



Published in final edited form as:

Cell Rep. 2019 May 07; 27(6): 1726–1741.e5. doi:10.1016/j.celrep.2019.04.056.

Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency

Khoa A. Tran^{1,5}, Stefan J. Pietrzak^{1,5}, Nur Zafirah Zaidan^{1,5}, Alireza Fotuhi Siahpirani¹, Sunnie Grace McCalla¹, Amber S. Zhou¹, Gopal Iyer², Sushmita Roy^{1,3}, and Rupa Sridharan^{1,4,6,*}

¹Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA

²Department of Human Oncology, University of Wisconsin-Madison, Madison, WI, USA

³Department of Biomedical Informatics, University of Wisconsin-Madison, Madison, WI, USA

⁴Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, WI, USA

⁵These authors contributed equally

⁶Lead Contact

SUMMARY

Elucidating the mechanism of reprogramming is confounded by heterogeneity due to the low efficiency and differential kinetics of obtaining induced pluripotent stem cells (iPSCs) from somatic cells. Therefore, we increased the efficiency with a combination of epigenomic modifiers and signaling molecules and profiled the transcriptomes of individual reprogramming cells. Contrary to the established temporal order, somatic gene inactivation and upregulation of cell cycle, epithelial, and early pluripotency genes can be triggered independently such that any combination of these events can occur in single cells. Sustained co-expression of Epcam, Nanog, and Sox2 with other genes is required to progress toward iPSCs. Ehf, Phlda2, and translation initiation factor Eif4a1 play functional roles in robust iPSC generation. Using regulatory network analysis, we identify a critical role for signaling inhibition by 2i in repressing somatic expression and synergy between the epigenomic modifiers ascorbic acid and a Dot1L inhibitor for pluripotency gene activation.

Graphical Abstract

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondence: rsridharan2@wisc.edu.

AUTHOR CONTRIBUTIONS

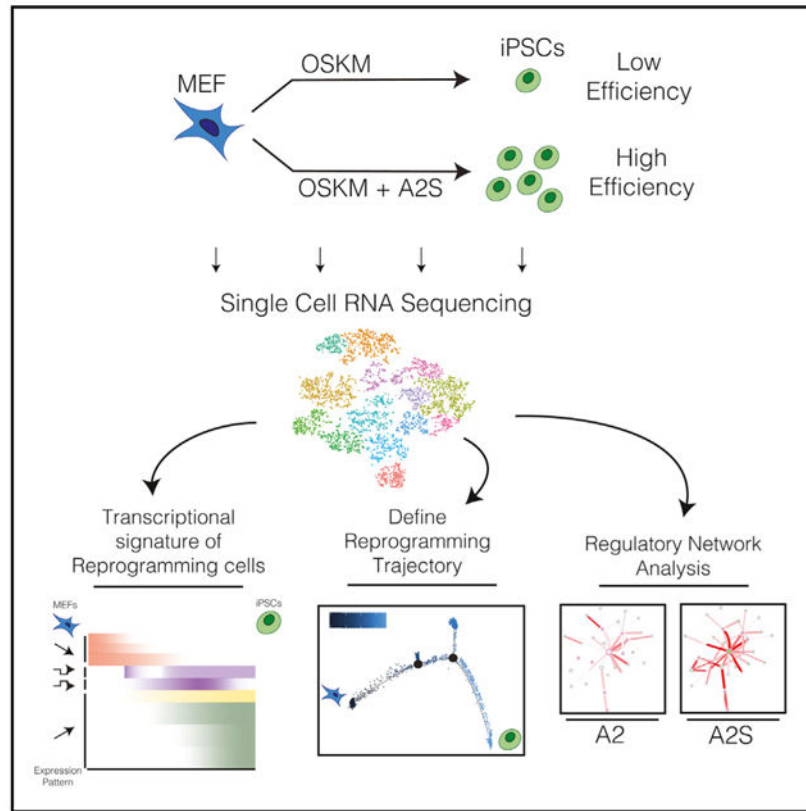
K.A.T., S.J.P., and N.Z.Z. performed data collection, analyzed and interpreted the data, and generated figures under the supervision of R.S. who designed the project and wrote the manuscript with S.R. A.F.Z. and S.G.M. performed co-expression and MERLIN analysis under the supervision of S.R. who designed the analysis. G.I. provided essential resources and critical review. A.S.Z. contributed to initial experiments and analysis.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.04.056>.

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.



In Brief

Tran et al. combine ascorbic acid, 2i, and Dot1l inhibition to robustly generate induced pluripotent stem cells. With single-cell transcriptomes, they define the transcriptional signature and key regulators of reprogramming cells. Using network analysis, they find 2i suppresses somatic while ascorbic acid and Dot1l inhibitor collaboratively upregulate pluripotency genes.

INTRODUCTION

Somatic cells can be reprogrammed to induced pluripotent stem cells (iPSCs) by the introduction of the transcription factors Oct4, Sox2, Klf4, and c-Myc (OSKM) (Takahashi and Yamanaka, 2006). Mouse iPSCs are functionally equivalent to embryonic stem cells (ESCs) because they pass all the tests of pluripotency, including tetraploid complementation (Zhao et al., 2009). The efficiency of reprogramming remains low at about 5% even when the reprogramming factors are inducibly expressed from a single locus in the mouse genome (Buganim et al., 2013). In addition, iPSC colonies appear at different times during the reprogramming process (Apostolou and Hochedlinger, 2013; Buganim et al., 2013; Papp and Plath, 2013). Identifying only those cells that successfully complete the reprogramming process versus those that fail to do so can reveal key mechanisms that make the reprogramming process inefficient. Although some markers, such as SSEA1, EPCAM, CD73, ICAM1, and CD44, enrich for successfully reprogramming cells (Lujan et al., 2015;

O'Malley et al., 2013; Polo et al., 2012), it is not yet possible to prospectively identify only the cells that will become iPSCs to follow them as they reprogram.

Transcriptional profiling of bulk reprogramming populations over time has led to the description of a temporal series of events with early downregulation of somatic cell expression followed by metabolic and cell cycle changes that culminates in the activation of the pluripotency gene regulatory network (Apostolou and Hochedlinger, 2013; Apostolou and Stadtfeld, 2018). Mouse embryonic fibroblasts (MEFs) undergo a mesenchymal-to-epithelial transition (MET) before pluripotency gene activation during reprogramming (Hussein et al., 2014; Li et al., 2010; Mikkelsen et al., 2008; Samavarchi-Tehrani et al., 2010). Importantly, whether all cells undergoing reprogramming have to trigger these programs in the same temporal order remains unknown. Due to the low efficiency and variable kinetics of obtaining iPSCs, reprogramming cultures will have heterogeneous expression profiles. Therefore, in population-based analyses of unsorted cells, expression signatures from cells that will successfully reprogram are obscured.

To overcome these issues with ensemble profiling, single-cell analysis of candidate factors in reprogramming MEFs has been performed both at the RNA and protein level. These studies have uncovered intermediate markers, a role for Ras-signaling, and a role for Sox2 in the deterministic activation of the pluripotency network. (Buganim et al., 2012; Kim et al., 2015; Lujan et al., 2015; Zunder et al., 2015). More recent experiments have focused on profiling cells during reprogramming in low-efficiency systems, including non-transgenic chemical reprogramming (Zhao et al., 2018; Guo et al., 2019; Schiebinger et al., 2019).

Reprogramming efficiency can be increased by the modulation of regulators that decrease chromatin compaction and those that perturb signaling pathways (Esteban et al., 2010; Huangfu et al., 2008; Ichida et al., 2009; 2014; Maherali and Hochedlinger, 2009; Mikkelsen et al., 2008; Onder et al., 2012; Shi et al., 2008; Silva et al., 2008; Tran et al., 2015). We and others have combined such epigenomic and signaling modulators and found that they synergistically increase reprogramming efficiency from OSKM-expressing cells (Bar-Nur et al., 2014; Tran et al., 2015; Vidal et al., 2014). In this study, we added SGC0946 (inhibitor of Dot1L, a histone H3K79 methyltransferase) along with our previous cocktail of ascorbic acid (vitamin C) and 2i (inhibitors to mitogen-activated protein [MAP] kinase and glycogen synthetase kinase), in conjunction with OSKM to reprogram MEFs to iPSCs at an efficiency of ~40% within 6 days. Although each small molecule has been used previously, to our knowledge this particular combination (called A2S [ascorbic acid, 2i, SGC] henceforth) has not been reported.

Using single-cell RNA sequencing (RNA-seq) analysis, we profiled reprogramming MEFs along a time course in both a regular serum-containing (fetal bovine serum [FBS]) and the A2S system. We found that early events, such as epithelial and cell cycle activation, are turned on independently. Surprisingly, all mesenchymal genes are not downregulated together in the same cells, and some genes, such as *Twist1*, can even be found expressed with early pluripotency marker *Nanog*. A large majority of the cells in FBS stop cycling partly due to senescence, which can be overcome by the addition of A2S. *Nanog*, *Oct4*, and even *Sox2* could be activated in individual cells, but what distinguished successful

reprogramming was the detectable coexpression of these genes in different modules. Nanog was found in a sub-cluster with Epcam, Sall4, and Tdgf1; Oct4 with Zfp42; and Sox2 with Utf1 and Dppa5a. The lack of detectable expression of some markers, such as Epcam, with other pluripotency genes correlated with cells reverting to an Epcam-negative state. Functional experiments provide a role for reprogramming-specific transient upregulation of transcription factors, such as Ehf; translation initiation (Eif4a1); and factors such as Phlda2 for reaching an iPSC state. By applying a network-based analytical framework to our single-cell data, we studied the effect of individual components of A2S on the acquisition of pluripotency. Our analysis identified that specific connections of the pluripotency network can only be made when both epigenomic modifiers are present, but without the suppression of somatic expression by the signaling inhibitors reprogramming efficiency is compromised. Thus, we have uncovered that reprogramming need not progress in discrete stages but instead is the result of co-occurring modulation of various networks.

RESULTS

Combining Epigenomic and Signaling Modifiers Leads to High-Efficiency Generation of Bona Fide iPSCs

We reprogrammed MEFs that have a doxycycline (dox)-inducible cassette containing a transgene with four reprogramming factors: Oct4, Sox2, Klf4, and c-Myc (OSKM). iPSC generation was monitored by immunofluorescence for NANOG at various time points. The NANOG+ colonies that remained after dox withdrawal are transgene-independent iPSCs (Brambrink et al., 2008; Stadtfeld et al., 2008). In FBS conditions, NANOG+ colonies emerged by day 6 and most were transgene independent by day 12 of reprogramming, yielding an efficiency of about 3.2% (Figure 1A; STAR Methods).

As very few cells successfully reprogram in FBS, we next sought to increase reprogramming efficiency to elucidate the transcriptional changes required for pluripotency acquisition. We have previously shown that the addition of ascorbic acid (AA) and 2i increases reprogramming efficiency of both embryonic and adult fibroblasts (Tran et al., 2015). A small-molecule screen of chemicals (data not shown) revealed that the addition of an inhibitor to the H3K79 methyltransferase Dot1L called SGC0946 (Jackson et al., 2016) to the AA+2i combination boosted iPSC generation from reprogrammable MEFs. By day 6, ~1,900 Nanog+ iPSC colonies were obtained at an efficiency of ~42% (STAR Methods) (Figure 1B). Beyond this time point, the colonies started merging with each other, and therefore, it was chosen as the endpoint for analysis. The A2S system also increased the kinetics of reprogramming because the NANOG+ colonies on day 4 were already transgene independent (Figure 1B) as compared to day 9 of FBS reprogramming (Figure 1A). To avoid biases from plating efficiencies (Schwarz et al., 2018), we further verified the efficiency by reprogramming MEFs as single cells. We found that transgene-independent colonies were obtained in ~40% of the wells in the A2S system (Figure 1C). Thus, the A2S combination of small molecules yielded a great increase in reprogramming efficiency and kinetics.

To determine whether iPSCs generated from the A2S system were bona fide, colonies were picked on day 6 from an A2S reprogramming experiment and could be passaged in FBS

without loss of pluripotency. These iPSCs were karyotypically normal and produced teratomas that were comprised of cells from all three germ layers (Figures S1A and S1B).

