Defining reprogramming checkpoints from single-cell analysis of induced pluripotency

Khoa A. Tran[1,2], Stefan J. Pietrzak[1,2], Nur Zafirah Zaidan[1,2], Alireza Siahpirani[1], Sunnie Grace McCalla[1], Gopal Iyer[1], Sushmita Roy[1], Rupa Sridharan[1]*

* - Corresponding author and lead contact – rsridharan2@wisc.edu

1- Wisconsin Institute for Discovery

2- These authors contributed equally to this work

University of Wisconsin-Madison

Abstract

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 novel combination of epigenetic 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 towards iPSCs. Ehf, Phlda2, and translation initiation factor Eif4a1 play novel 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 epigenetic modifiers ascorbic acid and a Dot1L inhibitor for pluripotency gene activation.

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; R. 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). These single-cell experiments have been restricted to analyses of pre-selected factors in less than a hundred cells or in low-efficiency systems including non-transgenic chemical reprogramming (Zhao et al., 2018), which may not capture the diversity of gene expression patterns in reprogramming cells.

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 epigenetic 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 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 henceforth) has not been reported.

Using single-cell RNA-seq analysis, we profiled reprogramming MEFs along a time course in both regular serum-containing (FBS) and the A2S system. We found that early events such as epithelial and cell cycle activation are turned on independently. Surprisingly, mesenchymal gene downregulation is not completed early, 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 co-expression 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. We found that without such co-expression, Epcam, which was previously found to enrich for successfully reprogramming cells, could in fact revert to an Epcam- state. Functional experiments prove a novel role for reprogramming-specific transient upregulation of transcription factors (Ehf), translation initiation factor (Eif4a1), and pluripotency stabilization gene (Phlda2). By applying a network-based analytical framework to our single-cell data, we studied the effect of individual components of our cocktail on the pluripotency regulatory network. Our analysis identified specific connections of the pluripotency network that can only be made when both the epigenetic modifiers are present, hence pinpointing the critical regulatory connections that drive reprogramming efficiency. 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 epigenetic 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, c-Myc and Klf4 (OSKM). iPSC generation was monitored by immunofluorescence for NANOG at various timepoints. 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% (Fig 1A, 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, ~1900 Nanog+ iPSC colonies were obtained at an efficiency of ~42% (Methods) (Fig 1B). Beyond this timepoint 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 since the NANOG+ colonies on day 4 were already transgene independent (Fig 1B) as compared to Day 9 of FBS reprogramming (Fig 1A). To avoid biases from plating efficiencies (Schwarz et al., 2014), 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 (Fig 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 (Fig S1A, B).

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 single-cell transcriptomics. We profiled reprogramming cells in FBS on days 3, 6, 9, 12; A2S on days 2, 4, 6 as well as the starting population of MEFs and endpoint of ESCs using a microfluidics-based droplet digital sequencing system (Bio-Rad-ddSeq, Methods). In addition, iPSCs that were generated from the A2S system were profiled to determine their similarity to ESCs. Since AA and 2i are known to change the expression profile of ESCs (Blaschke et al., 2013; Marks et al., 2012), we also sequenced ESCs 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 (Fig S1C and Methods). We used the Monocle2 program (Qiu et al., 2017a; 2017b)(Fig S1D, E) 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-SNE analysis revealed the iPSCs derived from A2S, when passaged in FBS clustered with ESCs grown in FBS and away from ESCs passaged in A2S (Fig 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 (Fig S1F).

A2S accelerates FBS reprogramming

The cells profiled from the time course analysis were grouped into 14 clusters (Fig 2A). The starting MEFs were heterogenous and occupied two clusters (Cluster 2 and Cluster 7) (Fig 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 (Fig 2A). Similarly, the day 2 of A2S samples predominated a single cluster (92% of Cluster 5) while the cells from day 4 and day 6 belonged to several clusters (Fig S2A). Therefore, at the beginning of reprogramming, the cells are more homogeneous than later

timepoints, irrespective of the efficiency of the system. The fact that cells from different timepoints cluster together based on similarity in gene expression profiles suggests that average expression from previous timepoint-based analysis warrants analysis by single-cell sequencing. A small fraction of cells from A2S were found in the FBS clusters and vice versa (Fig S2A). The entire reprogramming population also clustered away from ESCs grown in A2S (Fig S2B); therefore, ESCs grown in FBS were used as the endpoint for all subsequent analyses.

From previous bulk RNA sequencing 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 while 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 (Fig S3B). Similarly, the transient CD73 intermediate marker (Lujan et al., 2015) was rapidly acquired and downregulated (Fig 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)(Fig 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 (Fig S3B). Notably, the gene expression of Mbd3 and Gatad2a were not affected in A2S reprogramming (Fig 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 (Fig 2A), we examined the top 10% of differentially expressed genes between all the clusters. Since 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 (Fig 2B, Fig 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; R. 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 (Fig 2B) decreased expression of developmental signaling and cell migration genes including Tgfb3, Snai1, and Twist2 (Fig 3A). Larger fractions of cells retained expression of Id1, 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 (Fig 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 timepoints

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; R. 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. Surprisingly, Cdh1 upregulation was compatible with expression of mesenchymal genes, albeit in different proportions, as well as the somatic marker Thy1 (Fig 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 sequencing results by performing immunofluorescence for Twist1 and Cdh1 and found an overlap of both markers in the proportion predicted by the transcriptional data (Fig 3C). The trends of dual mesenchymal gene+/Cdh1+ cells were similar in A2S and FBS reprogramming (Fig S2D).

By performing a pairwise comparison between the earliest timepoints of the FBS and A2S time course (Cluster 3 vs Cluster 5, Fig 2B) 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, 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

Since iPSCs self-renew indefinitely, mechanisms that confer an ESC-like cell cycle improve reprogramming efficiency (Hanna et al., 2009; Marion 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- Fig 2B). 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, Fig 2B, 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 selection of cycling MEFs (Mikkelsen et al., 2008) for reprogramming rather than a true upregulation in all cells.

After this timepoint, 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), while a minority retained expression (Cluster 10) (Fig 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 (Fig 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 (Fig 3E).

