depletion of gene function using gene knockout or constitutive shRNA knockdown (Qin et al., 2014; Yang et al., 2014). Taken together, these findings highlight the need to transiently generate hypomorphic states of gene expression in order to functionally dissect multi-phasic somatic cell reprogramming.

EXPERIMENTAL PROCEDURES

Transfection, Staining, and Imaging in 384-Well Plates

Plates (384-well, Grenier) were coated with 15 μl Matrigel (BD) overnight at 37°C. siRNA (2.5 μl 500 nM siGENOME, Dharmacon) was added to the plates, which were then frozen at −20°C before use. For reverse transfection, a master mix of 0.04 μl DharmaFECT1 (Dharmacon) transfection reagent and 7.46 μl OptiMEM (Invitrogen) mix was added to each well of the 384-well plates containing the siRNA. 5 days post-OSKM, the DharmaFECT1 and siRNA mix was incubated for 20 min before 4,000 BJ fibroblasts were seeded in each well in 40 μl MEF medium.

For the genome-wide screen, the whole-genome Dharmacon SMARTpooled siRNA library targeting 21,121 human genes was printed on 67 Matrigel-coated 384-well plates, where each well contained a mixture of four siRNAs targeting a single gene. On each plate, we included negative controls (non-target siRNA) and positive controls (*ESET* siRNA and *OCT4* siRNA) in the designated wells. These controls were chosen because they are known to play an important role in somatic cell reprogramming; OCT4 (Park et al., 2008; Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007) is indispensable in most reprogramming assays, and ESET (Chen et al., 2013; Soufi et al., 2012) is a well-established barrier of reprogramming. Medium was changed every other day for 6 days. Subsequently, cells were washed once with PBS and fixed with 50 μl 4% paraformaldehyde (Sigma) for 20 min. Cells were then washed twice before 25 ml anti-human TRA-1-60 antibody (1:300, BD #560173) was added, and cells were incubated at 4°C overnight. Hoechst 3342 (1:20,000, Invitrogen) was added to each well and stained for 30 min. The cells were then washed once with PBS and covered in 50 ml PBS.

Cells were imaged with an IXU ultra plate-scanning confocal microscope (Molecular Devices) at ×10 magnification, and nine pictures were taken per well. Images obtained from multiple sites were combined to generate a

(B) MA plots of the de novo assembled transcripts in si*SFRS11* versus siNT and si*SMAD3* versus siNT.
(C) GO analysis of genes differentially spliced by SFRS11 shows the enrichment of the indicated biological processes (top) and protein domains (bottom). x axis represents the enrichment score (−log10(p-value)) of the ontology term.
(D) Sashimi plots demonstrating the differential splicing events occurring at the indicated genes. y axis represents the RPKM of the reads. The numbers within the histogram indicate the number of reads mapped to the corresponding junction. Exon skipping (*ZNF207*), mutually exclusive splicing (*MTA1*), 5′ alternative splicing (*HMGA3*), and 5′ alternative splicing (*ZNF462*) are shown.
(E) Analysis of exon-skipping events after depletion of *SFRS11*. 434 events were identified to have an FDR < 0.01 when compared to siNT control. Knockdown of *SFRS11* resulted in the increase of exon inclusion (310 events, gray) as compared to exon exclusion (124 events, blue).
(F) Analysis of mutually exclusive events after depletion of *SFRS11*. 114 events were identified to have an FDR < 0.01 when compared to siNT control. Exon 1 refers to the upstream exon that was included, while exon 2 refers to the downstream exon that was included. Knockdown of *SFRS11* resulted in increased inclusion of the downstream exon (72 events) as compared to upstream exon inclusion (42 events).
(G) Validation of alternative splicing events during reprogramming using PCR primers flanking alternatively spliced exons. Top panels illustrate exon skipping, whereas bottom panels illustrate mutually exclusive events.
(H) Regulation of *ZNF207* isoforms by SFRS11. Depletion of SFRS11 resulted in increased retention of *ZNF207* exon 9 and subsequent switching of isoform B to isoform A and C. n = 4; error bars indicate SD.
(I) Overexpression of SFRS11 WT and mutant constructs decreases reprogramming efficiency. n = 3; error bars indicate SD. *p < 0.05.
(J) SFRS11 binds to endogenous *ZNF207* transcripts. CLIP-qPCR of BJ fibroblasts overexpressing WT-SFRS11-HA, but not ΔRRM-SFRS11-HA, construct showing enrichment for *ZNF207* transcripts. Binding of the *ZNF207* transcript is dependent on the RNA-recognition motif (RRM). n = 3; error bars indicate SD.
(K) Gel electrophoresis analysis of CLIP-qPCR. Detection of *ZNF207* transcript was only possible with the pull-down of WT-SFRS11-HA protein in reprogramming BJ fibroblast cells.
(L) Overexpression of *ZNF207* isoforms isoform A and C enhanced reprogramming, while the expression of isoform B impeded the process. n = 3; error bars indicate SD. *p < 0.05, **p < 0.01, ***p < 0.001.
(M) Antisense oligonucleotides (AONs) induce *ZNF207* exon skipping, which in turn results in reduced reprogramming efficiency. scrAON is a scrambled negative control. AON 1 and 2 induced exon skipping of *ZNF207* exon 6, resulting in the isoform switch of C to A. AON 3–5 induce skipping of *ZNF207* exon 9, resulting in the isoform switch of C to B. n = 3; error bars indicate SD. ****p < 0.0001. Refer to Table S4 for sequences.
(N) Proposed model illustrating the role of SFRS11 in the regulation of reprogramming-specific alternative splicing. Biological processes that are significantly enriched are boxed. SFRS11 regulates splicing of both unknown and previously reported reprogramming regulators, including *ZNF207*, *DNMT1*, *MTA1*, *HIRA*, *MBNL1*, and *PPARG*.