Single-Cell RNA-Seq Time Course Confirms Heterogeneity of Reprogramming Populations

To dissect the intrinsic heterogeneity during FBS reprogramming and determine whether the A2S system accelerated or overcame the FBS reprogramming barriers, we performed singlecell transcriptomics. We profiled reprogramming cells in FBS on days 3, 6, 9, and 12; A2S on days 2, 4, and 6; as well as the starting population of MEFs and endpoint of ESCs using a microfluidics-based droplet digital sequencing system (Bio-Rad ddSeq, STAR Methods). In addition, iPSCs that were generated from the A2S system were profiled to determine their similarity to ESCs. Because AA and 2i are known to change the expression profile of ESCs (Blaschke et al., 2013; Marks et al., 2012), we also sequenced ESCs and iPSCs that had been passaged in A2S.

We obtained an average of about 55,000 reads and 13,000 uniquely identified transcripts per cell, which corresponded to a total 18,005 genes detected across all cells (Figure S1C; STAR Methods). We used the Monocle2 program (Qiu et al., 2017a, 2017b) (Figures S1D and S1E) to analyze the gene expression data and identified gene regulatory networks using the MERLIN algorithm (Chasman et al., 2016) to provide insights into the different factors that influence reprogramming efficiency. A t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis (STAR Methods) revealed the iPSCs derived from A2S when passaged in FBS clustered with ESCs grown in FBS and away from ESCs or iPSCs passaged in A2S (Figure 1D). This result further confirmed that the iPSCs had reached an ESC-like transcriptional state. As expected, ESCs cultured in A2S expressed blastocyst-enriched genes, such as *Dazl*, while also repressing the development-associated gene *Emb* and showed more homogeneous expression of naive marker *Tbx3* but not *Rex1* (Figure S1F).

A2S Accelerates FBS Reprogramming

The cells profiled from the time course analysis were grouped into 14 clusters (Figure 2A). The starting MEFs were heterogenous and occupied two clusters (cluster 2 and cluster 7) (Figure 2A). For the FBS samples, the cells on day 3 occupied a single cluster (77% of cluster 3) away from days 6, 9, and 12 reprogramming cells (Figure 2A). Similarly, the day 2 of A2S samples predominated a single cluster (92% of cluster 5), whereas the cells from day 4 and day 6 belonged to several clusters (Figure 2B; Figure S2A). Therefore, at the beginning of reprogramming, the cells are more homogeneous than later time points, irrespective of the efficiency of the system. The fact that cells from different time points cluster together based on similarity in gene expression profiles suggests that average expression from previous time-point-based analysis warrants analysis by single-cell sequencing. A small fraction of cells from A2S were found in the FBS clusters and vice versa (Figure 2B). The entire reprogramming population also clustered away from ESCs and iPSCs grown in A2S (Figure S2B). Distance in the t-SNE does not necessarily reflect the most differential gene clusters. However, given that the cells in reprogramming cultures were most similar in gene expression profile to pluripotent cells grown in serum, ESCs grown in FBS were used as the endpoint for all subsequent analyses.

From previous bulk RNA-seq and mass cytometry analysis, various cell surface markers have been identified that enrich for reprogramming cells that will transition to iPSCs (Lujan et al., 2015; Nefzger et al., 2017; O'Malley et al., 2013; Polo et al., 2012), although the same markers can have heterogeneous expression in ESCs (O'Malley et al., 2013). We reasoned that if A2S reprogramming was an accelerated version of FBS reprogramming, the same markers would be found in a greater proportion. The marker CD44 is high in MEFs, whereas ICAM1 is transiently increased in reprogramming cells (O'Malley et al., 2013). The CD44-/ICAM1+ population was two-fold greater in A2S by day 6 than FBS on day 12 (Figure S3B). Similarly, the transient CD73 intermediate marker (Lujan et al., 2015) was rapidly acquired and downregulated (Figure S3A). There was a greater decrease in the MEF-specific Thy1+ or Vcam+ cells in A2S as compared to FBS reprogramming (Polo et al., 2012; Schwarz et al., 2018) (Figure S3A). The Thy1-/Fut9+ (SSEA1) (Polo et al., 2012) and the Epcam+/Sca1-/Fut9+ (Schwarz et al., 2018) populations that are more predictive of cells that will complete reprogramming were both ~4-fold higher in A2S by day 6 as compared to FBS (Figure S3B). Notably, the gene expression of Mbd3 and Gatad2a were not affected in A2S reprogramming (Figure S3A). The absence of these proteins leads to high-efficiency reprogramming (Mor et al., 2018; Rais et al., 2013). Taken together, these results indicate that A2S improves the kinetics and efficiency of the route taken by FBS reprogramming cells.

To identify the genes that distinguished the clustering of single cells in the Monocle t-SNE analysis (Figure 2A), we examined the top 10% of differentially expressed genes between all the clusters. Because this is single-cell data, we measured both the percentage of cells displaying each of the four major patterns of expression between MEFs and ESCs as well as the average expression (Figure 2C; Figure S2C; Table S1). There was a net decrease in expression (groups A–D), which included genes in categories such as cell differentiation and migration; a reprogramming-related decrease (groups F–H), mainly composed of cell cycle, DNA replication, and spliceosome-related genes; a reprogramming-related increase (groups K–L); and a net increase from MEFs to ESCs (groups M–N), which included pluripotency genes. We also observed a fifth pattern (group O), which was made of ribosomal genes that displayed tremendous cell-cell variability but was expressed in all cells.

Mesenchymal and Epithelial Changes Are Independently Regulated

From bulk sequencing experiments, it is thought that downregulation of somatic cell gene expression, including the mesenchymal genes, are early events in reprogramming (Apostolou and Hochedlinger, 2013; Apostolou and Stadtfeld, 2018; Li et al., 2010; Samavarchi-Tehrani et al., 2010). We found that not all mesenchymal genes are rapidly decreased in all cells. The majority of the cells in group A (Figure 2C) decreased expression of developmental signaling and cell migration genes, including Tgfb3, Snai1, and Twist2 (Figure 3A). Larger fractions of cells retained expression of Id1 and Id2, and the mesenchymal factors Zeb1 and Zeb2 (group B). Expression of several collagens, Egr1 and Twist 1 (group C), was retained in an even higher proportion of cells than group B (Figure 3A). Thus, there are three different trends for populations to lose mesenchymal gene expression with a large majority of cells in FBS reprogramming still retaining MEF-like gene expression even at later time points

The mesenchymal MEFs have to transition to an epithelial state indicated by the upregulation of E-cadherin (Cdh1) (Apostolou and Hochedlinger, 2013; Apostolou and Stadtfeld, 2018; Li et al., 2010; Samavarchi-Tehrani et al., 2010). Given the differential proportion of mesenchymal genes that were turned off in individual cells, we determined the co-expression of Cdh1 with several mesenchymal genes. It should be noted that because of the limit of detection of single-cell transcriptomics, such analysis may underestimate the number of co-expressing cells. Surprisingly, Cdh1 upregulation was compatible with the expression of mesenchymal genes, albeit in different proportions, as well as the somatic marker Thy1 (Figure 3B). Instead, from our data, it is apparent that the mesenchymal gene downregulation and E-cadherin upregulation operate as different modules. For example, the downregulation of Snai1 does not automatically lead to Cdh1 expression. We orthogonally confirmed the RNA-seq results by performing immunofluorescence for Twist1 and Cdh1 and found an overlap of both markers in the proportion predicted by the transcriptional data (Figure 3C). The trends of dual mesenchymal gene+/Cdh1+ cells were similar in A2S and FBS reprogramming (Figure S2D).

By performing a pairwise comparison between the earliest time points of the FBS and A2S time course (cluster 3 versus cluster 5; Figure 2C), we found that FBS cells on day 3 still retained the expression of genes associated with system development (Col3a1) as well as signal transduction (Fgf7, Egr1, and Igfbp3) that were greatly reduced by day 2 of A2S reprogramming. Thus, the acceleration of reprogramming in A2S is partially derived from increasing the rate of downregulation of somatic genes.

Reprogramming-Specific Transient Gene Expression Patterns Are Important for Conversion to iPSCs

Because iPSCs self-renew indefinitely, mechanisms that confer an ESC-like cell cycle improve reprogramming efficiency (Hanna et al., 2009; Marión et al., 2009; Ruiz et al., 2011; Utikal et al., 2009). The starting population of MEFs heterogeneously expressed cell cycle markers to segregate into two different clusters (cluster 2 and 7; Figure 2C). Interestingly, both FBS day 3 and A2S day 2 reprogramming cells also expressed cell cycle genes, such as Mcm6, Bub1b, and Ccnb1 (groups F–H; Figure 2C; Table S1). Therefore, either the induction of the reprogramming factors upregulated these genes in the majority of MEFs or reprogramming was productively initiated only from those MEFs that were already cycling. The initial upregulation of cell cycle observed in bulk transcriptomic data may represent the selection of cycling MEFs (Mikkelsen et al., 2008) for reprogramming rather than a true upregulation in all cells.

After this time point, there was a dramatic difference in the way the two systems behaved. In the FBS clusters, the vast majority of the cells (76% of all FBS cells) downregulated cell cycle genes (clusters 4, 11, 13, and 14), whereas a minority retained expression (cluster 10) (Figure 3D). In contrast, in the A2S system, the vast majority of the cells still retain the expression of cell cycle genes and a small fraction (21% of all A2S cells, located within cluster 8) shut these genes off (Figure 3D). This result was corroborated by immunofluorescence for the cell cycle marker Ki67 with a rapid decline by day 6 of FBS reprogramming, which was not observed in A2S cells (Figure 3E).

Cell cycle gene expression upregulation was compatible with *Thy1*, *Zeb2*, and *Twist1* expression, as well as *Cdh1* in both FBS and A2S systems (Figure 3G; Figure S2D). This result suggests that the cell cycle can also be activated with continued somatic expression.

It is known that in FBS, most reprogramming cells experience reprogramming-induced senescence (Banito et al., 2009; Li et al., 2009; Mikkelsen et al., 2008). Corroborating this notion, the antiproliferative *Cdkn1c* gene was highly upregulated in FBS reprogramming cells but not in the A2S system (Figure 3E). By contrast, p53 transcription levels were maintained in the entire population. Senescence-associated genes, such as *Ink4a*, were also activated during A2S reprogramming and interleukin-6 (IL-6) remained unactivated (Figure S3A). Thus, the senescence block may be overcome by the lack of activation of *Cdkn1c* (Figure 3E). In this aspect, the A2S system in MEFs resembles a cohort of fast-cycling granulocytes—monocyte precursors that undergo non-stochastic reprogramming due to reduced levels of *Cdkn1c* (Guo et al., 2014).

Besides senescent genes, this third pattern of reprogramming-related upregulation (groups K and L; Figure 2C) was without a specific gene ontology. Because cell fate transitions are often orchestrated by transcription factors, chromatin-modifying proteins, or signaling molecules, we knocked down three genes belonging to these categories—*Ano1*, *Aldh3a1*, and *Ehf*—during reprogramming. Among these genes, the knock down of *Ehf* caused a decrease in A2S reprogramming efficiency (Figure 3H; Figure S2E). This suggests that transient upregulation of some genes is, in fact, required for reprogramming to iPSCs and does not represent a different lineage-specific endpoint.

Co-expression of Core Pluripotency Factors Are Independent of Each Other

The activation of genes highly expressed in ESCs (groups M and N; Figure 2C) was largely restricted to reprogramming clusters C9, C6, and C10 that already expressed cell cycle genes (Figure S4A). We examined the expression of known pluripotency genes within this group. *Epcam*, *Sall1*, and *Gdf3* were expressed in reprogramming clusters other than the ones with the most ESC-like characteristics (Figure 4A). This suggests that they can be activated in isolated cells and may not predict cells completing the transition to iPSCs. Surprisingly, *Sox2* was also expressed in cells other than the ones most similar to cluster 1, suggesting that its activation may not be sufficient to activate a cascade of deterministic pluripotency gene activation as previously suggested (Buganim et al., 2012) (Figure 4A). We next determined which genes were most prevalently expressed with the core pluripotency factors *Oct4*, *Sox2*, and *Nanog* in the reprogramming populations, while acknowledging the caveat that such analysis may be limited by the detection limit of single-cell transcriptional sequencing. *Nanog* was detected with *Sall4*, *Epcam*, and *Tdgf1* (*Cripto*) (Figure 4B). Within the population of *Nanog*-expressing cells, *Sall4* was equally expressed in both cluster 6 and cluster 9 (Figure 4B). However, *Tdgf1* expression was higher in cluster 9 cells, suggesting that *Tdgf1* may be more important for activating the rest of this subset (Figure 4B). On the other hand, although *Oct4* was activated with *Zfp42* (Figure S4B), *Sox2* was found with *Dppa5a* and *Utf1* and was part of a larger cluster that included *Tet1* and *Zscan10* (Figure 4C). In cluster 10 that is predominantly made of cells from FBS reprogramming, this larger

subset is heterogeneously activated. In contrast, in cluster 6 that mostly contains A2S reprogramming, the whole group was coordinately upregulated (Figure 4C).