Cell cycle gene expression upregulation was compatible with Thy1, Zeb2, and Twist1 expression, as well as Cdh1 in both FBS and A2S systems (Fig 3G, Fig 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; H. 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 (Fig 3E). By contrast p53 transcription levels were maintained, in the entire population (Fig S3A). Thus, the senescence block is overcome by the lack of activation of Cdkn1c (Fig 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-L, Fig 2B) was without a specific gene ontology. Since 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, the knockdown of Ehf caused a decrease in A2S reprogramming efficiency (Fig 3H, Fig 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-N, Fig 2B) was largely restricted to reprogramming clusters C9, C6, and C10 that already expressed cell cycle genes (Fig 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 (Fig 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) (Fig 4A). We next determined whether the core pluripotency factors Oct4, Sox2, and Nanog were co-expressed in the reprogramming populations. Interestingly, each of these factors was maximally co-expressed with a different set of genes. Nanog was co-expressed with Sall4, Epcam, and Tdgf1 (Cripto) (Fig 4B). In Nanog-expressing cells, Sall4 was equally expressed in both Cluster 6 and Cluster 9 (Fig 4B). However, Tdgf1 expression was higher in Cluster 9 cells, suggesting that Tdgf1 may be more important for activating the rest of this subset (Fig 4B). On the other hand, while Oct4 was activated with Zfp42 (Fig S4B), Sox2 was co-expressed with Dppa5a and Utf1 and was part of a larger cluster that included Tet1 and Zscan10 (Fig 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 (Fig 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 co-expressed with Lin28a and Phlda2, a gene involved in placental growth (Salas et al., 2004) and not previously implicated in pluripotency (Fig 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 knockdown and control, we found a 25% decrease in the number of DPPA4-positive colonies (Fig 4E). Therefore, the co-expression 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 homogenously 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 (Fig 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 (Fig 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) (Fig 5A). As expected, a larger fraction of day 6 (63%) cells were found in the part of the trajectory toward ESCs 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 towards ES cells from the ones that were found in the branch. At the early branchpoint, the cells that continue toward ESCs have a higher expression of epithelial genes, such as Cdh1 and Epcam (Fig 5B). At the later branchpoint, cells that continue have already activated the cell cycle and present high levels of Nanog as expected (Fig 5B, 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 (Fig 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 (Fig 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 (Fig 5C). Given that Epcam is found co-expressed with a subset of genes (Fig 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 (Fig 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 level of EPCAM expression on day 3 of A2S reprogramming (Fig 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 (Fig 5E). Taken together, these analyses suggest that without co-expression of other genes within the subset, cells may revert to an Epcam-negative state, while with co-expression, cells persist along the trajectory towards 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 (Fig 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 (Fig 5F). This decrease was not due to a change in the number of cells or increasing cell death (Fig 5F). Taken together, these data suggest that sustained expression of genes is affected by co-expression 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 epigenetic modulators – ascorbic acid, which is thought to regenerate 2-oxoglutarate-dependent epigenetic enzymes (Hore et al., 2016), and SGC, 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 (Fig 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 using the FBS+A2S scRNA-seq dataset collected in this study (Methods) using an expression-based network inference algorithm, MERLIN (Chasman et al., 2016). We focused on the ~1800 genes used to initially differentiate the Monocle clusters in the FBS+A2S dataset (Fig 2A) along with sufficiently expressed regulators such as transcription factors, chromatin remodelers and signaling proteins (Fig 6B, 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 a probabilistic prior, 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 (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, spanning 291 genes, the majority of which are statistically enriched for Gene Ontology processes and 4,962 interactions at a confidence of 0.8 or higher connecting 1,009 regulators to 1,628 target genes (Methods). The regulatory network captures known connections among the key pluripotency regulators and target genes (e.g. Esrrb→Klf4, Sox2→Klf4, Esrrb←→Sox2, Esrrb←→Nanog) and is comparable to performance seen when using bulk RNA-seq data (Methods), providing support to the relevance of this network.

MERLIN modules recapitulated the four patterns of expression from MEFs to ESCs (Fig 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. This would enable us to understand why each of the components was required 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, (Fig 6B) which included MEF-specific genes such as Col5a1 and Tagln, even on day 6 (Fig 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 (Fig 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) (Fig 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 (Fig 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, 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 (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 (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) (Fig 6D and Fig S5B, S5C). For the upregulated genes, several connections surrounding Nanog (Fig 6E) and Sox2 (Fig S6A) were absent in the A2 condition, while those around Epcam and Cdh1 were maintained (Fig 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, while a greater proportion of cells expressed Esrrb with Tdh in the AS condition (Fig 6E). In the A2S condition, all these connections are stronger and new connections such as the ones between Dppa5a, Klf2, and Dppa3 emerge (Fig 6E). The network surrounding DNA replication genes such as Mcm6 remains strong in any of the dual combinations (Fig 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. While 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 induced pluripotent stem cells 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 were thought to be temporally activated (Apostolou and Stadtfeld, 2018; Brambrink et al., 2008; Stadtfeld and Hochedlinger, 2010) (Fig 7). Since most studies have focused on MEFs as the starting cell type, an important early event is the mesenchymal to epithelial transition, 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 shutdown 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 Sca1 (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. Using such co-expression analysis, we also isolate Phlda2 as a novel marker of the stabilization phase of reprogramming.

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; Marion 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; Marion 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 novel 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.

Acknowledgements

This work was supported by the NIH R01GM113033 to R.S. and NIH-NIGMS R01GM117339 to S.R. K.A.T. was supported by the University of Wisconsin Fellowship (AOF), S.J.P by the NHGRI training grant to the Genomic Sciences Training Program under Award Number 5T32HG002760, and N.Z.Z. by the American Heart Association (AHA) predoctoral fellowship.

Competing interests

The authors declare that they have no competing interests.

References

Apostolou, E., Hochedlinger, K., 2013. Chromatin dynamics during cellular reprogramming. Nature 502, 462–471.

Apostolou, E., Stadtfeld, M., 2018. Cellular trajectories and molecular mechanisms of iPSC reprogramming. Current Opinion in Genetics & Development 52, 77–85.

Banito, A., Rashid, S.T., Acosta, J.C., Li, S., Pereira, C.F., Geti, I., Pinho, S., Silva, J.C., Azuara, V., Walsh, M., Vallier, L., Gil, J., 2009. Senescence impairs successful reprogramming to pluripotent stem cells. Genes & Development 23, 2134–2139.

Bar-Nur, O., Brumbaugh, J., Verheul, C., Apostolou, E., Pruteanu-Malinici, I., Walsh, R.M., Ramaswamy, S., Hochedlinger, K., 2014. Small Molecules Facilitate Rapid and Synchronous iPSC Generation. Nature Methods 11, 1170–1176.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., Blau, H.M., 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 463, 1042–1047.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., Lorincz, M.C., Ramalho-Santos, M., 2013. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature 500, 222–226.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., Jaenisch, R., 2008. Sequential Expression of Pluripotency Markers during Direct Reprogramming of Mouse Somatic Cells. Cell Stem Cell 2, 151–159.

Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K.J., Ji, F., Schwarz, B.A., Kalocsay, M., Elledge, S.J., Chen, Y., Sadreyev, R.I., Gygi, S.P., Hu, G., Shi, Y., Hochedlinger, K., 2018. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. Cell 172, 106–120.e21.

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

Buganim, Y., Faddah, D.A., Jaenisch, R., 2013. Mechanisms and models of somatic cell reprogramming. Nat Rev Genet 14, 427–439.

Chasman, D., Walters, K.B., Lopes, T.J.S., Eisfeld, A.J., Kawaoka, Y., Roy, S., 2016. Integrating Transcriptomic and Proteomic Data Using Predictive Regulatory Network Models of Host Response to Pathogens. PLoS Comput Biol 12, e1005013.

de Hoon, M.J.L., Imoto, S., Nolan, J., Miyano, S., 2004. Open source clustering software. Bioinformatics 20, 1453–1454.

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., Chen, K., Li, Y., Liu, X., Xu, J., Zhang, S., Li, F., He, W., Labuda, K., Song, Y., Peterbauer, A., Wolbank, S., Redl, H., Zhong, M., Cai, D., Zeng, L., Pei, D., 2010. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell Stem Cell 6, 71–79.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Trcka, D., Wrana, J.L., 2012. A Late Transition in Somatic Cell Reprogramming Requires Regulators Distinct from the Pluripotency Network. Cell Stem Cell 11, 769–782.

Guo, S., Zi, X., Schulz, V.P., Cheng, J., Zhong, M., Koochaki, S.H.J., Megyola, C.M., Pan, X., Heydari, K., Weissman, S.M., Gallagher, P.G., Krause, D.S., Fan, R., Lu, J., 2014. Nonstochastic Reprogramming from a Privileged Somatic Cell State. Cell 1–14.

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

Hore, T.A., Meyenn, von, F., Ravichandran, M., Bachman, M., Ficz, G., Oxley, D., Santos, F., Balasubramanian, S., Jurkowski, T.P., Reik, W., 2016. Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. Proc Natl Acad Sci U S A 113, 12202–12207.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44–57.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., Melton, D.A., 2008. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol 26, 795–797.