montage of different panels. Granule area, integrated fluorescent intensity, and nuclei number were quantified using MetaXpress Image Acquisition and Analysis software V2. The Z score was calculated using the formula $Z = (X - m)/SD$, where m is the mean of negative control and SD is the standard deviation of the whole population. X is the sample value calculated based on TRA-1-60+ area per number of cells. For further analysis, we excluded genes that reduced cell number by <50% of the average cell number per well in each plate. This was done to remove false-positive candidates that affected cell viability.

More details on production of viral supernatant, reprogramming assay, cell culture, differentiation of iPSC clones, immunostaining, RNA extraction, RNA sequencing, ChIP sequencing, AON transfection, microarray data, differential GO analysis, enrichment analysis, GSEA, and other procedures can be found in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the raw data from sequencing experiments reported in this paper is GEO: GSE68234.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.05.049.

AUTHOR CONTRIBUTIONS

C.-X.D.T., J.-W.C., and Z.-S.C. designed and performed research, analyzed data, and wrote the paper; H.F.W., H.C.G., D.M., S.S., G.Y.L.G., and E.K. designed and conducted research; L.Y., V.T., Y.-T.C., J.J.C., G.Q.D., K.B.W., C.A.E.F., H.L., Y.-P.L., and F.A.B. analyzed data; and Y.-H.L. designed research, analyzed data, and wrote the paper.

ACKNOWLEDGMENTS

We are grateful to Kenneth Zaret and Samantha Seah for helpful discussion. We thank Tan Yong-Si, Kin Xiao-Xuan, and Julien Maury for technical assistance. H.L. supported by grants from the National Institutes of Health (CA196631-01A1 and 1U54GM114838-01). Y.-H.L. is supported by the A*Star Investigatorship research award and grants JCO R09138 and JCO R09125. We are grateful to the Biomedical Research Council, Agency for Science, Technology and Research, Singapore for research funding.

Received: February 19, 2016
Revised: April 3, 2016
Accepted: May 11, 2016
Published: June 9, 2016

REFERENCES

Birmingham, A., Selfors, L.M., Forster, T., Wrobel, D., Kennedy, C.J., Shanks, E., Santoyo-Lopez, J., Dunican, D.J., Long, A., Kelleher, D., et al. (2009). Statistical methods for analysis of high-throughput RNA interference screens. Nat. Methods *6*, 569–575.

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.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat. Genet. *45*, 34–42.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

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

Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241–245.

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.

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.

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.

Lai, W.-H., Ho, J.C.-Y., Lee, Y.-K., Ng, K.-M., Au, K.-W., Chan, Y.-C., Lau, C.-P., Tse, H.-F., and Siu, C.-W. (2010). ROCK inhibition facilitates the generation of human-induced pluripotent stem cells in a defined, feeder-, and serum-free system. Cell. Reprogram. *12*, 641–653.

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.

Lin, T., Ambasudhan, R., Yuan, X., Li, W., Hilcove, S., Abujarour, R., Lin, X., Hahm, H.S., Hao, E., Hayek, A., and Ding, S. (2009). A chemical platform for improved induction of human iPSCs. Nat. Methods *6*, 805–808.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Ohta, S., Nishida, E., Yamanaka, S., and Yamamoto, T. (2013). Global splicing pattern reversion during somatic cell reprogramming. Cell Rep. *5*, 357–366.

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.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

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.

Qin, H., Blaschke, K., Wei, G., Ohi, Y., Blouin, L., Qi, Z., Yu, J., Yeh, R.-F., Hebrok, M., and Ramalho-Santos, M. (2012). Transcriptional analysis of pluripotency reveals the Hippo pathway as a barrier to reprogramming. Hum. Mol. Genet. *21*, 2054–2067.

Qin, H., Diaz, A., Blouin, L., Lebbink, R.J., Patena, W., Tanbun, P., LeProust, E.M., McManus, M.T., Song, J.S., and Ramalho-Santos, M. (2014). Systematic identification of barriers to human iPSC generation. Cell *158*, 449–461.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

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.

Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Schöler, H.R., and Ding, S. (2008). A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell *2*, 525–528.

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.

Stadtfeld, M., and Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. Genes Dev. *24*, 2239–2263.

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.

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

Yang, C.-S., Chang, K.-Y., and Rana, T.M. (2014). Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming. Cell Rep. *8*, 327–337.

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.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., et al. (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. Cell Stem Cell *3*, 475–479.