The most restricted pattern of expression included *Dppa4*, which is known to be a marker of the “stabilization” phase of reprogramming that occurs after the core pluripotency genes are activated (Golipour et al., 2012). *Dppa4* was detected in the subset with *Lin28a* and *Phlda2*, a gene involved in placental growth (Salas et al., 2004) (Figure 4D).

Intrigued by this finding, we depleted the levels of *Phlda2* during reprogramming. Interestingly, although the number of NANOG-expressing colonies remained similar between *Phlda2* knock down and control, we found a 25% decrease in the number of DPPA4-positive colonies (Figure 4E). Therefore, the coexpression of pluripotency factors within each subgroup may functionally predict regulators of transitions to the next stage toward pluripotency.

Similar to downregulation of MEF genes and activation of cell cycle, pluripotency gene activation is increased in a greater proportion of cells, to a higher extent and more homogeneously with co-expression partners in A2S as compared to FBS reprogramming.

Continued Mesenchymal Expression Is a Roadblock to High-Efficiency Reprogramming

From these analyses, it is clear that A2S is more efficient than FBS reprogramming in accelerating each of the four major patterns of expression. Therefore, examining the A2S system alone would help us identify genes that are bottlenecks to the completion of reprogramming in cells that are much further along the process. In fact, when we compared the differentially expressed genes that were only related to reprogramming in FBS or A2S alone, we found about 33% unique to the A2S system (Figure S5A). The ones that were solely found in A2S reprogramming were enriched for gene ontology terms, such as system development and cell differentiation, and included pluripotency genes, such as *Nanog* and *Oct4*. In contrast, the FBS-exclusive gene expression was dominated by cell cycle genes (Figure S5A). Therefore, we further examined the A2S cells by performing a trajectory analysis in which cells are arranged in pseudotime according to similarity in gene expression patterns (Trapnell et al., 2014) (Figure 5A). As expected, a larger fraction of day 6 (63%) cells were found in the part of the trajectory toward pluripotent cells than those that were found before the branchpoint.

We performed branched expression analysis modeling (BEAM) (Qiu et al., 2017a) to identify the genes that were overrepresented in cells that continued along the trajectory toward ESCs from the ones that were found in the branch. We note that ontogenically MEFs cannot convert to ESCs but use the trajectory to determine a path toward pluripotency. At the early branchpoint, the cells that continue toward ESCs have a higher expression of epithelial genes, such as *Cdh1* and *Epcam* (Figure 5B). At the later branchpoint, cells that continue have already activated the cell cycle and present high levels of *Nanog* as expected (Figures 5B and 5C). Surprisingly, the mesenchymal gene *Twist1* was found to be a gene that influences the branchpoint decision even at this late point in the pseudotime trajectory (Figure 5C) and was even found to be co-expressed with *Nanog*. Although *Nanog* levels

were similar in cells at the beginning of branch 2, cells that stall have a higher level of Twist1 co-expression than those that continue (Figure 5D).

From population-based studies, cells that express Epcam during intermediate phases of reprogramming have a greater probability of completing the process (Polo et al., 2012). In the branchpoint analysis, several cells that exit the trajectory express high levels of Epcam but at the end of the branch have decreased expression rather than maintained levels (Figure 5C). Given that Epcam is found co-expressed with a subset of genes (Figure 4B), we wondered whether the expression of Epcam was influenced by expression levels of other genes within its subset. In fact, we found that Epcam⁺ cells that continue along to complete pluripotency co-expressed higher levels of Nanog, Tdgf1, and Sall4 than those that stall at the branchpoint (Figure 5D). This result suggests that activation of all the genes within a subset is important to sustain initial expression. Because single-cell analysis destroys the cell, the cells at the end of the branchpoint could represent those that never expressed Epcam and are at the end of the trajectory due to covariance with other genes. Therefore, we sorted cells based on the level of EPCAM expression on day 3 of A2S reprogramming (Figure 5E). After allowing reprogramming to continue for an additional 3 days, we found that 7.5% of the high- and 16.6% of medium-expressing EPCAM cells gave rise to an EPCAM-negative population (Figure 5E). Taken together, these analyses suggest that without co-expression of other genes within the subset, cells may revert to an Epcam-negative state, whereas with co-expression, cells persist along the trajectory toward an ESC-like state.

A reverse pattern to Epcam is observed for the branchpoint gene, translation initiation factor Eif4a1. Here, after an initial downregulation, cells that successfully remain on the trajectory upregulate gene expression (Figure 5C). Eif4a1 is a part of the translation initiation complex along with the closely related protein Eif4a2 (Modelska et al., 2015; Williams-Hill et al., 1997). To determine if Eif4a1 had a causal role in obtaining iPSCs, we depleted its levels using RNA interference during A2S reprogramming. Interestingly, depletion of Eif4a1 severely compromised the efficiency of reprogramming (Figure 5F). This decrease was not due to a change in the number of cells or increasing cell death (Figure 5F). Taken together, these data suggest that sustained expression of genes is affected by coexpression of other factors and is required for completing the process to a productive pluripotent state.

A2S Concurrently Enhances Downregulation of MEF Genes and Upregulation of ESC Genes

The chemicals we used for high-efficiency reprogramming include signaling inhibitors and two epigenomic modulators—AA, which is thought to regenerate 2-oxoglutarate-dependent chromatin-modifying enzymes (Hore et al., 2016), and SGC0946, an inhibitor of Dot1L-mediated histone H3K79 methylation (Jackson et al., 2016). To understand the relative contribution of each component, we subjected MEFs to every dual combination of chemicals and assessed reprogramming efficiency on day 6. We found that SGC+2i (S2) yielded approximately half the NANOG⁺ colonies of the A2S combination, whereas AA+2i (A2) and AA+SGC (AS) were only 6.6% and 10.4% efficient, respectively, on day 6 of reprogramming (Figure 6A). Irrespective of the dual combinations that were used, the iPSC

colonies remained NANOG+ after dox withdrawal. Exposure to each individual component had lower effects on enhancing reprogramming efficiency (data not shown).

We performed single-cell RNA-seq on reprogramming MEFs that had been subjected to each dual combination on day 4 and day 6 and compared the profiles to FBS and A2S reprogramming. Because none of the dual combinations were able to achieve the high efficiency of the A2S system, we hypothesized that each dual combination likely rewires some components of the gene regulatory network controlling the transcriptional dynamics of reprogramming. Therefore, we first reconstructed the putative regulatory network by using the FBS+A2S single cell RNA-seq (scRNA-seq) dataset collected in this study (STAR Methods) using an expression-based network inference algorithm, MERLIN (Chasman et al., 2016). We focused on the ~1,800 genes used to initially differentiate the Monocle clusters in the FBS+A2S dataset (Figure 2A) along with sufficiently expressed regulators, such as transcription factors, chromatin remodelers, and signaling proteins (Figure 6B; STAR Methods). MERLIN is based on a probabilistic framework that predicts the regulators of a target gene based on the ability of the regulator's mRNA levels to explain the variation in a target gene's expression level. Using probabilistic modeling, MERLIN allows regulators to control target genes with similar expression levels to have non-identical regulatory programs. Furthermore, target genes are grouped into modules based on their co-expression and shared regulatory program (STAR Methods). Thus, there are two outputs of MERLIN: (1) modules that represent characteristic patterns of expression of genes and (2) networks that specify the regulators of individual genes as well as modules. The MERLIN analysis produced 15 modules with 5 or more genes. There were 4,962 interactions between 1,009 regulators and 1,628 target genes at a stringent confidence of 0.8 or higher (STAR Methods). The regulatory network captures known connections among the key pluripotency regulators and target genes (e.g., $Esrrb \rightarrow Klf4$, $Sox2 \rightarrow Klf4$, $Esrrb \leftarrow Sox2$, $Esrrb \leftarrow Nanog$) and is comparable to the performance seen when using bulk RNA-seq data (STAR Methods), providing support to the relevance of the interactions.

MERLIN modules recapitulated the four patterns of expression from MEFs to ESCs (Figure 6B). We compared the expression patterns of genes in these modules in cells treated with A2S and each dual combination to identify key similarities and differences in expression pattern across these treatments to enable us to define the requirement of each component for successful reprogramming. We found that compared to A2S, the AS combination that omitted 2i continued to have a high expression of modules M1 through M4, (Figure 6B) which included MEF-specific genes, such as *Col5a1* and *Tagln*, even on day 6 (Figure 6C). This trend was even more obvious for genes that are aberrantly upregulated in the early days of reprogramming (module M8) and included genes such as *Oasl2* and *Egr1* (Figure 6C). For the cell cycle genes that are transiently downregulated in FBS reprogramming (modules M5 through M7), every dual combination could activate these genes (e.g., *Mcm6* and *Ccnb1*) (Figure 6C). The A2 combination was compromised in activating pluripotency genes (modules M9 through M11). Contrary to earlier reports, *Dot1L* inhibition does not increase *Cdh1* levels (Figure 6C) any more than the combinations that do not include this small molecule (Onder et al., 2012). Interestingly, the AS combination was as good at activating several genes of the pluripotency cluster as S2 but still resulted in a smaller number of iPSC colonies (Figure 6A), likely due to the continued expression of somatic genes because of the

failure to downregulate the MEF program. However, neither AS nor S2 was as good as A2S at activating pluripotency, suggesting synergistic effects of the triple combination.

We next used the high-confidence inferred regulatory network as a scaffold to estimate the relative strengths of the regulatory connections in each condition in order to identify which components of the network were present in each of the combinations (STAR Methods). Briefly, we used this network structure to fit a regression model for each gene in each condition and used the regression weight to estimate the edge strength (STAR Methods). The regression weight is reflective of the strength of the regulatory connection between a regulator and a target gene and provides information that might not be obvious from the absolute level of expression of a gene. Hence, although a gene node could be less expressed in one condition, its connections with regulators can be stronger if the expression of its regulators can explain its expression variation. We found that there were several sub-networks that had different strength in the dual combinations compared to the A2S combination. For the modules that do not turn off somatic genes or transiently upregulated gene expression, the connections between the regulators *Oas2l* and *Trim30*, or between *Col5a1* and *Col1a2* were retained only in the AS condition (lacking 2i) (Figure 6D; Figures S5B and S5C). For the upregulated genes, several connections surrounding *Nanog* (Figure 6E) were absent in the A2 condition, whereas those around *Epcam* and *Cdh1* were maintained (Figure S5E). For the more restricted pluripotency genes, S2 and AS differ in the kinds of connections that were made; for example, *Pou5f1* was better correlated with *Dppa3* in S2, whereas a greater proportion of cells expressed *Esrrb* with *Tdh* in the AS condition (Figure 6E). In the A2S condition, all these connections are stronger and new connections, such as the ones between *Dppa5a*, *Klf2*, and *Dppa3*, emerge (Figure 6E). The network surrounding DNA replication genes, such as *Mcm6*, remains strong in any of the dual combinations (Figure S5D).

Taken together, these results indicate that any combination of small molecules is able to overcome the senescence block faced by cells in FBS reprogramming. 2i is required for the downregulation of both MEF genes and transiently upregulated genes. Although A2 is sufficient to activate epithelial genes, SGC is required for the activation of pluripotency genes that emerge late. However, only in the presence of both AA and SGC, the rewiring of the pluripotency network is complete.