Hussein, S.M.I., Puri, M.C., Tonge, P.D., Benevento, M., Corso, A.J., Clancy, J.L., Mosbergen, R., Li, M., Lee, D.-S., Cloonan, N., Wood, D.L.A., Munoz, J., Middleton, R., Korn, O., Patel, H.R., White, C.A., Shin, J.-Y., Gauthier, M.E., Cao, K.-A.L., Kim, J.-I., Mar, J.C., Shakiba, N., Ritchie, W., Rasko, J.E.J., Grimmond, S.M., Zandstra, P.W., Wells, C.A., Preiss, T., Seo, J.-S., Heck, A.J.R., Rogers, I.M., Nagy, A., 2014. Genome-wide characterization of the routes to pluripotency. Nature 516, 198–206.

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., Huangfu, D., Akutsu, H., Liu, D.R., Rubin, L.L., Eggan, K., 2009. A small-molecule inhibitor of TGF-Beta signaling replaces Sox2 in reprogramming by inducing Nanog. Cell Stem Cell 5, 491–503.

Ichida, J.K., Tcw, J., Williams, L.A., Carter, A.C., Shi, Y., Moura, M.T., Ziller, M., Singh, S., Amabile, G., Bock, C., Umezawa, A., Rubin, L.L., Bradner, J.E., Akutsu, H., Meissner, A., Eggan, K., 2014. Notch inhibition allows oncogene-independent generation of iPS cells. 10, 632–639.

Jackson, S.A., Olufs, Z.P.G., Tran, K.A., Zaidan, N.Z., Sridharan, R., 2016. Alternative Routes to Induced Pluripotent Stem Cells Revealed by Reprogramming of the Neural Lineage. Stem Cell Reports 6, 302–311.

Kim, D.H., Marinov, G.K., Pepke, S., Singer, Z.S., He, P., Williams, B., Schroth, G.P., Elowitz, M.B., Wold, B.J., 2015. Single-Cell Transcriptome Analysis Reveals Dynamic Changes in lncRNA Expression during Reprogramming. Cell Stem Cell 16, 88–101.

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

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., Qin, B., Xu, J., Li, W., Yang, J., Gan, Y., Qin, D., Feng, S., Song, H., Yang, D., Zhang, B., Zeng, L., Lai, L., Esteban, M.A., Pei, D., 2010. A mesenchymal-to-

epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell 7, 51–63.

Liang, G., He, J., Zhang, Y., 2012. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. Nature Cell Biology 14, 457–466.

Lujan, E., Zunder, E.R., Ng, Y.H., Goronzy, I.N., Nolan, G.P., Wernig, M., 2015. Early reprogramming regulators identified by prospective isolation and mass cytometry. Nature 521, 352–356.

Maherali, N., Hochedlinger, K., 2009. Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Current Biology 19, 1718–1723.

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

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., Stunnenberg, H.G., 2012. The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell 149, 590–604.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., Meissner, A., 2008. Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49–55.

Modelska, A., Turro, E., Russell, R., Beaton, J., Sbarrato, T., Spriggs, K., Miller, J., Graf, S., Provenzano, E., Blows, F., Pharoah, P., Caldas, C., Le Quesne, J., 2015. The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. Cell Death Dis 6, e1603-e1612.

Mor, N., Rais, Y., Sheban, D., Peles, S., Aguilera-Castrejon, A., Zviran, A., Elinger, D., Viukov, S., Geula, S., Krupalnik, V., Zerbib, M., Chomsky, E., Lasman, L., Shani, T., Bayerl, J., Gafni, O., Hanna, S., Buenrostro, J.D., Hagai, T., Masika, H., Vainorius, G., Bergman, Y., Greenleaf, W.J., Esteban, M.A., Elling, U., Levin, Y., Massarwa, R., Merbl, Y., Novershtern, N., Hanna, J.H., 2018. Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. Cell Stem Cell 23, 412–425.e10.

Nefzger, C.M., Rossello, F.J., Chen, J., Liu, X., Knaupp, A.S., Firas, J., Paynter, J.M., Pflueger, J., Buckberry, S., Lim, S.M., Williams, B., Alaei, S., Faye-Chauhan, K., Petretto, E., Nilsson, S.K., Lister, R., Ramialison, M., Powell, D.R., Rackham, O.J.L., Polo, J.M., 2017. Cell Type of Origin Dictates the Route to Pluripotency. Cell Reports 21, 2649–2660.

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., Lander, E.S., Armstrong, S.A., Daley, G.Q., 2012. Chromatin-modifying enzymes as modulators of reprogramming. Nature 483, 598–602.

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., Kaji, K., 2013. High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature 499, 88–91.

Papp, B., Plath, K., 2013. Epigenetics of Reprogramming to Induced Pluripotency. Cell 152, 1324–1343.

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., Bar-Nur, O., Cheloufi, S., Stadtfeld, M., Figueroa, M.E., Robinton, D., Natesan, S., Melnick, A., Zhu, J., Ramaswamy, S., Hochedlinger, K., 2012. A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell 151, 1617–1632.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., Trapnell, C., 2017a. Single-cell mRNA quantification and differential analysis with Census. Nat Meth 14, 309–315.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., Trapnell, C., 2017b. Reversed graph embedding resolves complex single-cell trajectories. Nat Meth 14, 979–982.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., Maza, I., Mor, N., Baran, D., Weinberger, L., Jaitin, D.A., Lara-Astiaso, D., Blecher-Gonen, R., Shipony, Z., Mukamel, Z., Hagai, T., Gilad, S., Amann-Zalcenstein, D., Tanay, A., Amit, I., Novershtern, N., Hanna, J.H., 2013. Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502, 65–70.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C.O., Forrest, A.R.R., Gough, J., Grimmond, S., Han, J.-H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C.R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R.D., Tegner, J., Lenhard, B., Teichmann, S.A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D.A., Ideker, T., Hayashizaki, Y., 2010. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140, 744–752.

Ruiz, S., Panopoulos, A.D., Herrerias, A., Bissig, K.-D., Lutz, M., Berggren, W.T., Verma, I.M., Izpisua Belmonte, J.C., 2011. A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. Current Biology 21, 45–52.

Salas, M., John, R., Saxena, A., Barton, S., Frank, D., Fitzpatrick, G., Higgins, M.J., Tycko, B., 2004. Placental growth retardation due to loss of imprinting of Phlda2. Mechanisms of Development 121, 1199–1210.

Saldanha, A.J., 2004. Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246–3248.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., 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.

Schwarz, B.A., Bar-Nur, O., Silva, J.C.R., Hochedlinger, K., 2014. Nanog Is Dispensable for the Generation of Induced Pluripotent Stem Cells. Current Biology 24, 347–350.

Schwarz, B.A., Cetinbas, M., Clement, K., Walsh, R.M., Cheloufi, S., Gu, H., Langkabel, J., Kamiya, A., Schorle, H., Meissner, A., Sadreyev, R.I., Hochedlinger, K., 2018. Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming. Cell Stem Cell 23, 289–305.e5.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13, 2498–2504.

Shi, Y., Desponts, C., Do, J.T., Hahm, H.S., Scholer, H.R., 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.

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

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., Carey, M., Garcia, B.A., Plath, K., 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nature Cell Biology 15, 872–882.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., Plath, K., 2009. Role of the Murine Reprogramming Factors in the Induction of Pluripotency. Cell 136, 364–377.

Stadtfeld, M., Hochedlinger, K., 2010. Induced pluripotency: history, mechanisms, and applications. Genes & Development 24, 2239–2263. doi:10.1101/gad.1963910

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

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

Tran, K.A., Jackson, S.A., Olufs, Z.P.G., Zaidan, N.Z., Leng, N., Kendziorski, C., Roy, S., Sridharan, R., 2015. Collaborative rewiring of the pluripotency network by chromatin and signalling modulating pathways. Nature Communications 6, 1–14.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381–386.