DISCUSSION

Reprogramming of somatic cells to iPSCs has been studied using bulk sequencing of reprogramming populations as well as those sorted on the basis of cell surface markers (Apostolou and Hochedlinger, 2013; Hussein et al., 2014; Lujan et al., 2015; Mikkelsen et al., 2008; O'Malley et al., 2013; Polo et al., 2012). These studies have led to an understanding of reprogramming trajectories taken by the majority of the cells. Here, by applying single-cell transcriptional sequencing, we find that there is overlapping expression of genes that was thought to be temporally activated (Apostolou and Stadtfeld, 2018; Brambrink et al., 2008; Stadtfeld and Hochedlinger, 2010) (Figure 7). Because most studies have focused on MEFs as the starting cell type, an important early event is the MET, a process amenable to acceleration (Liang et al., 2012; Zhou et al., 2017). Surprisingly, here

we find that mesenchymal genes are not all downregulated at the same stage. The frequently used marker of the epithelial transition *Cdh1* can be upregulated in cells that continue to express mesenchymal genes, such as *Twist1*. Thus, our study demonstrates that in order to increase the rate of reprogramming, it may be worthwhile to focus on other small molecules that can reliably and consistently shut down mesenchymal gene expression. We also find that another epithelial gene, *Epcam*, can be downregulated in a few cells if it is not co-expressed with other pluripotency genes. This result mirrors the recent finding that the reliability of *Epcam* as a marker is enhanced by co-expression with *SSEA1* and without *Scal* (Schwarz et al., 2018). Such co-expression is valuable for sustaining the expression not only of *Epcam* but also of the pluripotency factors, which can be activated in isolated cells even in FBS reprogramming. This includes *Sox2*, which was identified by candidate sequencing to start a cascade of deterministic pluripotency (Buganim et al., 2012). We find that the level of *Sox2* expression is higher when found in cells also expressing *Dppa5a* and *Utf1*.

It has also been noted that somatic cell nuclear transfer tends to activate the *Oct4* locus earlier than has ever been observed for reprogramming (Bhutani et al., 2010). One reason for this may be that genes such as *Ehf* that are transiently upregulated may have a role in restructuring the gene networks in a way that makes the next step conducive to reach the pluripotent state. Co-opting basic translational machinery (Brumbaugh et al., 2018), such as the regulation of *Eif4a1*, a device used by cancer cells (Modelska et al., 2015; Wolfe et al., 2014), may also be important for reprogramming, increasing the parallels between cancer and pluripotency.

The small molecules that we have used contribute differentially to the pluripotency network. One way that any combination of the small molecules works is by decreasing the number of cells that display senescence gene expression. A greater number of cycling cells increases reprogramming efficiency (Hanna et al., 2009; Marión et al., 2009; Ruiz et al., 2011; Utikal et al., 2009). Previous studies have genetically modulated the levels of cell cycle control genes, such as *p53*, to affect this change (Hanna et al., 2009; Marión et al., 2009; Utikal et al., 2009). We now provide a chemical method that can be transiently applied to overcome the senescence barrier. By applying a network analysis method, we also identify the connections of these molecules. We find that the addition of *2i* suppresses some aberrantly expressed genes and allows for faster downregulation of MEF markers. *AA* and *SGC* work together to reinforce the pluripotency program. The modulation of the dose and timing of these factors could be harnessed in the future to rationally enhance reprogramming efficiency further.

STAR ★ METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rupa Sridharan (rsridharan2@wisc.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Primary MEFs—Male and female MEFs were isolated from E13.5 time-mated embryos as described in Tran et al., (2015) from reprogrammable mice (Sridharan et al., 2013) homozygous for the Oct4-2A-Klf4-2A-IRES-Sox2-2A-c-Myc (OKSM) transgene at the Col1a1 locus and either homozygous or heterozygous for the reverse tetracycline transactivator (rtTA) allele at the Rosa26 locus. MEFs were maintained in MEF media (DMEM, 10% FBS, L-glutamine, Pen/Strep, NEAA, 2-mercaptoethanol). Mice were maintained according to protocol approved by the UW-Madison IACUC.

Mouse Embryonic Stem Cells—Murine ESCs (V6.5 line, male) were maintained in ESC media (knockout DMEM, 15% FBS, L-glutamine, Pen/Strep, NEAA, 2-mercaptoethanol, and leukemia inhibitory factor) on a feeder layer of irradiated MEFs.

METHOD DETAILS

Reprogramming—MEFs were thawed and maintained in ESC media for 2 days before plating. On day -1, 5000 cells were plated onto 0.1% gelatin-coated coverslips in 6-well plates. 24 hours post-plating (day 0), cells were counted to determine the number of cells adhered to the coverslip. This number was used to calculate reprogramming efficiency (Figure 1A). On day 0, MEFs were treated with 2 µg/mL doxycycline to induce OKSM expression and irradiated MEFs were added. For A2S and dual combination reprogramming, 50 µg/mL of ascorbic acid (Sigma A8960) and 5 µM SGC0946 (ApexBio A4167) were added on Day 0. 3 µM CHIR-99021 (Stemgent 04-0004-10) and 1 µM PD-0325901 (Stemgent 04-0006-10) (2i) were added 12 hours post-doxycycline induction. Media containing doxycycline and small molecules was changed every two days. Efficiency of reprogramming was determined by Nanog immunofluorescence either on day of fixing as indicated, or after withdrawal of doxycycline and small molecules for an additional 4 days. Two or more biological replicates were performed for each set of reprogramming experiments. iPSC colonies were isolated from reprogramming culture on day 6 and maintained in either regular ESC media or A2S-containing ESC media on irradiated MEFs for several passages. For single-cell reprogramming, MEFs were infected with pMX-tdTomato retrovirus and FACS-sorted into 96-well plates as single tdTomato+ cell per well on irradiated MEFs. FBS and A2S reprogramming were performed as above. Doxycycline and chemicals were removed on day 11 and AP-positive wells were scored on day 15.

Immunofluorescence—Immunofluorescence was performed as described in Sridharan et al., (2009). Briefly, cells were fixed with 4% paraformaldehyde-PBS, followed by permeabilization with 0.5% TritonX-PBS and stained with antibodies in blocking buffer (1X PBS with 5% normal donkey serum, 0.2% Tween-20, and 0.2% fish skin gelatin). Nanog (CosmoBio RCAB0002P), Dppa4 (ThermoFisher Scientific PA5-47530), Cdh1 (Ebioscience 14-3249-82), and Twist1 (Novus Biologicals, NBP2-37364SS) antibodies were used at 1:100 dilution, while Ki67 (Abcam ab15580) was used at 1:200. Imaging and colony counts were performed on Nikon Eclipse Ti using NIS Elements software.

Flow cytometry—MEFs were induced to reprogram in the A2S condition as above, but without irradiated MEFs. On day 3, cells were harvested with trypsin, resuspended to a

single-cell suspension and stained with Epcam antibody (CD326) – PE conjugated (BD PharMingen 563477) at 1 μ l per 5 \times 10⁵ cells for 1 hour before being sorted using BD FACS Aria II. Epcam+ cells were re-plated and allowed to reprogram for an additional 3 days before another FACS was performed on day 6.

siRNA transfection—siRNA purchased from Integrated DNA Technologies or GE Life Science were transfected using Dharmafect reagent (GE Life Sciences) according to manufacturer's instructions. For Eif4a1 and Ehf knockdown experiments, siRNA was added on the day of plating at 0.5nM. siRNA was added every 48 hours and concentration was increased gradually up to 40nM to account for increasing cell numbers. For the Eif4a1 experiment, live cell counts were performed every day using Trypan Blue exclusion. For the Phlda2 experiments, siRNA was added at days 4 and 5 at 50nM and 75nM respectively. Two siRNAs were combined for Phlda2. The following siRNAs were used: Eif4a1 siRNA #1 mm.Ri.Eif4a1.13.1, Eif4a1 siRNA#2 mm.Ri.Eif4a1.13.2, Ehf siRNA #1 mm.Ri.Ehf.13.1, Ehf siRNA #2 mm.Ri.Ehf.13.2, Phlda2 siRNA#1 mm.Ri.Phlda2.13.1, Phlda2 siRNA#2 mm.Ri.Phlda2.13.2, Non-Targeting siRNA D-001810-01-50. To evaluate knockdown efficiency, qRT-PCR was performed using primers listed in Table S2.

Single-Cell RNA-sequencing—To ensure optimal viability of cells during droplet formation, cells were washed once with DPBS, followed by a media change 12 hours prior to single cell isolation. On the day of single cell isolation, cells in 6-well plates were washed five times with DPBS, dislodged with 1 mL 0.25% trypsin-EDTA and neutralized with 1 mg/ml trypsin inhibitor (Sigma Aldrich T6522). Cells were filtered through a 35 μ m nylon mesh (Corning 352235) and centrifuged at 300 \times g for 3 min. Pelleted cells were gently washed with DPBS and pelleted again at 300 \times g, 3 min, RT. Cells were resuspended in 1 mL 0.1% BSA-DPBS (ThermoFisher 15260037) and gently pipetted 20-50 times. Single-cell suspension was confirmed under the microscope and cell concentration and viability were measured on a Bio-Rad TC20. Cells were diluted to a final concentration of about 2500 cells/uL in 0.1% BSA-DPBS.

Single-Cell Isolation and Library Preparation—Single-cell encapsulation was performed on a ddSEQ Single-Cell Isolator (BioRad 12004336), with reagents provided in the SureCell WTA 3' Library Prep Kit (Illumina 20014279), according to manufacturer's instructions. Briefly, approximately 12,500 cells in single-cell resuspension were mixed with Cell Enzyme Mix containing reverse transcriptase. A ddSEQ cartridge was primed with Priming Solution before Barcode Suspension Mix, Cell Suspension Mix, and encapsulation oil were loaded onto the cartridge and into the Isolator. Generated single-cell droplets were transferred to a pre-chilled plate and run on a thermal cycler to begin reverse transcription of mRNA. Droplets were subsequently disrupted, and first-strand library cDNA was used for second strand synthesis. Quality of pre-amplified libraries was confirmed on High Sensitivity DNA Chips on the Agilent Technology 2100 Bioanalyzer. Libraries with a minimum of 1.8 ng DNA were tagged with DNA adapters from the SureCell WTA 3' Library Prep Kit and amplified.

Next-Generation Sequencing and Genome Alignment—Between 7 and 9 libraries were multiplexed per lane on an Illumina HiSeq2500 Rapid Run (2×75), with a mean of over 280 million reads per lane. Fastq files (bcl2fastqv2.19) were generated, either through Illumina BaseSpace – the Illumina computing environment for sequencing data analysis – or through the University of Wisconsin-Madison Bioinformatics Resource Center, and uploaded to Illumina BaseSpace. Sequences were aligned to *Mus musculus* 10 (mm10) genome using Spliced Transcripts Alignment to a Reference (STAR), available through the SureCell RNA Single Cell App v1.1.0 on BaseSpace. On average, 85.48% reads per sample aligned to the genome, and 2.06% reads per sample aligned to abundant features (mitochondria, small non-coding RNA, ribosomal RNA). A unique molecular identifier (UMI) per cell plot was generated using BaseSpace, which indicates the total number passing filter. A drop in the knee plot indicated a transition to empty beads, in which a cell barcode contained low UMI counts. This drop serves as the threshold for calling cells that pass the sample-specific knee filter, and all subsequent analyses were performed with cells passing this filtering step. In total, we isolated 8,334 cells, and on average, 260 cells passed knee filter per sample, with a median of 53,497 genic reads, 13,100 genic UMIs and 4,274 genes detected per cell passing filter. Several libraries were re-sequenced in order to achieve a sequencing a depth of approximately 50,000 reads per cell for each sample.

Bioinformatic analysis

t-SNE Clustering: We used Monocle2 v2.6.3 on R version 3.4.3 (Kite-Eating Tree) <http://cole-trapnell-lab.github.io/monocle-release/docs/> (Qiu et al., 2017b, 2017a) to analyze the data obtained after alignment. We initially plotted the distribution of UMI counts within each cell and filtered out any cells with UMI counts outside a range determined by:

$$10^{\wedge}(\text{mean}(\log_{10}(\text{Total number of UMIs within all cells of dataset})) \\ + / - 2 * \text{standard deviation}(\log_{10}(\text{Total number of UMIs within all cells of dataset})))$$

Out of the 4,374 cells passing filter, 4,167 cells were within the optimal UMI range and used for downstream Monocle analysis. Genes that were not expressed in at least 1 cell were excluded from analysis. Principal component analysis (PCA) was then performed to identify the variance explained by each component of the cell dataset (cds).

- (1) `cds <- detectGenes(cds, min_expr = 0.1)`
- (2) `fData(cds)$use_for_ordering <- fData(cds)$num_cells_expressed > 0.1 * ncol(cds)`
- (3) `plot_pc_variance_explained(cds, return_all = F)`

We reduced the number of dimensions to the number of PC components that explained the most variance, before the PCA components began to level off. Together, these components explained at least 50% of the variance for each dataset.

In order to remove the irradiated feeder MEFs from our analysis, we performed a t-distributed stochastic neighbor embedding (t-SNE clustering) using data from ESCs grown on a feeder MEF layer in FBS and A2S. Irradiated MEFs formed a separate cluster and

could be identified by expression of MEF markers in the ESC samples. We used the cell IDs of the FBS-ESCs that were found with these irradiated MEFs to identify where these cells are located in an initial clustering of all FBS, A2S reprogramming, MEF, and FBS-ESCs samples. The cells associated with this cluster were removed, resulting in the total of 4,374 cells that were used in the Monocle pipeline. A table with the Cell IDs that were removed from the analysis is available on GEO under entry GSE108222.

t-SNE Cluster Analysis: To identify genes important for defining clusters within the MEF reprogramming, DE analysis was performed between all 14 clusters within the t-SNE plot of Figure 2A.

To determine the distribution of cells from each sample that fall into each cluster, phenotypic data (cell barcode ID, sample, pseudotime, cluster number) was extracted for each cluster and sample. The composition of a cluster or sample was then calculated by percentage or mean of the population.

The top 10% of DEGs from this list were used in generating a heatmap to visualize percentage of cells within each cluster and sample that express these genes. Gene patterns were identified by k-means clustering into 15 groups using Cluster3 software and visualized by Java TreeView (de Hoon et al., 2004; Saldanha, 2004)(Figure 2C). Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al., 2009) was used to functionally annotate groups of genes from heatmaps.

Generating pseudotime trajectory: To order cells by pseudotime, the gene expression of each cell has to be compared to a standard. We chose to use the top 5% of differentially expressed genes (DEG) between seven t-SNE clusters from only the A2S reprogramming samples and the MEFs and ESCs. This ensured that we were not comparing only established cell types or gene expression at specific time points. Using MEFs as the starting point, Monocle defined a pseudo-reprogramming time trajectory, termed pseudotime, where cells are linearly ordered relative to their progress or change in gene expression relative to the starting population. Lengths of the trajectory between each branchpoint were used to define state by the Monocle algorithm.

- (4) `diff_test_resClusterDE <- differentialGeneTest(cds, fullModelFormulaStr = "~Cluster," cores = detectCores())`
- (5) `SetOrdering(cds, ordering_genes = Top 5% DE Genes)`
- (6) `reduceDimensions(cds)`
- (7) `orderCells(cds)`
- (8) `plot_cell_trajectory(cds, color_by = "Phenotype Data")`

Branched expression analysis modeling (BEAM) was performed to identify genes involved in the decision-making process of progressing along the trajectory or to a branch. Genes involved in BEAM with a q-value less than $1e-40$ were then plotted along pseudotime to visualize relative expression of genes as cells progress to either branchpoint or toward the end of the trajectory.

- (9) `BEAM_cds <- BEAM(cds, branch_point = 1/2, cores = detectCores())`
- (10) `plot_genes_branched_heatmap(cds[row.names(subset(BEAM_cds, qval < 1e-40)),], branch_point = 1, num_clusters = 10, cores = detectCores(), use_gene_short_name = T, show_rownames = T)`
- (11) `plot_genes_branched_pseudotime(cds_subset, branch_point = 2, color_by = "Cluster," ncol = 1)`
- (12) `diff_test_res_PseudotimeDE <- differentialGeneTest(cds, fullModelFormulaStr = "~Pseudotime," cores = detectCores())`

Co-expression Analysis: To determine how a pair of two different genes are co-expressed within the cell population, Monocle's cell type hierarchy function was implemented. Cells expressing a particular gene were identified using the following command:

- (13) `GeneName_id <- row.names(subset(fData(cds), gene_short_name == "Gene Name"))`
- (14) `cth <- newCellTypeHierarchy()`
- (15) `cth <- addCellType(cth, "Gene Name 1 Positive" classify_func = function(x) {x[GeneName1_id,] > 0 })`
- (16) `cth <- addCellType(cth, "Gene Name 2 Positive," classify_func = function(x) {x[GeneName2_id,] > 0 })`
- (17) `cds <- classifyCells(cds, cth, 0.1)`

Visualizing the t-SNE plot based on cell type will identify cells that are positive for one of the genes of interest, those that are double positive (labeled Ambiguous), and those that are double negative (labeled Unknown). The phenotype data table contains information on cell type, which allows us to determine how prevalent each cell type is within each cluster, sample, etc. To visualize how different genes are expressed in cells that are known to be positive for a particular gene, we also generated violin plots. Note that due to the sequencing depth of single cell RNA-Seq (also known as "Dropout") co-expression may be underestimated. After defining a cell type using the above command (Gene 1+ cells), we use the following code to produce violin plots:

- (18) `Gene1_C1 <- cds[pData(cds)$CellType == "Gene1+" & pData(cds)$Cluster == "1"]`
- (19) `Gene1_C1_table <- as.data.frame(pData(Gene1_C1))`
- (20) `Gene1_C1_table$Identifier <- row.names(Gene1_C1_table)`
- (21) `Gene1_C1_Id <- row.names(Gene1_C1_table)`
- (22) `cds_log <- log(exprs(cds)+1)`
- (23) `t <- as.data.frame(cds_log)`
- (24) `Gene1_C1_counts <- t[, colnames(t) %in% Gene1_C1_Id]`
- (25) `Gene1_C1_counts_all <- as.data.frame(t(Gene1_C1_counts))`

- (26) `Gene1_C1_counts_all$CellType < - "Gene1+ C1"`
- (27) `ggplot(Gene1_counts, aes(x = CellType, y = Gene2)) + geom_violin()`

Constructing gene regulatory network (MERLIN): MERLIN is based on a probabilistic graphical model representation of a regulatory network and uses a probabilistic graph structure prior to enable genes in the same module to have similar but not identical regulators. To infer networks, we used the top 10% of differentially expressed genes identified by Monocle, and added a list of 445 known transcription factors, signaling proteins, and chromatin remodelers, as well as genes known to be involved in early stem cell state specification, which resulted in 2,100 genes in 4,633 cells. We applied MERLIN in a stability selection framework. Briefly, we created 100 subsamples by randomly selecting 2,317 cells for each, and ran MERLIN independently on each subsample. As initial cluster assignments for genes, we used k-means with 10 clusters. We used the following default options for running MERLIN: -5 for sparsity, 4 for modularity prior and 0.6 for redefining modules. The outputs of MERLIN comprise a regulatory network as well as module assignments for input genes. We next obtained consensus networks and consensus modules as described in Chasman et al., (2016). Each edge in the consensus network has a confidence value that indicates the percentage of subsamples in which that edge was inferred. Consensus modules are defined by applying hierarchical clustering to a co-clustering matrix (which is the fraction of subsamples' where a pair of genes were in the same MERLIN module). We identified a total of 15 modules with at least 5 genes spanning 291 genes. We associated each consensus module with regulators based on a significant overlap (hypergeometric test, FDR < 0.05) of regulator targets from the 80% confidence network. Furthermore, we assessed the inferred modules for enrichments of Gene Ontology processes, and found 12 of the 15 consensus modules to be enriched.

Visualizing inferred networks for each module: For a given module, we selected all incoming edges to that module from our 80% confidence network. Next, we selected cells from each condition (A2, S2, AS, and A2S, day 4 or day 6) and applied a linear regression model to predict the expression of the target gene as a function of its regulators in the 80% confidence network. We visualized these using the program Cytoscape (Shannon et al., 2003). Briefly, the edge color corresponds to regression coefficient of that regulator for the target (from -0.5 (blue) to 0 (white) to 0.5 (red)). Edge width corresponds to edge confidence (from 80% (1) to 100% (5)). Node color corresponds to percentage of cells in which that gene was expressed (from 0% (white) to 100% (green)). Node border is pink if the gene is in the given module, and gray if it is not.

Dual Combinations: A Monocle cell dataset was created using single-cell data from all dual combination reprogramming experiments as well as the data from A2S days 4 and 6. Jitter plots were generated in Monocle to illustrate expression of specific genes in each different condition. The MERLIN algorithm was applied to the dual combination and A2S RNA-seq data to generate regulatory networks for the defined modules in each reprogramming condition.

QUANTIFICATION AND STATISTICAL ANALYSIS

Information on replicates for each experiment can be found in the figure legends. p and q values for differentially expressed genes of single cell RNA-sequencing data were calculated from likelihood ratio tests on the parallel arrays of models generated through monocle.

DATA AND SOFTWARE AVAILABILITY

All single-cell RNA-seq data have been submitted to the National Center for Biotechnology Information Gene Expression Omnibus database and can be accessed at GEO: GSE108222.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by NIH-NIGMS R01GM113033 and the Shaw Scientist Award to R.S. and NIH-NIGMS R01GM117339 to S.R. K.A.T. was supported by the Advanced Opportunity Fellowship from UW-Madison and NSF GRFP-DGE 1256259, S.J.P. by NHGRI 5T32HG002760, N.Z.Z. by the SCRMC fellowship from UW-Madison and American Heart Association 18PRE34080337, S.G.M.C. by the Data Sciences Initiative of UW-Madison and NIGMS T32GM007133, and A.S.Z. by NIGMS T32GM008688. We thank Mike Ducat of Biorad and Alasdair Reid and Melissa Pourpak of Illumina for providing the ddSeq instrument and experimental support; Molly Zeller and Joshua Hyman of the UW-Madison Biotechnology Center for assisting with sequencing and Profs. James Thomson and Krishanu Saha and members of the Sridharan lab for critical reading of the manuscript.

REFERENCES

- Apostolou E, and Hochedlinger K (2013). Chromatin dynamics during cellular reprogramming. *Nature* 502, 462–471. [PubMed: 24153299]
- Apostolou E, and Stadtfeld M (2018). Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Curr. Opin. Genet. Dev.* 52, 77–85. [PubMed: 29925040]
- Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V, Walsh M, et al. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.* 23, 2134–2139. [PubMed: 19696146]
- Bar-Nur O, Brumbaugh J, Verheul C, Apostolou E, Pruteanu-Malinici I, Walsh RM, Ramaswamy S, and Hochedlinger K (2014). Small molecules facilitate rapid and synchronous iPSC generation. *Nat. Methods* 11, 1170–1176. [PubMed: 25262205]
- Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, and Blau HM (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463, 1042–1047. [PubMed: 20027182]
- Blaschke K, Ebata KT, Karimi MM, Zepeda-Martínez JA, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A, et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222–226. [PubMed: 23812591]
- Brambrink T, Foreman R, Welstead GG, Lengner CJ, 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. [PubMed: 18371436]
- Brumbaugh J, Di Stefano B, Wang X, Borkent M, Forouzmand E, Clowers KJ, Ji F, Schwarz BA, Kalocsay M, Elledge SJ, et al. (2018). Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. *Cell* 172, 106–120.e21. [PubMed: 29249356]
- Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, 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. [PubMed: 22980981]