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

Vidal, S.E., Amlani, B., Chen, T., Tsirigos, A., Stadtfeld, M., 2014. Combinatorial modulation of signaling pathways reveal cell-type specific requirements for highly efficient and synchronous iPSC reprogramming. Stem Cell Reports 1–11.

Williams-Hill, D.M., Duncan, R.F., Nielsen, P.J., Tahara, S.M., 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.

Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M., Roderick, J.E., Van der Meulen, J., Schatz, J.H., Rodrigo, C.M., Zhao, C., Rondou, P., de Stanchina, E., Teruya-Feldstein, J., Kelliher, M.A., Speleman, F., Porco, J.A.J., Pelletier, J., Ratsch, G., Wendel, H.-G., 2014. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature 513, 65–70.

Zhao, T., Fu, Y., Zhu, J., Liu, Y., Zhang, Q., Yi, Z., Chen, S., Jiao, Z., Xu, X., Xu, J., Duo, S., Bai, Y., Tang, C., Li, C., Deng, H., 2018. Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. Cell Stem Cell 23, 31–45.

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., Zhou, Q., 2009. iPS cells produce viable mice through tetraploid complementation 461, 86–90.

Zhou, Z., Yang, X., He, J., Liu, J., Wu, F., Yu, S., Liu, Y., Lin, R., Liu, H., Cui, Y., Zhou, C., Wang, X., Wu, J., Cao, S., Guo, L., Lin, L., Wang, T., Peng, X., Qiang, B., Hutchins, A.P., Pei, D., Chen, J., 2017. Kdm2b Regulates Somatic Reprogramming through Variant PRC1 Complex-Dependent Function. Cell Reports 21, 2160–2170.

Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M., Nolan, G.P., 2015. A Continuous Molecular Roadmap to iPSC Reprogramming through Progression Analysis of Single-Cell Mass Cytometry. Cell Stem Cell 16, 323–337.

Figure Legends

Figure 1: Combining epigenetic 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 arrow.
Bottom – Number of NANOG+ colonies counted at each indicated time point or after 4 additional days after doxycycline was removed (Withdrawal). Bars represent standard deviation between two replicate samples. Right panel – immunofluorescence images of NANOG. Scale bar = 250 um.

B) Top – Schematic of A2S reprogramming experiment. Cells were harvested and immunofluorescence performed on the days indicated by arrow.
Bottom – Number of NANOG+ colonies counted at each indicated time point or after 4 additional days after doxycycline was removed (Withdrawal). Bars represent standard deviation 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.

D) Monocle clustering plot showing ESCs cultured in A2S or FBS conditions and iPSCs cultured in 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) Heatmap representing the percentage of cells expressing the top 10% differentially expressed genes that define the 14 t-SNE clusters in Fig 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 according to their progression along reprogramming. Significant gene ontology terms associated with a specific group are labeled on the right.

Figure 3: Reprogramming-specific gene expression patterns are important for conversion to iPSCs

A) t-SNE plots based on Fig 2A highlighting the expression of MEF-associated genes that are downregulated as cells transition from MEFs to iPSCs. 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.
C) Left – t-SNE plots based on Fig 2A illustrating co-expression of Cdh1 with Twist1. Right – Immunofluorescent staining for Cdh1 and Twist1. Percentage of Cdh1+/Twist1+ colonies shown below image. Scale bar = 10 um.
D) t-SNE plots based on Fig 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 um.
F) t-SNE plot based on Fig 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.
H) Number of NANOG+ colonies on day 4 of A2S reprogramming after siRNA-mediated knockdown of Ehf. Bars represent standard deviation between two replicate samples.

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 Fig 2A, C10, C6, C9, and in all clusters other than C1, C10, C6 and C9.

B) Left – Co-expression measured by Jaccard index clustering of genes in Group N from Fig 2B for genes within Box 1 from Fig S4B in clusters C10, C6, C9, and C1. Right – 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 Fig S4B.

D) Same as B for genes within Box 3 of Fig S4B.

E) Reprogramming results upon knockdown of Phlda2 during reprogramming.
Left – Number of NANOG+ and DPPA4+ colonies seen when Phlda2 is knocked down compared to a non-targeting siRNA control. Bars represent standard deviation between two replicate samples.
Middle – Knockdown efficiency of the Phlda2 siRNAs compared to a non-targeting control. Bars represent standard deviation between two replicate samples.
Right – Immunofluorescence images for representative DPPA4+ and DPPA4- colonies. Scale bar = 50 um.

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.
Right- trajectory colored by individual sample.

B) Heatmaps for clustering of genes that define the branchpoints (q-value < 1E-40) 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 towards 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 when Eif4a1 is knocked down compared to a non-targeting siRNA control. Bars represent standard deviation between two replicate samples.
Right – Number of cells present in well at each day during Eif4a1 knockdown 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 Fig 2A (left) or duration of chemical combination exposure (right). MERLIN modules are labeled as M1 through M11.

C) Jitter plots of representative genes from expression patterns in Fig 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 -.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; downregulation, transient downregulation, transient upregulation, and gene upregulation. Mesenchymal genes are downregulated independently of each other and can be compatible with epithelial (Cdh1) or early pluripotency (Nanog) gene expression. Transiently regulated genes consist of cell cycle and anti-proliferative genes. Upregulation of pluripotency genes involves co-expression of core 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 while only the combination of A2S can lead to complete rewiring of the pluripotency network.