- Buganim Y, Faddah DA, and Jaenisch R (2013). Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* 14, 427–439. [PubMed: 23681063]
- Chasman D, Walters KB, Lopes TJS, Eisfeld AJ, Kawaoka Y, and Roy S (2016). Integrating Transcriptomic and Proteomic Data Using Predictive Regulatory Network Models of Host Response to Pathogens. *PLoS Comput. Biol.* 12, e1005013. [PubMed: 27403523]
- de Hoon MJL, Imoto S, Nolan J, and Miyano S (2004). Open source clustering software. *Bioinformatics* 20, 1453–1454. [PubMed: 14871861]
- Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, Li W, Weng Z, Chen J, Ni S, et al. (2010). Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* 6, 71–79. [PubMed: 20036631]
- Golipour A, David L, Liu Y, Jayakumaran G, Hirsch CL, Trcka D, and Wrana JL (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* 11, 769–782. [PubMed: 23217423]
- Guo S, Zi X, Schulz VP, Cheng J, Zhong M, Koochaki SHJ, Megyola CM, Pan X, Heydari K, Weissman SM, et al. (2014). Nonstochastic reprogramming from a privileged somatic cell state. *Cell* 156, 649–662. [PubMed: 24486105]
- Guo L, Lin L, Wang X, Gao M, Cao S, Mai Y, Wu F, Kuang J, Liu H, Yang J, et al. (2019). Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell RNA-Seq. *Mol. Cell* 73, 815–829.e7. [PubMed: 30772174]
- Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creighton MP, van Oudenaarden A, and Jaenisch R (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 462, 595–601. [PubMed: 19898493]
- Hore TA, von Meyenn F, Ravichandran M, Bachman M, Ficiz G, Oxley D, Santos F, Balasubramanian S, Jurkowski TP, and Reik W (2016). Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. *Proc. Natl. Acad. Sci. USA* 113, 12202–12207. [PubMed: 27729528]
- Huang W, Sherman BT, and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. [PubMed: 19131956]
- Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, and Melton DA (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* 26, 795–797. [PubMed: 18568017]
- Hussein SMI, Puri MC, Tonge PD, Benevento M, Corso AJ, Clancy JL, Mosbergen R, Li M, Lee D-S, Cloonan N, et al. (2014). Genomewide characterization of the routes to pluripotency. *Nature* 516, 198–206. [PubMed: 25503233]
- Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, Di Giorgio FP, 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. [PubMed: 19818703]
- Ichida JK, Tcw J, Williams LA, Carter AC, Shi Y, Moura MT, Ziller M, Singh S, Amabile G, Bock C, Umezawa A, Rubin LL, Bradner JE, Akutsu H, Meissner A, and Eggan K (2014). Notch inhibition allows oncogene-independent generation of iPS cells. *Nat. Chem. Biol.* 10, 632–639. [PubMed: 24952596]
- Jackson SA, Olufs ZPG, Tran KA, Zaidan NZ, and Sridharan R (2016). Alternative Routes to Induced Pluripotent Stem Cells Revealed by Reprogramming of the Neural Lineage. *Stem Cell Reports* 6, 302–311. [PubMed: 26905202]
- Kim DH, Marinov GK, Pepke S, Singer ZS, He P, Williams B, Schroth GP, Elowitz MB, and Wold BJ (2015). Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell* 16, 88–101. [PubMed: 25575081]
- Li H, Collado M, Villasante A, Strati K, Ortega S, Cañamero M, Blasco MA, and Serrano M (2009). The *Ink4/Arf* locus is a barrier for iPS cell reprogramming. *Nature* 460, 1136–1139. [PubMed: 19668188]
- 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. [PubMed: 20621050]

- Liang G, He J, and Zhang Y (2012). Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat. Cell Biol.* 14, 457–466. [PubMed: 22522173]
- Lujan E, Zunder ER, Ng YH, Goronzy IN, Nolan GP, and Wernig M (2015). Early reprogramming regulators identified by prospective isolation and mass cytometry. *Nature* 521, 352–356. [PubMed: 25830878]
- 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. [PubMed: 19765992]
- Marión RM, Strati K, Li H, Murga M, Blanco R, Ortega S, Fernandez-Capetillo O, Serrano M, and Blasco MA (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPSC cell genomic integrity. *Nature* 460, 1149–1153. [PubMed: 19668189]
- Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A, and Stunnenberg HG (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590–604. [PubMed: 22541430]
- Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, and Meissner A (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454, 49–55. [PubMed: 18509334]
- Modelska A, Turro E, Russell R, Beaton J, Sbarrato T, Spriggs K, Miller J, Gräf S, Provenzano E, Blows F, et al. (2015). The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. *Cell Death Dis.* 6, e1603–e1612. [PubMed: 25611378]
- Mor N, Rais Y, Sheban D, Peles S, Aguilera-Castrejon A, Zviran A, Elinger D, Viukov S, Geula S, Krupalnik V, et al. (2018). Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. *Cell Stem Cell* 23, 412–425. [PubMed: 30122475]
- Nefzger CM, Rossello FJ, Chen J, Liu X, Knaupp AS, Firas J, Paynter JM, Pflueger J, Buckberry S, Lim SM, et al. (2017). Cell Type of Origin Dictates the Route to Pluripotency. *Cell Rep.* 21, 2649–2660. [PubMed: 29212013]
- O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR, Linnarsson S, and Kaji K (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* 499, 88–91. [PubMed: 23728301]
- Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483, 598–602. [PubMed: 22388813]
- Papp B, and Plath K (2013). Epigenetics of reprogramming to induced pluripotency. *Cell* 152, 1324–1343. [PubMed: 23498940]
- Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, 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. [PubMed: 23260147]
- Qiu X, Hill A, Packer J, Lin D, Ma Y-A, and Trapnell C (2017a). Singlecell mRNA quantification and differential analysis with Census. *Nat. Methods* 14, 309–315. [PubMed: 28114287]
- Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, and Trapnell C (2017b). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. [PubMed: 28825705]
- Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502, 65–70. [PubMed: 24048479]
- Ruiz S, Panopoulos AD, Herrerías A, Bissig K-D, Lutz M, Berggren WT, Verma IM, and Izpisua Belmonte JC (2011). A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr. Biol.* 21, 45–52. [PubMed: 21167714]
- Salas M, John R, Saxena A, Barton S, Frank D, Fitzpatrick G, Higgins MJ, and Tycko B (2004). Placental growth retardation due to loss of imprinting of Phlda2. *Mech. Dev.* 121, 1199–1210. [PubMed: 15327781]
- Saldanha AJ (2004). Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20, 3246–3248. [PubMed: 15180930]

- Samavarchi-Tehrani P, Golipour A, David L, Sung H-K, Beyer TA, Datti A, Woltjen K, Nagy A, and Wrana JL (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 7, 64–77. [PubMed: 20621051]
- Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, et al. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* 176, 928–943.e22. [PubMed: 30712874]
- Schwarz BA, Cetinbas M, Clement K, Walsh RM, Cheloufi S, Gu H, Langkabel J, Kamiya A, Schorle H, Meissner A, et al. (2018). Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming. *Cell Stem Cell* 23, 289–305.e5. [PubMed: 30017590]
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. [PubMed: 14597658]
- Shi Y, Despons C, Do JT, Hahm HS, Schöler HR, and Ding S (2008). Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* 3, 568–574. [PubMed: 18983970]
- Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, and Smith A (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* 6, e253. [PubMed: 18942890]
- Sridharan R, Tchieu J, Mason MJ, 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. [PubMed: 19167336]
- Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1 γ in reprogramming to pluripotency. *Nat. Cell Biol* 15, 872–882. [PubMed: 23748610]
- Stadtfield M, and Hochedlinger K (2010). Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* 24, 2239–2263. [PubMed: 20952534]
- Stadtfield M, Maherali N, Breault DT, and Hochedlinger K (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* 2, 230–240. [PubMed: 18371448]
- 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. [PubMed: 16904174]
- Tran KA, Jackson SA, Olufs ZPG, Zaidan NZ, Leng N, Kendziorski C, Roy S, and Sridharan R (2015). Collaborative rewiring of the pluripotency network by chromatin and signalling modulating pathways. *Nat. Commun* 6, 6188. [PubMed: 25650115]
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386. [PubMed: 24658644]
- Utikal J, Polo JM, Stadtfield M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, and Hochedlinger K (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* 460, 1145–1148. [PubMed: 19668190]
- Vidal SE, Amlani B, Chen T, Tsirigos A, and Stadtfield M (2014). Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. *Stem Cell Reports* 3, 574–584. [PubMed: 25358786]
- Williams-Hill DM, Duncan RF, Nielsen PJ, and Tahara SM (1997). Differential expression of the murine eukaryotic translation initiation factor isogenes eIF4A(I) and eIF4A(II) is dependent upon cellular growth status. *Arch. Biochem. Biophys.* 338, 111–120. [PubMed: 9015395]
- Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* 513, 65–70. [PubMed: 25079319]
- Zhao X-Y, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Guo C-L, Ma Q-W, Wang L, Zeng F, and Zhou Q (2009). iPSC cells produce viable mice through tetraploid complementation. *Nature* 461, 86–90. [PubMed: 19672241]

- Zhao T, Fu Y, Zhu J, Liu Y, Zhang Q, Yi Z, Chen S, Jiao Z, Xu X, Xu J, et al. (2018). Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. *Cell Stem Cell* 23, 31–45.e7. [PubMed: 29937202]
- Zhou Z, Yang X, He J, Liu J, Wu F, Yu S, Liu Y, Lin R, Liu H, Cui Y, et al. (2017). Kdm2b Regulates Somatic Reprogramming through Variant PRC1 Complex-Dependent Function. *Cell Rep.* 21, 2160–2170. [PubMed: 29166607]
- Zunder ER, Lujan E, Goltsev Y, Wernig M, and Nolan GP (2015). A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* 16, 323–337. [PubMed: 25748935]

Highlights

- Ascorbic acid, 2i, and Dot1l inhibitor rapidly and robustly convert MEFs to iPSCs
- Pluripotent and somatic gene expression is compatible in single cells
- Continued mesenchymal gene expression inhibits successful reprogramming
- Network analysis reveals distinct roles for each chemical component on gene expression

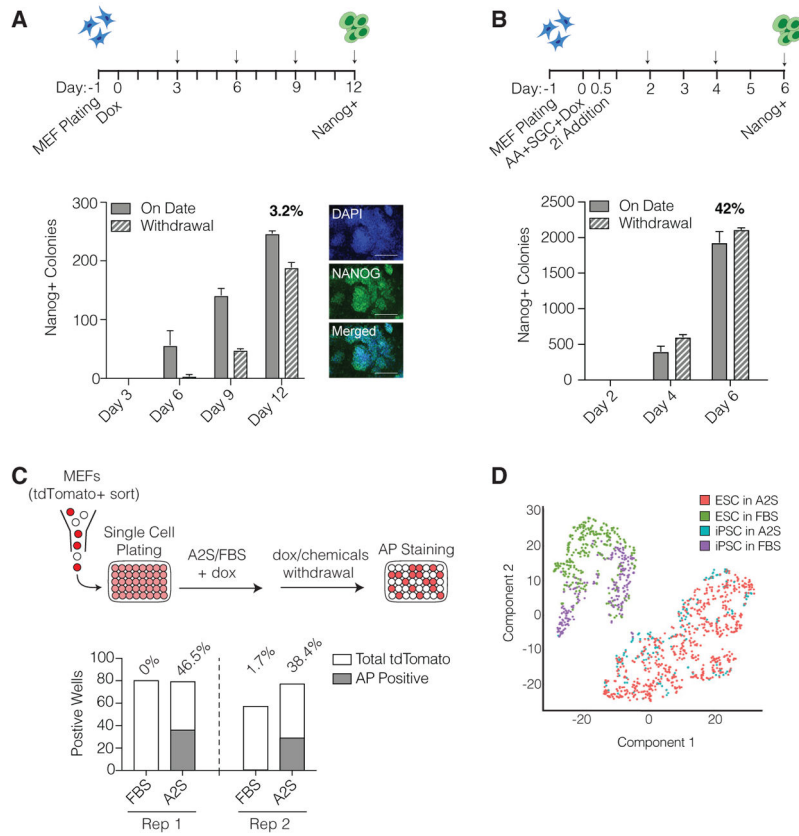


Figure 1. Combining Epigenomic and Signaling Modifiers Leads to High-Efficiency Generation of Bona Fide iPSCs

(A) Top: schematic of FBS reprogramming experiment. Cells were harvested and immunofluorescence performed on the days indicated by the arrows. Bottom: number of NANOG+ colonies counted at each indicated time point (on date) or after 4 additional days after doxycycline (dox) was removed (withdrawal). Bars represent SD between two replicate samples. Right panel – immunofluorescence images of NANOG. Scale bar, 250 μ m.

(B) Top: schematic of A2S reprogramming experiment. Cells were harvested and immunofluorescence performed on the days indicated by the arrows. Bottom: number of NANOG+ colonies counted at each indicated time point (On Date) or after 4 additional days after dox was removed (withdrawal). Bars represent SD between two replicate samples.

(C) Top: schematic of single-cell reprogramming experiment. MEFs infected with tdTomato virus were sorted and plated in a 96-well plate. Dox-independent colonies were stained with alkaline phosphatase (AP). Bottom: number of AP+ wells observed in each condition. Percentages indicate how many of the wells were AP+ out of the total number of wells with tdTomato+ cells. Data from two independent experiments are presented.

(D) Monocle clustering plot showing ESCs or iPSCs cultured in A2S or FBS media.

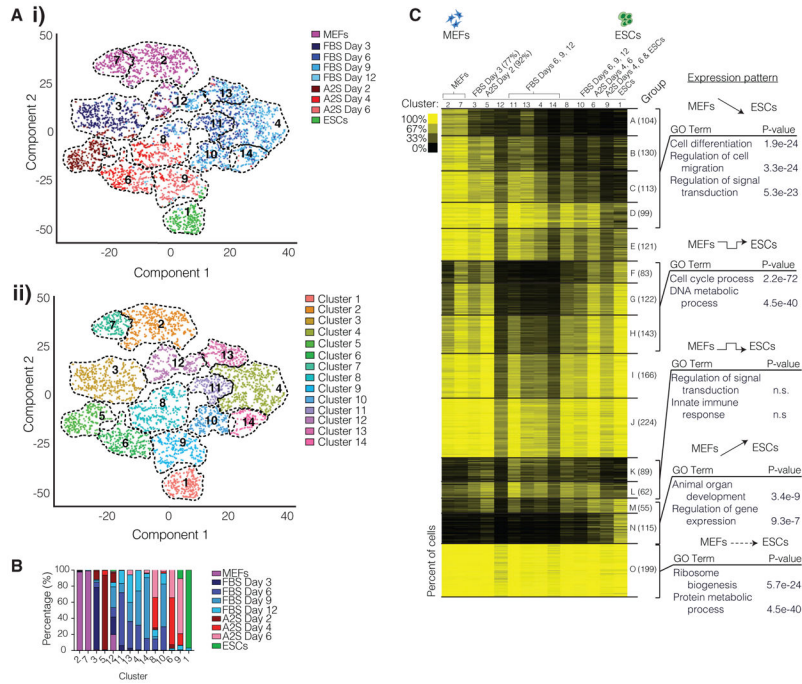


Figure 2. A2S Accelerates FBS Reprogramming

(A) Monocle t-SNE plots showing clustering of reprogramming cells from FBS and A2S, MEFs, and FBS-cultured ESCs. Samples were grouped into 14 clusters. Cells colored by sample (i) and cluster (ii).

(B) Graph showing the composition of each cluster from Figure 2A by sample.

(C) Heatmap representing the percentage of cells expressing the top 10% differentially expressed genes that define the 14 t-SNE clusters in Figure 2A. Each row represents a single gene. Genes were grouped by k-means into 15 groups labeled A to O, and the number of genes within each group are in parentheses. The 14 t-SNE clusters labeled 1–14 are presented in columns approximating their similarity to ESCs. Significant gene ontology terms associated with a specific group are labeled on the right. n.s., not significant. Arrows indicate pattern of expression change between MEFs and ESCs.

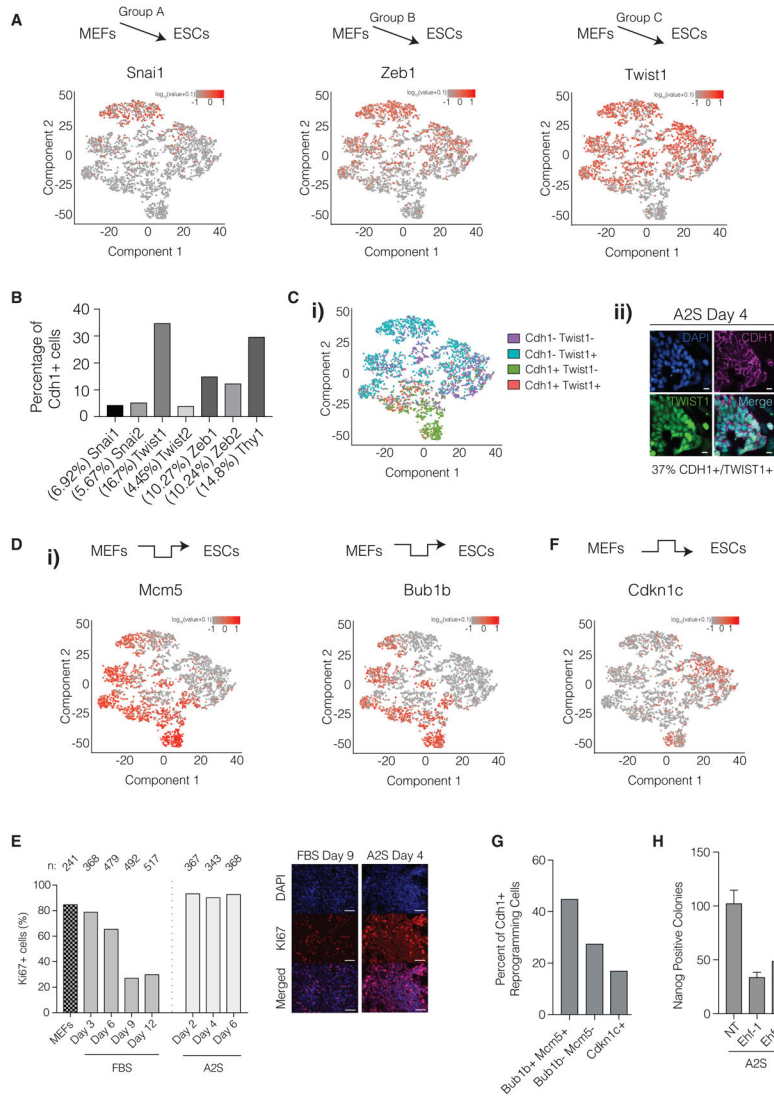


Figure 3. Reprogramming-Specific Gene Expression Patterns Are Important for Conversion to iPSCs

(A) t-SNE plots based on Figure 2A highlighting the expression of MEF-associated mesenchymal genes that are downregulated as cells transition from MEFs to pluripotency. Top schematic indicates the pattern of expression.

(B) Percentage of Cdh1+ cells that also co-express the indicated MEF genes on the x axis. The percentage of MEF gene-expressing cells that express Cdh1 is presented in brackets on the x axis. Note that because of the limit of detection of single-cell transcriptional analysis, co-expression may be underestimated.

(C) (i) t-SNE plots based on Figure 2A illustrating co-expression of Cdh1 with Twist1. Note that because of the limit of detection of single-cell transcriptional analysis, co-expression may be underestimated. (ii) Immunofluorescent staining for CDH1 and TWIST1. Percentage of CDH1+/TWIST1+ colonies on A2S day 4 shown below image. Scale bar, 10 μ m.

(D) t-SNE plots based on Figure 2A highlighting the expression of DNA replication and cell-cycle-associated genes. Top schematic indicates the pattern of expression.

(E) Left: percentage of cells that are Ki67+ at each indicated reprogramming time point in FBS or A2S systems. Right: immunofluorescent staining of Ki67 during FBS and A2S reprogramming (day 9 and day 4, respectively). Scale bar, 50 μ m.

(F) t-SNE plot based on Figure 2A for the anti-proliferation gene *Cdkn1c*. Top schematic indicates the pattern of expression.

(G) Percentage of *Cdh1*+ cells that co-express cell cycle or anti-proliferative genes. Note that because of the limit of detection of single-cell transcriptional analysis co-expression may be underestimated.

(H) Number of NANOG+ colonies on day 4 of A2S reprogramming after small interfering RNA (siRNA)-mediated knock down of *Ehf*. Error bars represent SD of two replicates.

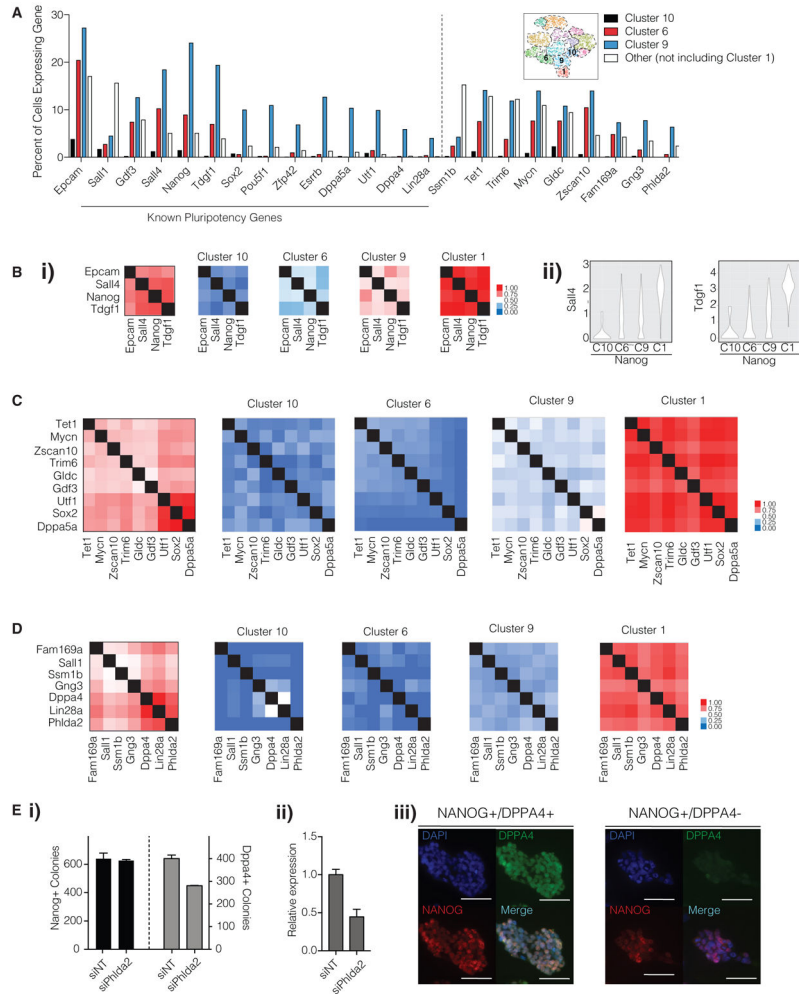


Figure 4. Co-expression Clusters of Core Pluripotency Factors with Specific Subsets
 (A) Percentage of cells expressing each representative pluripotency-associated gene within the t-SNE clusters from Figure 2A, namely, C10, C6, C9, and in all clusters other than C1, C10, C6, and C9.
 (B) (i) Co-expression measured by Jaccard index clustering of genes in group N from Figure 2C for genes within Box 1 from Figure S4B in clusters C10, C6, C9, and C1. Note that because of the limit of detection of single-cell transcriptional analysis, co-expression may be underestimated. (ii) Violin plots depicting the level of expression of Sall4 and Tdgf1 in Nanog+ cells in clusters C10, C6, C9, and C1.
 (C) Same as (B) for genes within Box 2 of Figure S4B.
 (D) Same as (B) for genes within Box 3 of Figure S4B.
 (E) Reprogramming results upon knockdown of Phlda2 during A2S reprogramming. (i) Number of NANOG+ and DPPA4+ colonies on day 6 of A2S reprogramming after siRNA-mediated knock down of Phlda2. Error bars represent SD of two replicates. (ii) Knock down efficiency of the Phlda2 siRNAs compared to a nontargeting control. Bars represent SD between two replicate samples. (iii) Immunofluorescence images for representative NANOG +/DPPA4+ and NANOG+/- DPPA4- colonies. Scale bar, 50 μ m.