Figure 1

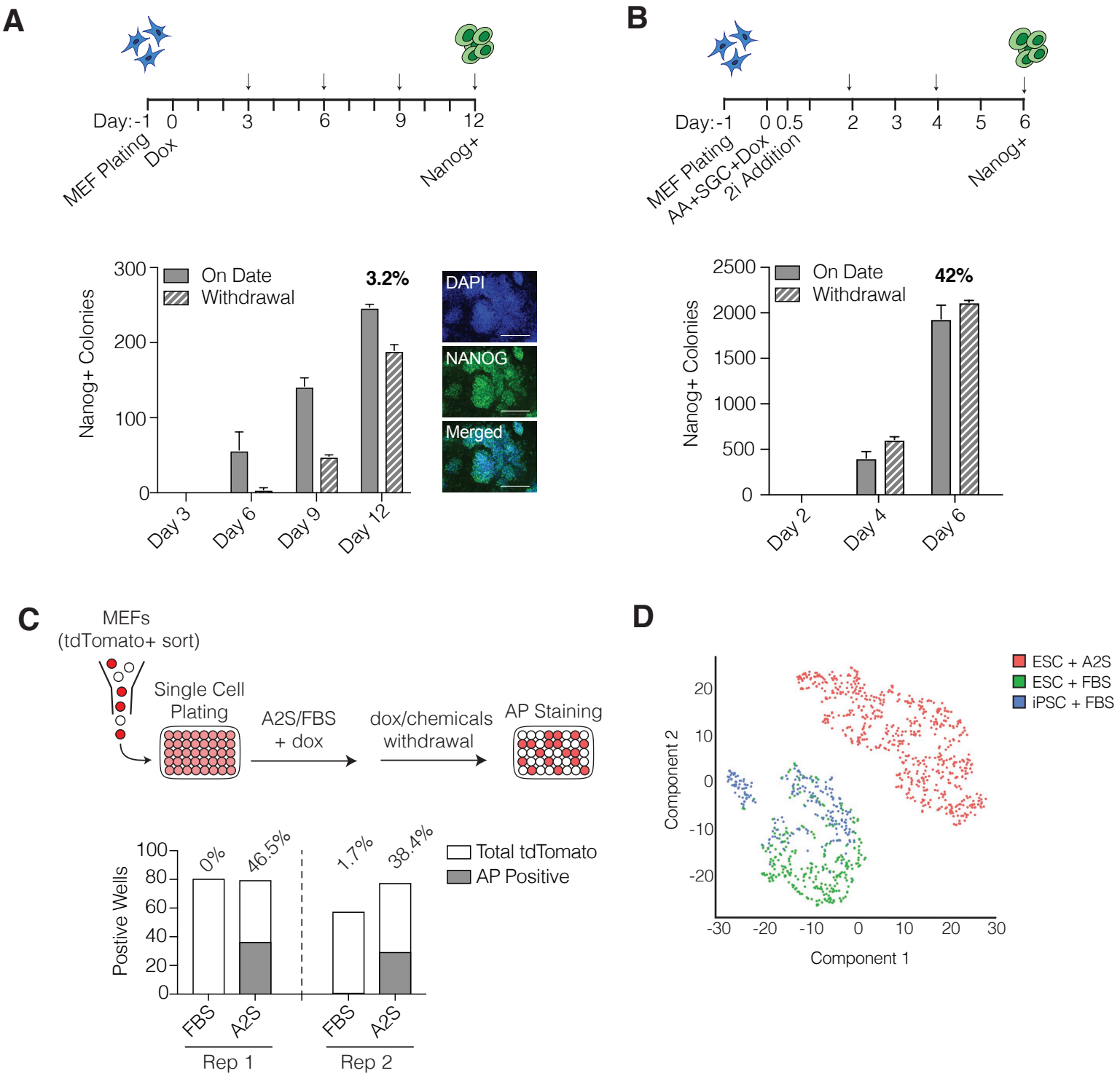


Figure 2

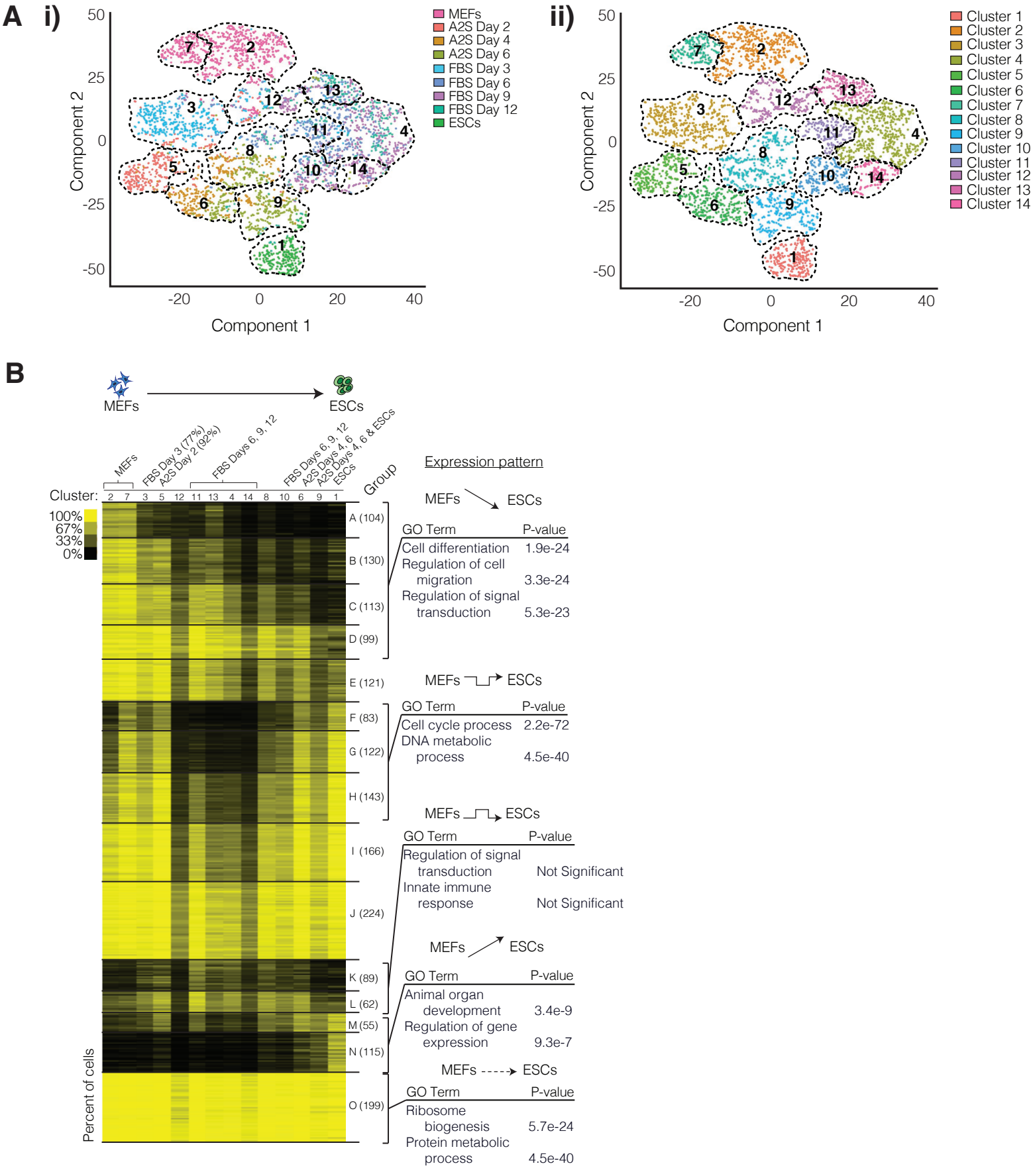


Figure 3

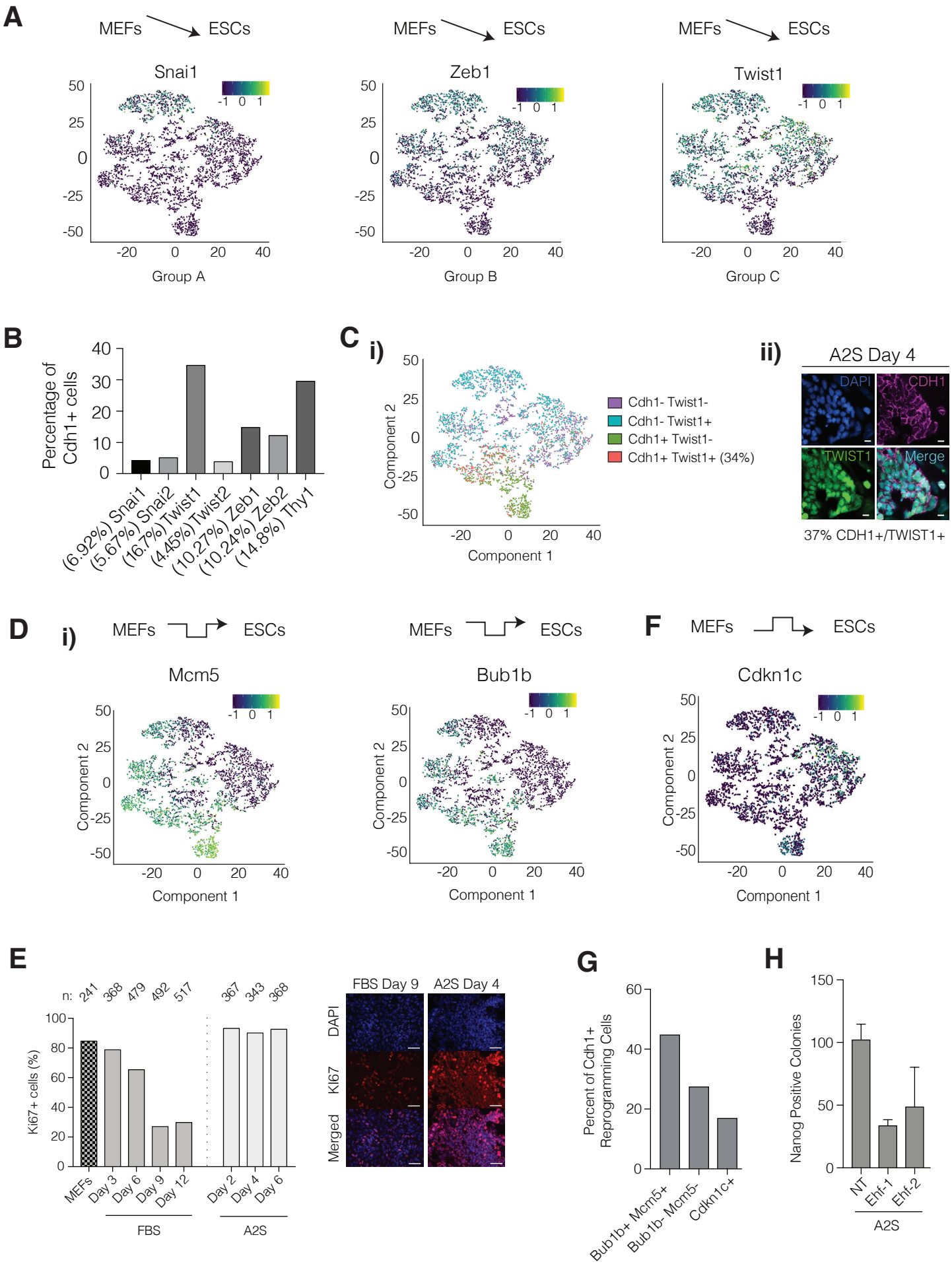
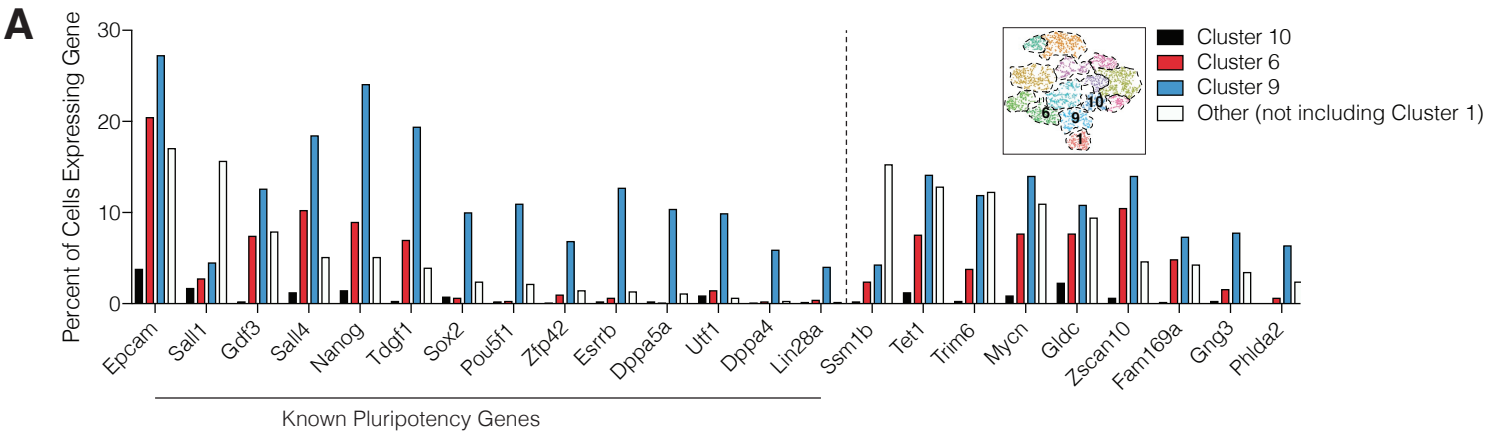
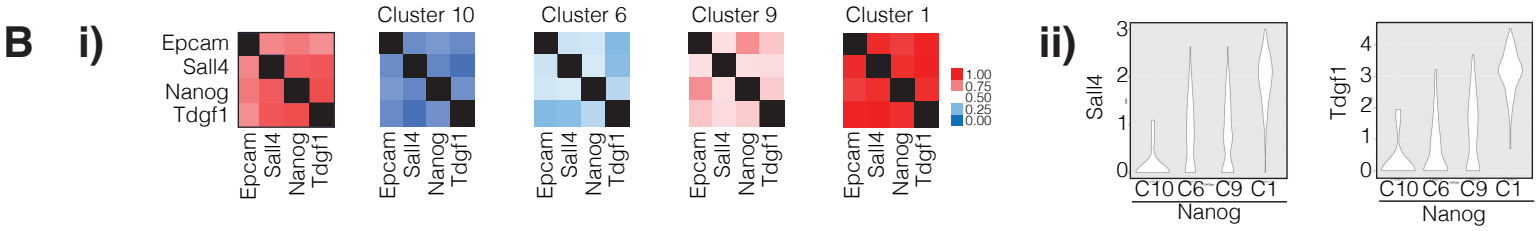
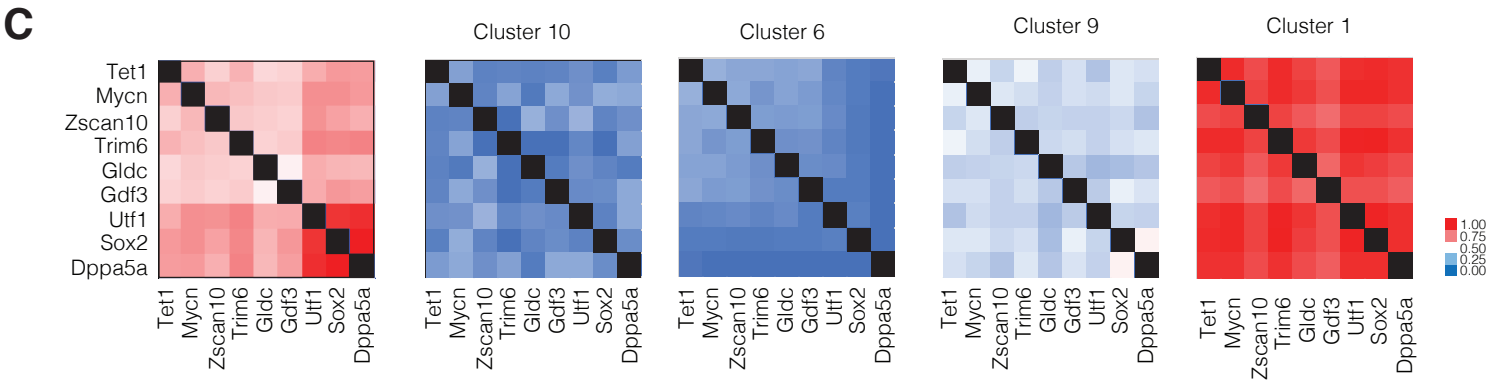
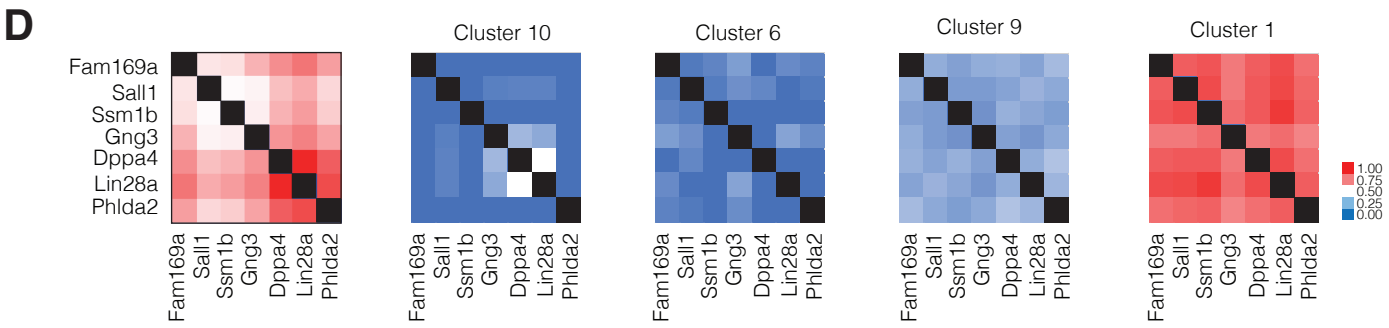


Figure 4









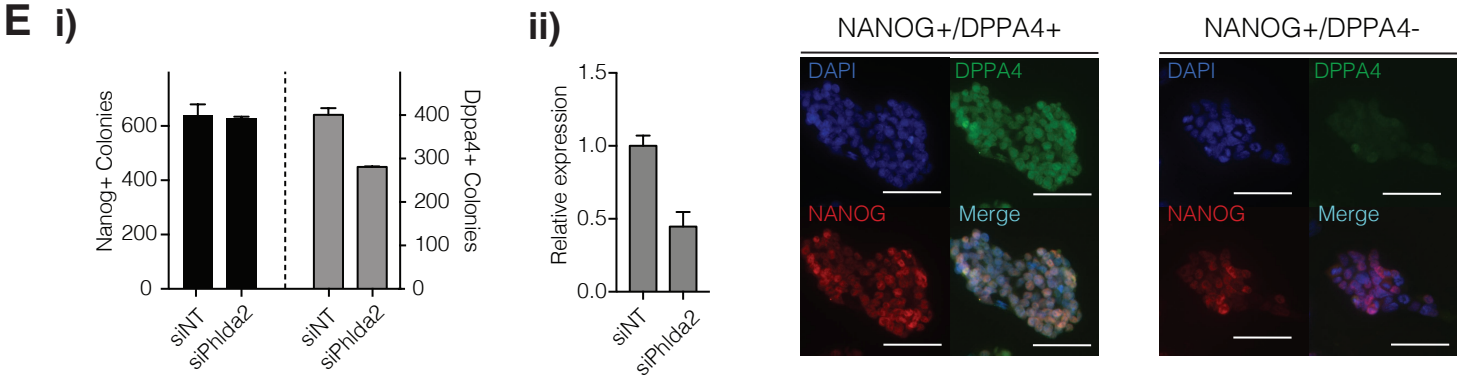
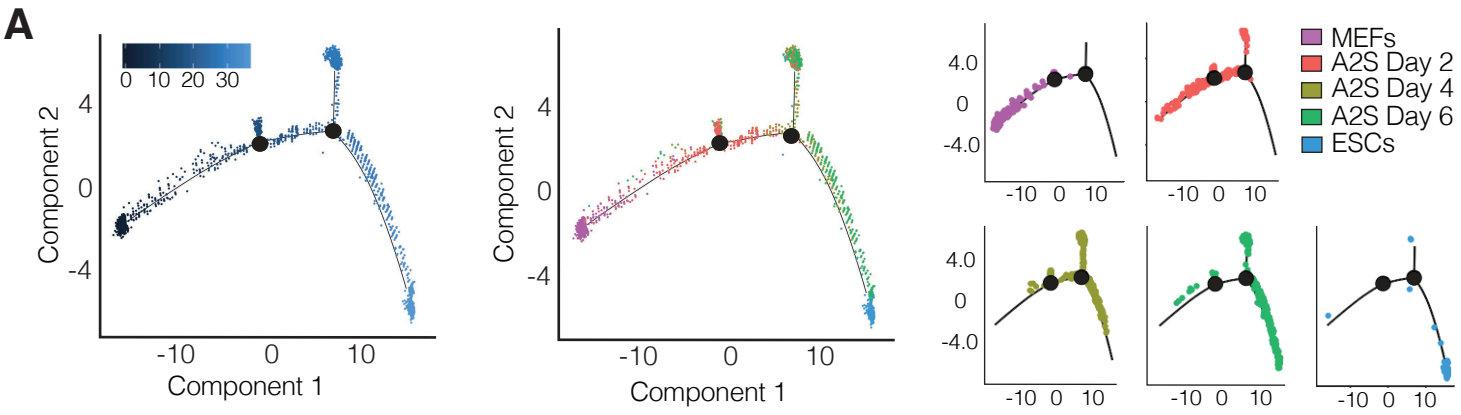
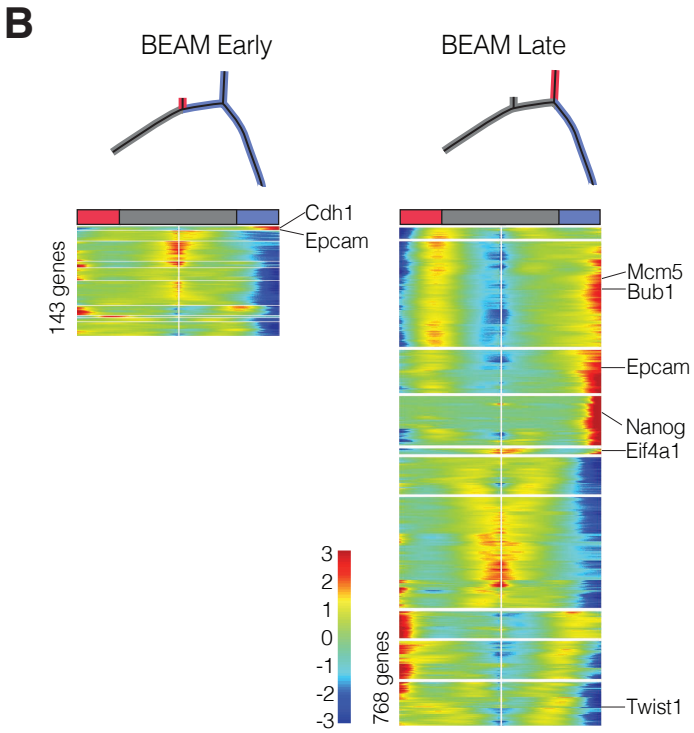
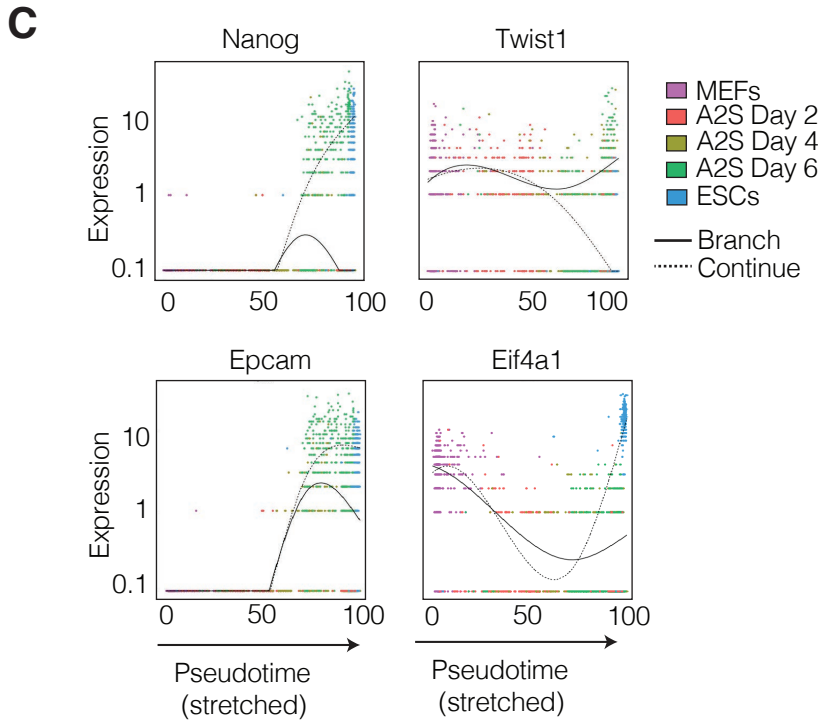
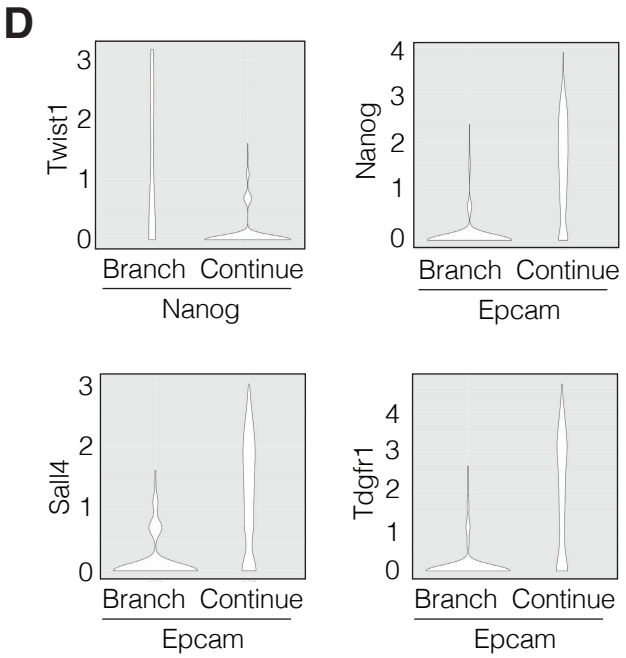


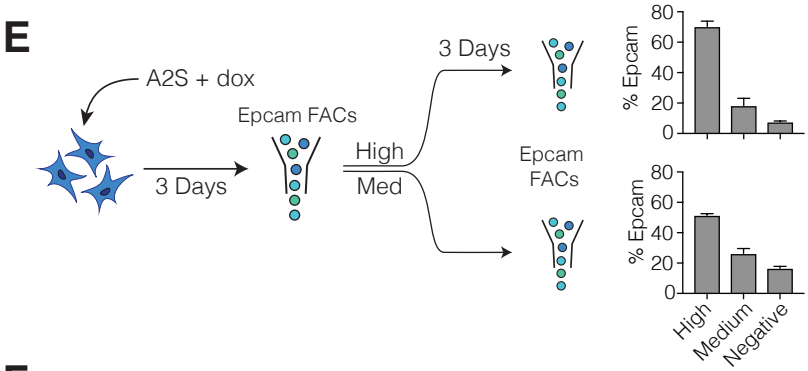
Figure 5











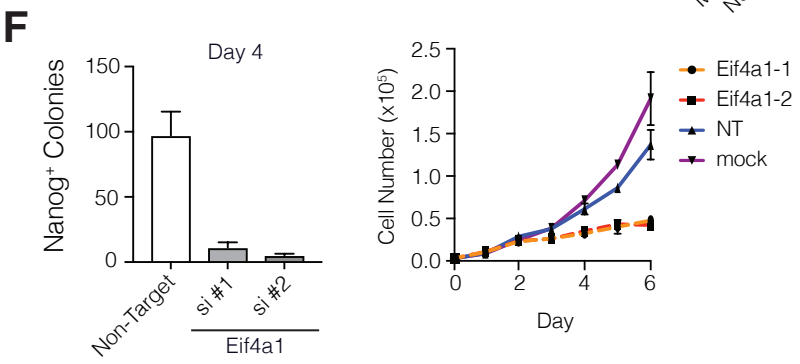


Figure 6

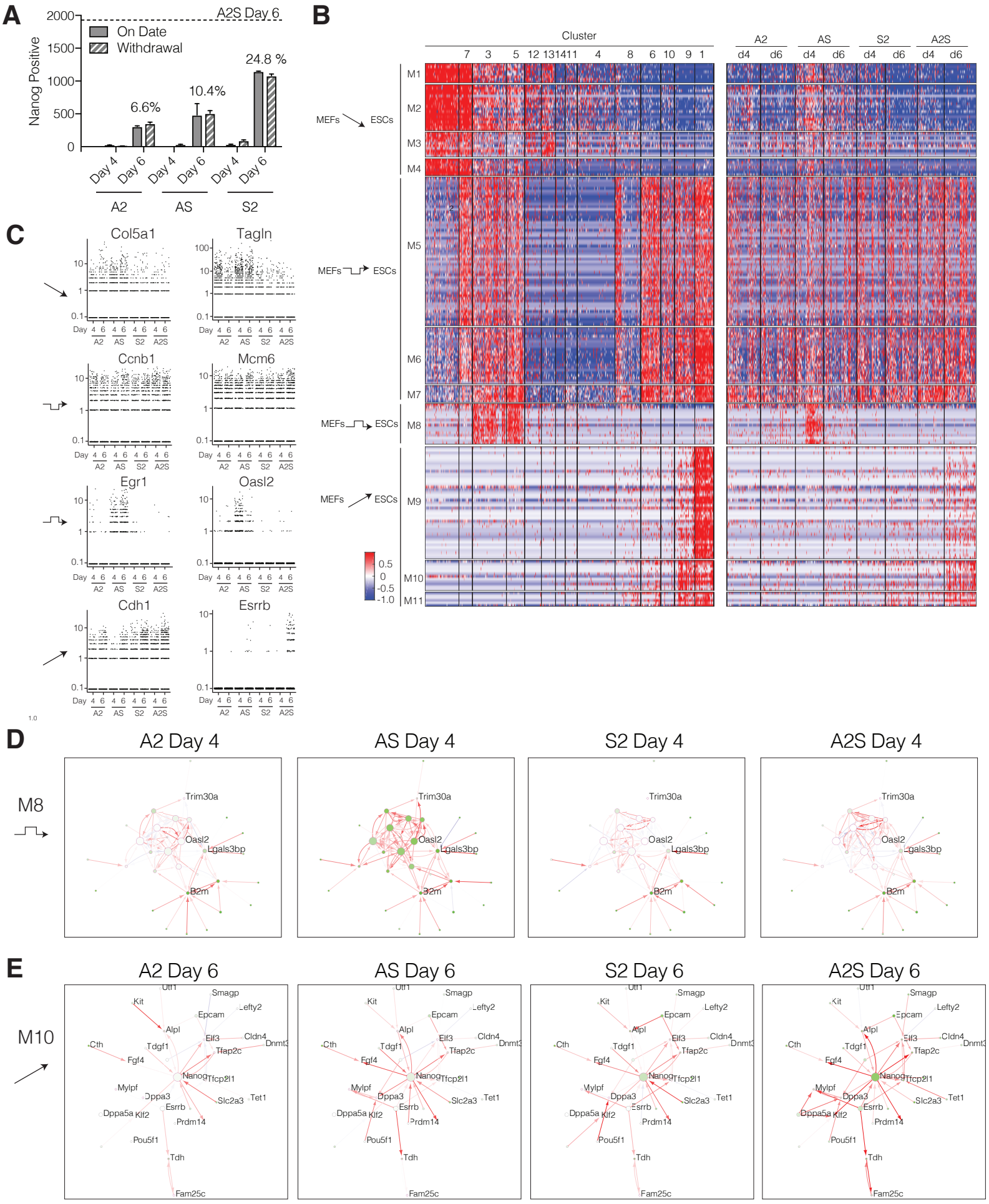


Figure 7