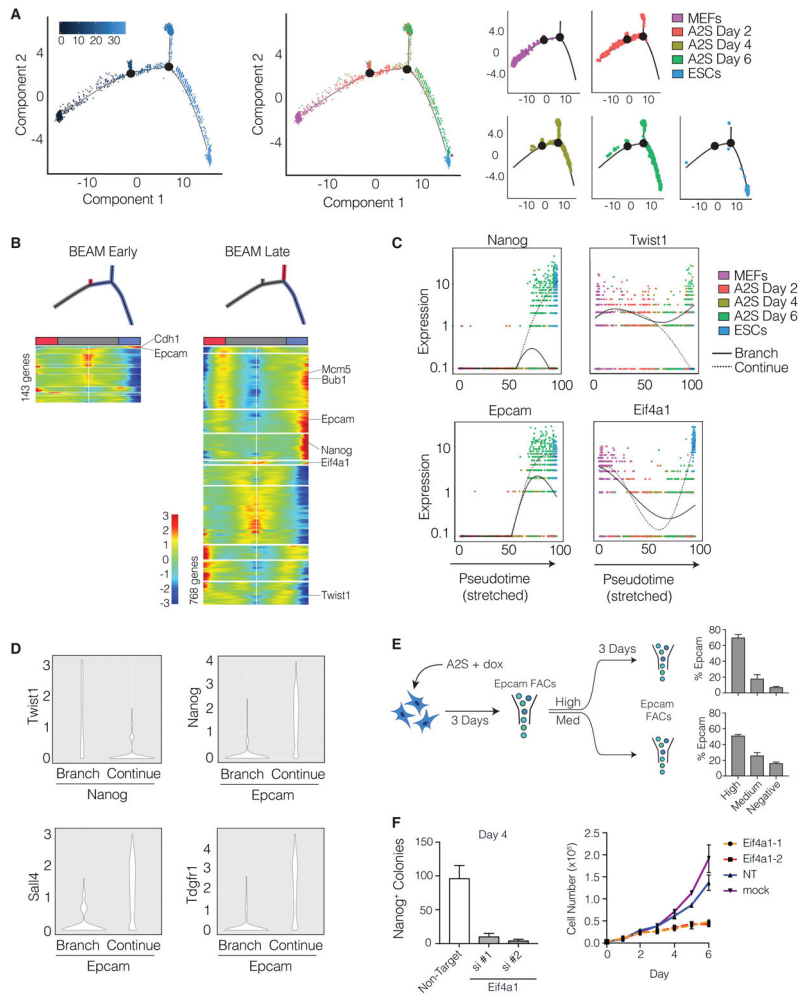


Figure 5. Roadblocks to High-Efficiency Reprogramming

(A) Pseudotime trajectory generated by Monocle for the A2S reprogramming system. Left - trajectory colored by pseudotime. Middle - trajectory colored by sample. Asterisk indicates that MEFs cannot ontogenically convert to ESCs, but pseudotime reflects transition to a pluripotent state. Right - trajectory colored by individual sample.

(B) Heatmaps for clustering of genes that define the branchpoints (q-value, <math><1E-40</math>) from BEAM analysis for early branch (left panel) and late branch (right panel). Center of the gray bar above heatmap is the start of the branchpoint. Red represents cells at the end of the branchpoint. Blue represents cells at the end of the continuing branch.

(C) Pseudotime plots that display how the expression of the representative genes differs as cells either exit at the late branchpoint (solid line) or continue along the path toward successful reprogramming (dashed line) colored by sample.

(D) Violin plots depicting the level of expression of Twist1 in Nanog+ cells (top left) and the expression of Nanog, Sall4, and Tdgf1 in Epcam+ cells in both the late branch and in the continuing segment of the trajectory.

(E) Left: schematic of EPCAM sort experiment. MEFs were reprogrammed in A2S conditions for 3 days and sorted based on EPCAM expression (high or medium). These two

populations underwent 3 more days of reprogramming and were sorted again based on high, medium, or no expression of EPCAM. Right: graphs depicting the percentage of the day 6 population that have high, medium, or no EPCAM expression from cells that were EPCAM-high on day 3 (top) or medium on day 3 (bottom).

(F) Left: number of NANOG+ colonies on day 4 of A2S reprogramming after siRNA-mediated knock down of Eif4a1. Error bars represent SD of two replicates. Right: cell counts on each day of Eif4a1 knock down reprogramming experiment.

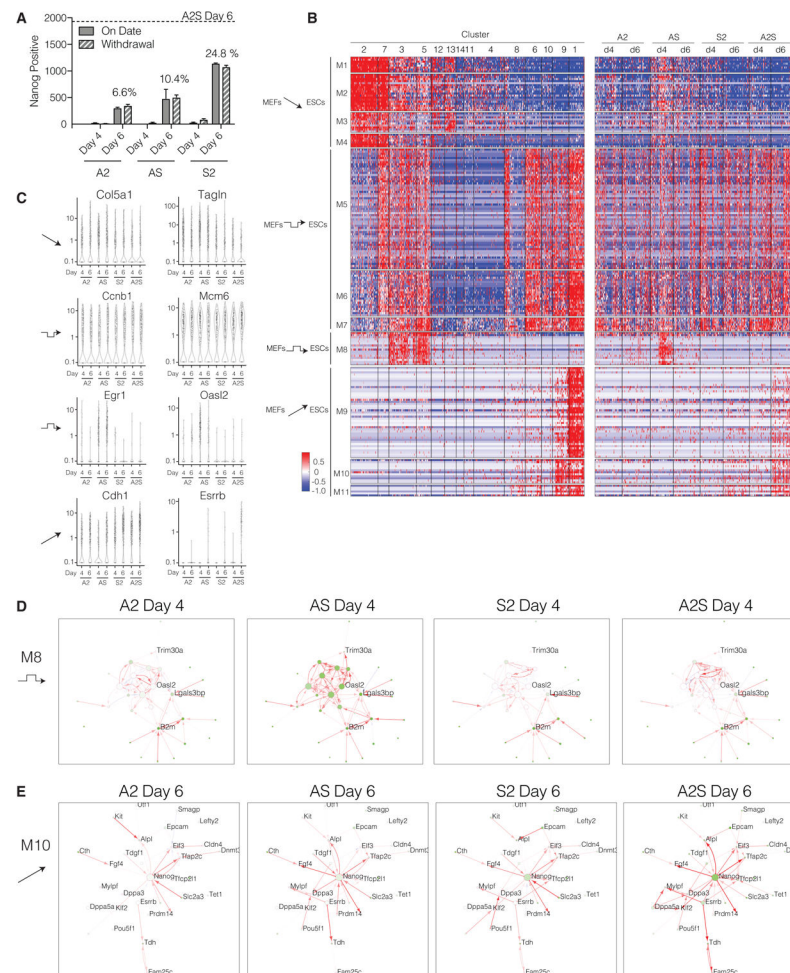


Figure 6. A2S Concurrently Enhances Downregulation of MEF Genes and Upregulation of ESC Genes

(A) NANOG⁺ colonies on specified day or after 4 days of dox withdrawal in each dual combination (A2, AS, and S2). Dashed line: NANOG⁺ colonies on day 6 of A2S. Bars represent standard deviation between two replicate samples.

(B) Heatmap generated from the MERLIN module analysis indicating the level of expression for the differentially expressed genes from the FBS+A2S analysis. Each row is a separate gene. Values are normalized to zero mean from the FBS and A2S reprogramming. Each column is a separate cell grouped based on the clusters in Figure 2A (left) or duration of chemical combination exposure (right). MERLIN modules are labeled as M1 through M11.

(C) Violin plots of representative genes from expression patterns in Figure 6B.

(D) Network wiring of regulatory connections inferred using MERLIN, colored by each reprogramming condition for the genes of a transiently expressed module. The edge color corresponds to the regression coefficient between the regulator and target connected by the edge (ranging from -0.5 (blue) to 0 (white) to 0.5 (red)) estimated using the data from the specific treatment. Edge width corresponds to edge confidence (from 80% [1] to 100% [5]). Node color corresponds to percentage of cells in a condition in which that gene was

expressed (from 0% [white] to 100% [green]). Node border indicates gene membership in a module: pink if the gene is in the given module and gray if it is not. The node size is proportional to the out-degree of the node. Network corresponds to M8.

(E) Same as (D) for genes in an upregulated, pluripotency-associated gene module (M10).

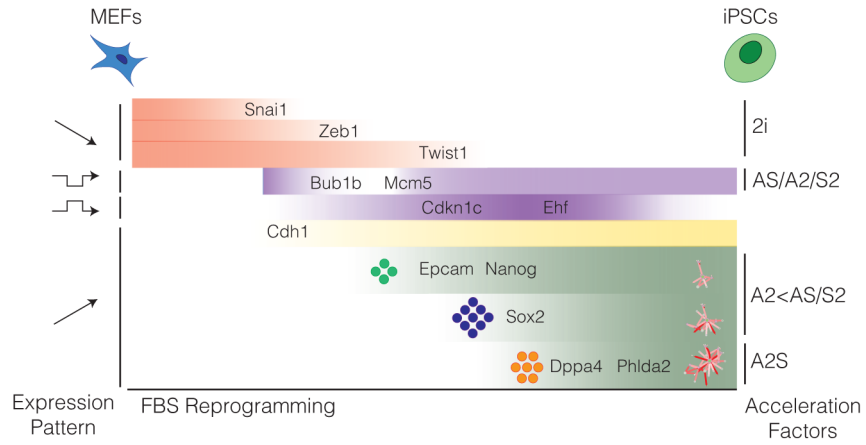


Figure 7. Model Depicting Regulation of Key Genes during MEF Reprogramming
 Four general gene expression patterns are observed during MEF reprogramming: down-regulation, transient downregulation, transient up-regulation, and gene upregulation. Mesenchymal genes are downregulated independently of each other and their expression is compatible with epithelial (Cdh1) or early pluripotency (Nanog) gene expression. Transiently regulated genes include cell cycle and anti-proliferative genes. Completion of reprogramming is enhanced by co-expression of markers, such as EpCAM with pluripotency genes (represented by colored circles), and the complete activation of the pluripotency network (represented by red and white networks). The addition of acceleration factors can impact specific gene expression patterns, whereas only the combination of A2S can lead to complete rewiring of the pluripotency network.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Nanog	CosmoBio	RCAB0002P; RRID: AB_10706358
Anti-Dppa4	ThermoFisher Scientific	PA5-47530; RRID: AB_2606899
Anti-Cdh1	Ebioscience	14-3249-82; RRID: AB_1210458
Anti-Twist1	Novus Biologicals	NBP2-37364SS
Anti-Ki67	Abcam	Ab15580; RRID: AB_443209
Anti-Epcam(CD326)-PE conjugated	BD PharMingen	563477; RRID: AB_2738233
Chemicals, Peptides, and Recombinant Proteins		
Ascorbic Acid	Sigma	A8960
SGC0946	ApexBio	A4167
CHIR99021	Stemgent	04-0004-10
PD0325901	Stemgent	04-0006-10
Critical Commercial Assays		
ddSEQ Single-Cell Isolator	Bio-Rad	12004336
SureCell WTA 3' Library Prep Kit	Illumina	20014279
Deposited Data		
Single-cell RNaseq raw data	NCBI GEO	GEO: GSE108222
Experimental Models: Cell Lines		
Mouse embryonic fibroblasts isolated from 129SV/Jae/C57BL6J mice carrying Col1A:tetO-OSKM/ tetO-OSKM Rosa26:M2rtTA/ M2rtTA or M2rtTA/ wildtype	Sridharan Lab (Sridharan et. al. 2013)	N/A
Mouse embryonic stem cell line V6.5	Laboratory of R. Jaenisch	N/A
Experimental Models: Organisms/Strains		
Mouse: 129SV/Jae/C57BL6J, Col1A: OSKM ^{tetO} / OSKM ^{tetO} R26: M2rtTA/ M2rtTA or M2rtTA/ wildtype	Sridharan Lab (Sridharan et. al. 2013)	N/A
Oligonucleotides		
siRNA targeting: Eif4a1 #1	IDT	mm.Ri.Eif4a1.13.1
siRNA targeting: Eif4a1 #2	IDT	mm.Ri.Eif4a1.13.2
siRNA targeting: Ehf #1	IDT	mm.Ri.Ehf.13.1
siRNA targeting: Ehf #2	IDT	mm.Ri.Ehf.13.2
siRNA targeting: Phlda2 #1	IDT	mm.Ri.Phlda2.13.1
siRNA targeting: Phlda2 #2	IDT	mm.Ri.Phlda2.13.2
siRNA targeting: Non-Targeting	IDT	D-001810-01-50
Primers for RT-PCR, see Table S2	This paper	N/A
Recombinant DNA		
pMX-tdTomato	This paper	N/A
Software and Algorithms		
Monocle2 v2.6.3	Qiu et al., 2017a, 2017b	http://cole-trapnell-lab.github.io/monocle-release/docs/
Cluster3	de Hoon et al., 2004	http://bonsai.hgc.jp/~mdehoon/software/cluster/
Java TreeView	Saldanha, 2004	http://jtreeview.sourceforge.net/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Database for Annotation, Visualization, and Integrated Discovery (DAVID)	Huang et al., 2009	https://david.ncifcrf.gov/
MERLIN	Chasman et al., 2016	N/A
Cytoscape	Shannon et al., 2003	https://cytoscape.org

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript