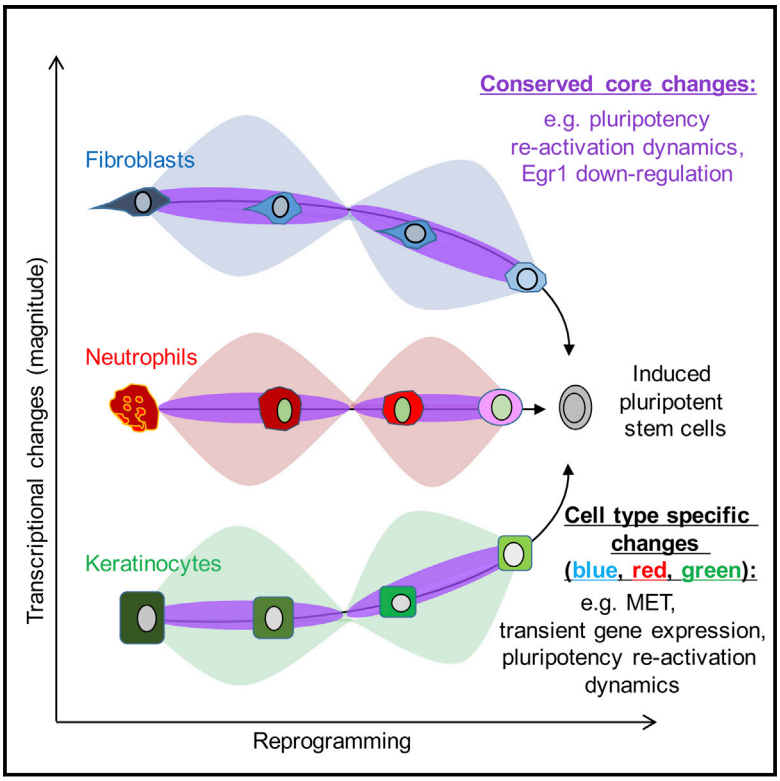
Report

Cell Reports

Cell Type of Origin Dictates the Route to Pluripotency

Graphical Abstract



Highlights

- Cell-type-specific molecular reprogramming trajectories converge at the iPSC stage
- Two transcriptional waves are common to all reprogramming pathways
- Transient changes during reprogramming are mostly cell-type specific
- *Egr1* downregulation is part of a conserved transcriptional program

Authors

Christian M. Nefzger, Fernando J. Rossello, Joseph Chen, ..., David R. Powell, Owen J.L. Rackham, Jose M. Polo

Correspondence

owen.rackham@duke-nus.edu.sg (O.J.L.R.), jose.polo@monash.edu (J.M.Polo)

In Brief

Nefzger et al. find that the molecular reprogramming trajectories of fibroblasts, neutrophils, and keratinocytes have a cell-type-specific component that only fully converges in induced pluripotent stem cells. The authors also identify universal changes shared by all three cell types, including two transcriptional waves and a conserved transcriptional program involving *Egr1* downregulation.

Nefzger et al., 2017, Cell Reports *21*, 2649–2660
December 5, 2017
https://doi.org/10.1016/j.celrep.2017.11.029

CellPress

Cell Reports
Report

Cell Type of Origin Dictates the Route to Pluripotency

Christian M. Nefzger,[1,2,3,9] Fernando J. Rossello,[1,2,3,9] Joseph Chen,[1,2,3] Xiaodong Liu,[1,2,3] Anja S. Knaupp,[1,2,3] Jaber Firas,[1,2,3] Jacob M. Paynter,[1,2,3] Jahnvi Pflueger,[4,5] Sam Buckberry,[4,5] Sue Mei Lim,[1,2,3] Brenda Williams,[3,6] Sara Alaei,[1,2,3] Keshav Faye-Chauhan,[1,2,3] Enrico Petretto,[8] Susan K. Nilsson,[3,6] Ryan Lister,[4,5] Mirana Ramialison,[2] David R. Powell,[7] Owen J.L. Rackham,[8,*] and Jose M. Polo[1,2,3,10,*]

[1]Department of Anatomy and Developmental Biology, Monash University, Wellington Road, Clayton, VIC 3800, Australia
[2]Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Wellington Road, Clayton, VIC 3800, Australia
[3]Australian Regenerative Medicine Institute, Monash University, Wellington Road, Clayton, VIC 3800, Australia
[4]ARC Center of Excellence in Plant Energy Biology, University of Western Australia, Perth, WA 6009, Australia
[5]Harry Perkins Institute of Medical Research, Perth, WA 6009, Australia
[6]Biomedical Manufacturing, CSIRO Manufacturing, Clayton, VIC 3169, Australia
[7]Monash Bioinformatics Platform, Monash University, Wellington Road, Clayton, VIC 3800, Australia
[8]Program in Cardiovascular and Metabolic Disorders, Duke-National University of Singapore Medical School, 8 College Road, 169857 Singapore, Singapore
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: owen.rackham@duke-nus.edu.sg (O.J.L.R.), jose.polo@monash.edu (J.M.Polo)
https://doi.org/10.1016/j.celrep.2017.11.029

SUMMARY

Our current understanding of induced pluripotent stem cell (iPSC) generation has almost entirely been shaped by studies performed on reprogramming fibroblasts. However, whether the resulting model universally applies to the reprogramming process of other cell types is still largely unknown. By characterizing and profiling the reprogramming pathways of fibroblasts, neutrophils, and keratinocytes, we unveil that key events of the process, including loss of original cell identity, mesenchymal to epithelial transition, the extent of developmental reversion, and reactivation of the pluripotency network, are to a large degree cell-type specific. Thus, we reveal limitations for the use of fibroblasts as a universal model for the study of the reprogramming process and provide crucial insights about iPSC generation from alternative cell sources.

INTRODUCTION

In 2006, Takahashi and Yamanaka described that mature cells can be reprogrammed to a pluripotent state, called induced pluripotent stem cells (iPSCs), by the forced expression of four transcription factors, namely, *Oct4*, *Klf4*, *Sox2*, and *cMyc* (OKSM), with vast implications for the regenerative medicine field (Takahashi and Yamanaka, 2006). Recent landmark studies have provided detailed molecular roadmaps for the reprogramming process of fibroblasts (Hussein et al., 2014; O'Malley et al., 2013; Polo et al., 2012) to iPSCs. However, the molecular events that underpin the reprogramming process of other cell types are still largely unknown, and to what extent aspects of reprogramming are universal or cell-type specific has not been properly addressed yet. For example, although Yamanaka and colleagues showed that a variety of human cell types pass through a primitive streak-like state late during reprogramming (Takahashi et al., 2014), whether this applies across species boundaries and to other cell types is still unclear. To fill this knowledge gap, we functionally and molecularly characterized the reprogramming pathways of mouse neutrophils and keratinocytes in addition to fibroblasts. We reveal universal transcriptional trends, but we also show significant cell-type-specific changes that appear to be dependent on the cell type of origins' transcriptional networks, including crucial processes like mesenchymal-to-epithelial transition (MET), reactivation of the pluripotency network, and the extent of developmental reversion. Accordingly, we show that the nuclear reprogramming process of somatic cells into the induced pluripotent state has a universal and cell-type-specific component, underscoring the complexity of such processes.

RESULTS

Distinct Cell-Type-Specific Changes in the Cell Surface Marker Profile during Reprogramming

To dissect whether the reprogramming pathways of different cell types follow common or different transcriptional dynamics, we first defined the cellular reprogramming pathways of fibroblasts, neutrophils, and keratinocytes. We used an OKSM reprogrammable mouse model (doxycycline [dox] controlled OKSM expression; harboring an *Oct4-Gfp* reporter) under the same reprogramming conditions for each cell type. We focused on hematopoietic cell types because they have a well-defined hierarchy (Bryder et al., 2006; Weissman and Shizuru, 2008) and on keratinocytes because the MET is a critical aspect of the fibroblast reprogramming process (Li et al., 2010; Samavarchi-Tehrani et al., 2010), and keratinocytes cells are already in an epithelial state. As shown previously by ourselves and others, fibroblasts undergoing reprogramming first activate

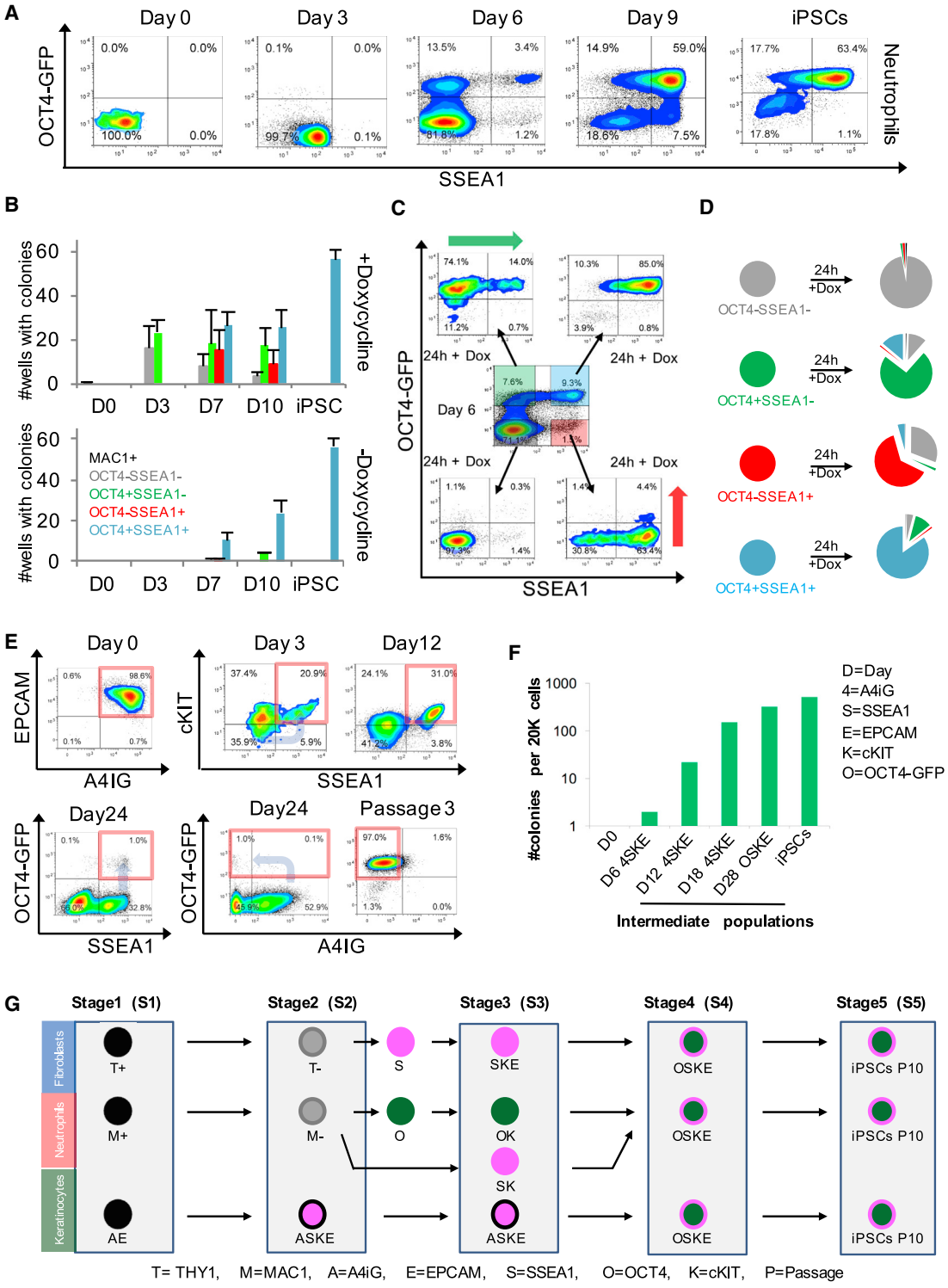


Figure 1. Dissection of Neutrophil and Keratinocyte Reprogramming Pathways

(A) Flow cytometry plots for OCT4-GFP and SSEA1 expression for reprogramming neutrophils at different time points.

(B) On the indicated days, reprogramming neutrophil cultures were subfractioned according to the detection of MAC1, OCT4, and SSEA1 (note that cells positive for either OCT4 or SSEA1 were always MAC1 negative), and the ability of subfractions to give rise to alkaline positive (AP+) colonies in the presence or absence of dox was assessed (n = 3 independent experiments; error bars represent SD).

(legend continued on next page)

the pluripotency-associated cell surface marker SSEA1 and then activate the endogenous *Oct4* locus (Brambrink et al., 2008; Polo et al., 2012; Stadtfeld et al., 2008; Wernig et al., 2007). Interestingly, fluorescence-activated cell sorting (FACS)-purified hematopoietic cell types (neutrophils, macrophages, B cells, and their progenitors) show the reverse pattern under reprogramming conditions, with the vast majority of cells undergoing reprogramming becoming OCT4-GFP$^+$ before expressing SSEA1 (Figure S1A). A minor subpopulation of hematopoietic cells (ranging from 0.5%–4.0%; Figure 1A; Figure S1A) appeared to follow the canonical reactivation sequence of fibroblasts (Brambrink et al., 2008; Polo et al., 2012; Stadtfeld et al., 2008; Wernig et al., 2007). Because these seemingly alternative reprogramming pathways were most pronounced in neutrophils, we selected them for in-depth characterization. After 3 days in reprogramming conditions, a subset of neutrophils lost their identity marker MAC1, coinciding with an increase in cell size (Figure S1B). Using cell tracking studies, we show that Mac1$^-$ cells gave rise to OCT4-GFP$^+$/SSEA1$^-$ and OCT4-GFP$^-$/SSEA1$^+$ cells, both of which were able to generate iPSCs (Figures 1B–1D). However, only when cells had acquired both markers and completed reprogramming did they become independent of OKSM transgene expression (Figure 1B). The cell surface markers cKIT and EPCAM were able to further subdivide the identified neutrophil reprogramming intermediates (Figure S1C). These results show that neutrophils can reach pluripotency by using partially different cellular pathways and that the timing at which these markers appear can be different from fibroblasts.

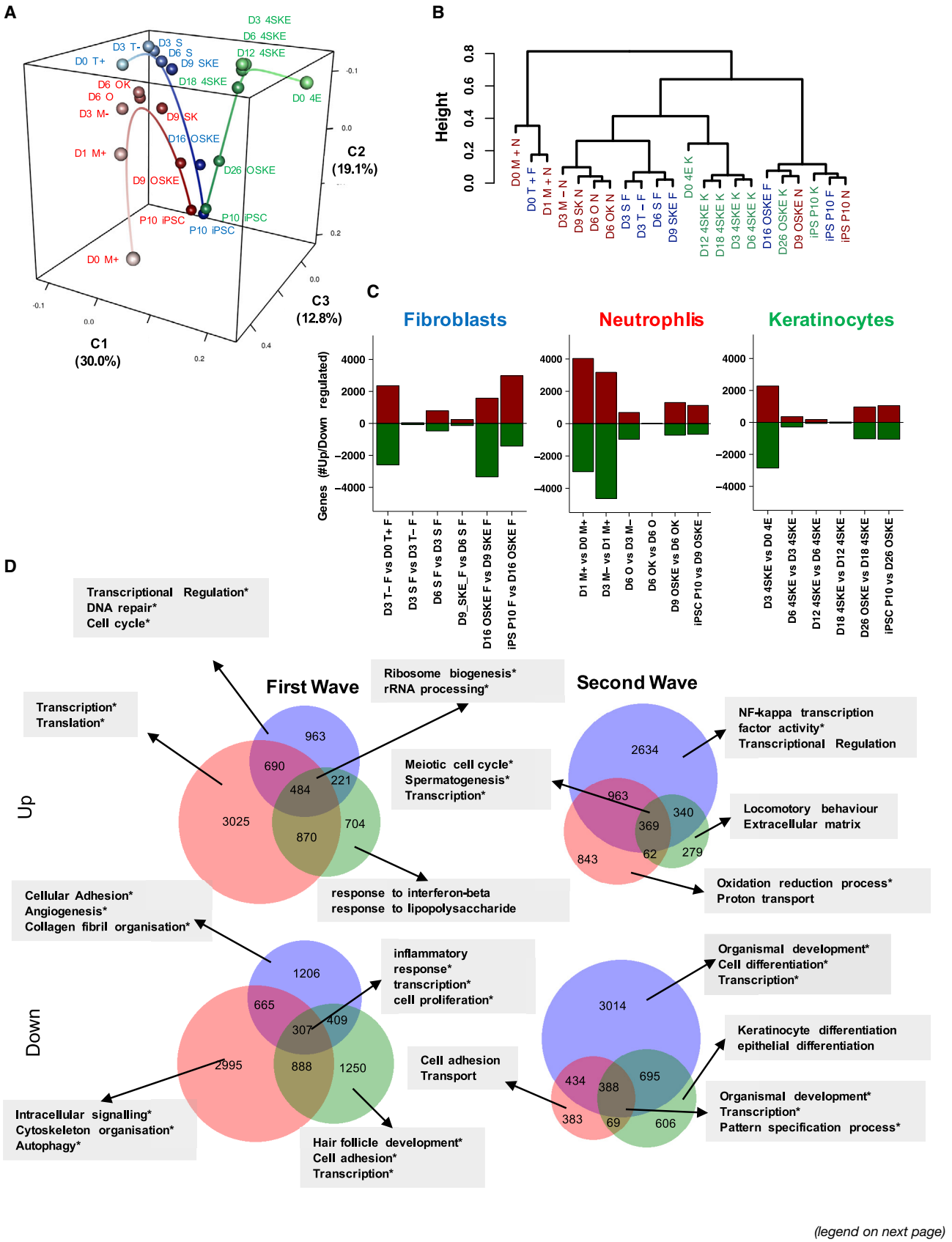
Reprogramming of keratinocytes (Figure 1E; Figures S2A and S2B) also revealed a cell-type-specific pattern of changes in cell surface markers. A subset of cells became both SSEA1$^+$ and cKIT$^+$ by day 3 of reprogramming without losing expression of the keratinocyte identity markers alpha 4 integrin (A4IG) or EPCAM, contrary to what was observed in fibroblast and neutrophil reprogramming (Figure 1E; Figure S2B). This population gradually gained reprogramming potential and eventually progressed toward an OCT4-GFP$^+$ state (day 24–28) while becoming A4IG negative (Figures 1E and 1F; Figure S2B). Remarkably, despite the fact that keratinocytes have no MET barrier to overcome, they required the longest period of time to activate both key iPSC markers SSEA1 and OCT4-GFP compared with both fibroblasts (~12–16 days) and neutrophils (~6 days) (Figures 1A and 1E; Figures S1C, S2B, and S3A).

Spanning-tree progression analysis of density-normalized events (SPADE) clusters phenotypically similar cells into nodes (Qiu et al., 2011) and has been used to interrogate, infer, and visualize cellular hierarchies and transitions based on the expression of cell surface markers in a range of systems, including nuclear reprogramming (Zunder et al., 2015), hematopoiesis (Bendall et al., 2011), and the intestinal epithelium (Nefzger et al., 2016). A SPADE analysis on flow cytometry data for these cell types during reprogramming suggests that the cell surface marker profiles of reprogramming intermediates gradually become similar, but only converge upon becoming fully reprogrammed cells (Figure 1G; Figure S3B). Altogether, our results reveal that different cell types activate and lose different surface markers with different dynamics, and unlike previous fibroblast studies (Polo et al., 2012; Stadtfeld et al., 2008; Wernig et al., 2007), reactivation of endogenous OCT4 is not a reliable marker to indicate transgene independence (Figure 1B). Moreover, we provide more evidence for the existence of alternative pathways during reprogramming of a given cell type (Chantzoura et al., 2015; Jackson et al., 2016). Despite the observed differences in the reprogramming process of iPSCs derived from all three cell types, we observed that they were fully pluripotent and able to give rise to derivatives of all three germ layers (Figure S3C).

The Reprogramming Process of Fibroblasts, Neutrophils, and Keratinocytes Is Dominated by Two Transcriptional Waves

To characterize these changes beyond the expression of cell surface markers, we performed whole transcriptome sequencing (RNA sequencing [RNA-seq]) on the three cell types, their reprogramming intermediates, and iPSCs (respective time points/cell surface maker profiles are indicated in Figure 2A). While the isolated intermediate populations for neutrophils and keratinocytes are enriched for successfully reprogramming cells (Figures 1A–1F), they likely retain a small degree of heterogeneity as we have shown previously for reprogramming fibroblast intermediates, purified via a similar strategy as used in this study (Polo et al., 2012). Correspondence analysis (COA) on the resulting datasets demonstrated a clear molecular connectivity within each cell type, which shows that gene expression profiles of intermediate cells become progressively similar to those of iPSCs (Figure 2A), confirming that reprogramming occurs molecularly via an ordered sequence of events for all the cell types. Furthermore, COA, as well as an unsupervised hierarchical clustering analysis (Figures 2A and 2B), revealed that the reprogramming pathways of all three cell types only fully converged at the end of the reprogramming process. This demonstrates that the transcriptional pathways that early intermediates follow are different for each cell type. Interestingly, intermediates for

(C and D) Cultures of reprogramming day 6 neutrophils were FACS subfractioned into OCT4$^-$/SSEA1$^-$ (gray), OCT4$^-$/SSEA1$^-$ (green), OCT4$^-$/SSEA1$^+$ (red), and OCT4$^+$/SSEA1$^+$ (blue) populations, cultured in isolation for 24 hr (with dox), and reassessed for Oct4-GFP and Ssea1 expression levels followed by (D) quantification. In (C), red and green arrows indicate that a subset of OCT4$^+$/SSEA1$^-$ and OCT4$^-$/SSEA1$^+$ positive cells transitioned to an OCT4$^+$/SSEA1$^+$ state.
(E) Flow cytometry plots for reprogramming keratinocytes at indicated time points. The reprogramming intermediates are indicated by the red frames.
(F) Number of alkaline phosphatase-positive colonies derived from 2 × 10^4 purified reprogramming intermediates (as defined in E by red frames) from the indicated time points from a representative experiment.
(G) Overview of changes in cell surface marker profile during reprogramming in fibroblasts, neutrophils, and keratinocytes. For comparative purposes, the reprogramming process of all cell types was subdivided into 5 stages. Cell surface marker profiles were indicated by using the following abbreviations: T, THY1 (black ring or circle); M, MAC1 (black ring or circle); 4, A4iG (black ring or circle); E, EPCAM; S, SSEA1 (pink ring or circle); O, OCT4 (green ring or circle); p, passage.
See also Figures S1–S3.



(legend on next page)

the alternative neutrophil pathways (MAC1$^-$/OCT4$^+$/SSEA1$^-$, MAC1$^-$/OCT4$^-$/SSEA1$^+$ cells; see Figure 1A) were transcriptionally similar with *Fut9*, the enzyme responsible for the SSEA1 epitope, the most differentially expressed gene (Table S1). Although these differences in transcriptome confirm that neutrophils can follow different molecular pathways while they undergo reprogramming, for the rest of our analyses, we focused on the dominant reprogramming pathway (OCT4 preceding activation of SSEA1). As previously identified in fibroblasts, we also observed two major transcriptional waves in the reprogramming pathways of neutrophils and keratinocytes (Figures 2C and 2D). Importantly, most genes that changed in both waves were cell-type specific (Figure 2D). Furthermore, genes associated with the identity of each cell type (and therefore mostly cell-type specific) were downregulated in fibroblasts, neutrophils, and keratinocytes, but we also observed that the upregulated genes were largely cell-type specific as well (Figure 2D).

In summary, the three investigated cell types follow a series of molecular events, which, during the early stages, were dominated by cell-type-specific changes. However, we identify the presence of two transcriptional waves, which, despite involving mostly different genes in each cell type, is seemingly a universal feature of the reprogramming process.

The General Transcriptional Changes of All Three Cell Types Are Largely Cell- Type Specific during Reprogramming

Next, we investigated the dynamic changes occurring during the reprogramming of these cell types. For comparative purposes, we divided the reprogramming process of each cell type into five stages (Figure 1G), which were matched according to their locations in the COA (Figure 2A). Fuzzy c-means clustering identified highly correlated gene clusters that were changing in a coordinated manner over time in each cell type (Figure S4A). By comparing the clusters from each cell type with each other, we uncovered a transcriptional bias (Figure 3A). Gene clusters associated with transcriptional upregulation were more correlated between cell types than gene clusters for transient reactivation or downregulation (Figure 3A). To quantify this, we compared the degree of transcriptional correlation between different sets of gene dynamics (upregulated, downregulated, and transiently activated) for each pairwise comparison between cell types (Figures 3B–3F; Table S2). Overall, we found that 24% of dynamic gene changes are highly correlated (Pearson's correlation > 0.7) (Figure 3B), but in the case of genes that are upregulated, this increased to 28% of genes (Figure 3D), whereas this decreased in downregulated genes to 14% (Figure 3C). Surprisingly, genes that get transiently reactivated in each cell type were poorly correlated (15%; Figure 3E), indicating that transient transcriptional events during reprogramming are mostly cell-type specific. Of note, 54% of genes that do change in all cell types during reprogramming (893 genes) are highly correlated (Figure 3F), implying that cell-type-specific transcriptional changes associated with the original cell identity mask a partially conserved core of gene expression changes shared by all three cell types. Gene ontology (GO) analysis of these conserved genes revealed significant association with germ cell development, transcription, cell differentiation, as well stem cell maintenance, including the crucial epigenetic remodeling enzyme Tet1 (Figure 3F; Table S2). In summary, these results show that the kinetics of a significant proportion of transcriptional changes during reprogramming are cell-type specific, which indicates that during reprogramming, different sets of genes follow particular dynamics in different cell types.

Germ Layer of Origin Genes Follow Specific Transcriptional Dynamics

To further investigate whether loss of the original cell identity masks conserved transcriptional dynamics, we discarded genes that were differentially expressed in at least one pairwise comparison between the starting somatic cell populations. We identified 3,496 genes that had comparable expression levels on day 0, and by default at the end of reprogramming when all cells reached the iPS state. A COA of the expression profile for this gene subset in each of the three cell types (Figure 4A) unmasked cell-type-specific transcriptional trajectories for keratinocytes compared with fibroblasts and neutrophils (in component three). Furthermore, the top genes involved in this separation are significantly associated with ectodermal and mesodermal development (Figures 4B–4D), implying that the transcriptional kinetics of some genes, although expressed at comparable levels in all three starting populations, may be influenced by the germ layer of origin.

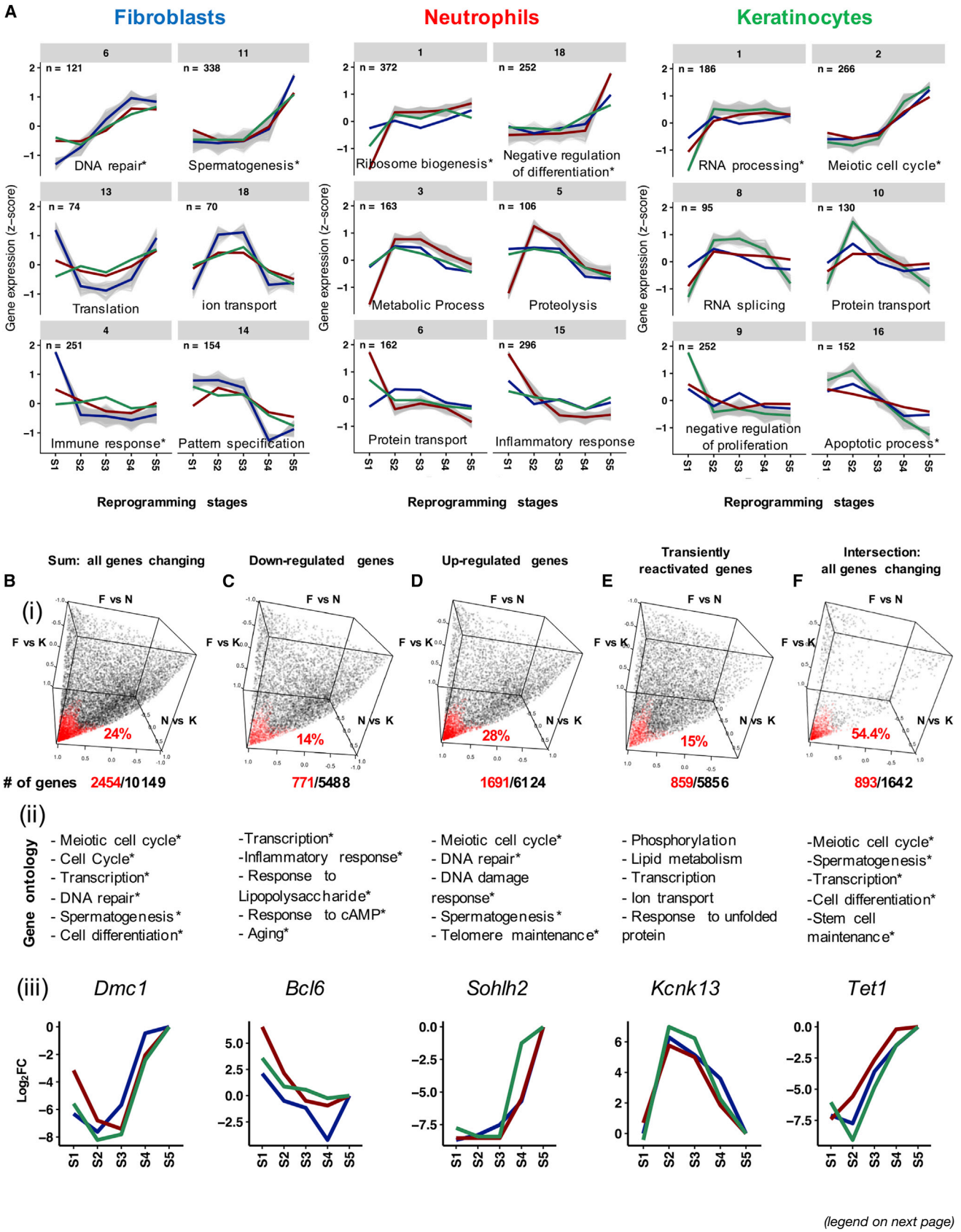
Yamanaka and colleagues in human reprogramming and our work in mouse fibroblasts have previously postulated that late-stage reprogramming cells pass through a primitive streak-like state, an early embryonic structure that precedes the formation of the three germ layers (Takahashi et al., 2014). We further confirm this with our current mouse dataset for fibroblasts, which shows a clear transient upregulation of primitive streak genes at stage 4 (Figure 4E). However, at least in the mouse system, this does not appear to be a universal phenomenon as was previously thought since both neutrophils and keratinocytes do not go through a similar transition (Figure 4E). In addition to primitive streak genes, we identified a specific cluster of transiently reactivated genes in fibroblasts (but not in the other two cell types) that are associated with gastrulation (Figure 4F). We also observed that some genes associated with extraembryonic

Figure 2. General Molecular Changes during Reprogramming

(A–C) Correspondence analysis (COA; A) and unsupervised hierarchical clustering (B) of combined RNA sequencing datasets of fibroblasts, neutrophils, and keratinocytes. (C) Stacked bar plots depicting the number of genes that are up- or downregulated from one transition to the next. Cell surface marker profiles and isolation days of reprogramming intermediates were indicated by using the following abbreviations: T, THY1; M, MAC1; 4, A4iG; E, EPCAM; S, SSEA1; O, OCT4; K, cKIT; p, passage, D, day.

(D) Set analysis of genes that are up- or downregulated during the first (left) or second (right) transcriptional wave of reprogramming; asterisks indicate associated GO categories for which $p < 0.05$ (Benjamini-Hochberg).

See also Table S1.



(legend on next page)

endoderm development peak transiently in reprogramming fibroblasts (Figure 4G). Furthermore, we show that genes associated with germ layer commitment were transiently upregulated at stage 4, an observation most pronounced in fibroblasts alone (Figure 4H), providing further evidence that this transition through a state reminiscent of early embryonic development is specific to fibroblasts.

Because a MET has been described as an essential and early aspect of fibroblast reprogramming, and since keratinocytes are already in an epithelial state, we assessed whether the widespread transcriptional changes during reprogramming also entail a temporary loss of epithelial genes in this cell type (Figure 4I). Interestingly, reprogramming of keratinocytes indicates that a MET per se is not an obligatory process during iPSC generation; mesenchymal fibroblasts and neutrophils need to lose their cell-specific signatures, including their mesenchymal genes, and later undergo epithelialization, which is seen as a MET. However, keratinocytes do not go through an epithelial-to-mesenchymal transition (EMT) followed by a MET (i.e., loss of the epithelial signature during the phase of downregulation of cell-type-specific genes only to regain these genes later), and thus are able to maintain their original epithelial gene signature throughout the reprogramming process (Figure 4I). In contrast to this event, keratinocytes expressed ICAM1, a cell surface marker that has previously been used to isolate reprogramming intermediates together with the somatic marker CD44 (O'Malley et al., 2013), at high levels prior to reprogramming, but lost this marker transiently at stage 2 before gradually re-expressing it on route to the iPSC state (Figure S4B). CD73 and CD200, antigens implicated as early reprogramming markers for fibroblast intermediates in a different study (Lujan et al., 2015), reached similar expression levels in all three cell types at stages 3 and 2 respectively (Figure S4B).

In summary, we show that a significant proportion of transcriptional changes during reprogramming are cell-type specific, including genes that are not differentially expressed between the starting populations. Furthermore, we provide evidence that a transition through a primitive streak-like state is not a universal attribute of the mammalian reprogramming process.

Somatic Identity Loss and Reactivation of the Pluripotency Network Are Cell-Type-Specific Processes with Conserved Aspects Shared by the Three Cell Types

Finally, we sought to compare key events that all successfully reprogramming cells (independent of cell type) need to undergo to become iPSCs, namely (1) loss of original cell identity and (2) reactivation of the pluripotency network. (1) By using Mogrify (Rackham et al., 2016), we determined the transcription factor (TF) networks that govern cell identity in the starting cell types and studied how these networks were lost over time. In a stage-matched context, loss of the fibroblast identity network is slower compared with neutrophils and keratinocytes (Figure 5A; Table S3). However, it appears that the speed of identity loss is not predictive of actual reprogramming time, as keratinocytes take >10 days longer to become iPSCs compared with fibroblasts (Figure S3A). We then used the same algorithm to identify core TFs that are specifically downregulated in the initial phase of reprogramming in each cell type's pathway as well as core TFs that are downregulated in all three cell types (Figure 5B). Although distinct TFs appear to get downregulated in a cell-type-specific manner during reprogramming to the pluripotent state (e.g., *HoxB7*, *Pparg*, and *Hmga2* in fibroblasts, *Runx2*, *CebpB*, and *Ep300* in neutrophils, and *HoxB5*, *Sox9*, and *Ahr* in keratinocytes), *Egr1* and *Stat3* were identified as core identity TFs for all these cell types (Figure 5B). However, as downregulation of *Stat3* in iPSCs relative to stage 1 was mild, but very drastic for *Egr1* (>32-fold, Figure S5A) and *Egr1* overexpression has previously been shown to inhibit reprogramming (Worringer et al., 2014), we focused on *Egr1* to functionally validate this in fibroblasts. Interestingly, *Egr1* knockdown boosted reprogramming when cells were infected at low MOIs, whereas at higher MOIs, a large proportion of refractory cells were killed early during reprogramming, thereby giving lower efficiencies when measured as a proportion of the number of starting cells, but further enriching cultures for SSEA1$^+$ reprogramming intermediates (Figure 5C; Figures S5B and S5C), and therefore giving rise to almost pure cultures. In the later stages of reprogramming (day 6 and beyond), *Egr1* knockdown enhanced (albeit did not accelerate) the transition of SSEA1$^+$ cells toward an SSEA1/EPCAM double-positive state by ~3-fold, indicating that *Egr1* knockdown can impact reprogramming both at early and late stages of the process (Figure 5D; Figure S5D).

(2) In comparison with the largely cell-type-specific identity loss, the activation dynamics of 48% of core pluripotency network genes were highly correlated (Pearson's correlation > 0.7) between the different cell types (Figure 5E; Table S3). This was the case despite the fact that only 30%–33% of direct OCT4, KLF4, and SOX2 (OKS) target genes were highly correlated between the different cell types (Table S3; Figures S5E–S5P), implying that OKS genes, in addition to promoting

Figure 3. Cell-Type-Specific and Conserved Aspects of Reprogramming

(A) Within each cell type's dataset, representative gene clusters were identified that follow certain kinetics (upregulation, downregulation, or transient reactivation; see Figure S4A for all clusters; specific cluster number is indicated in gray panels). Cluster member genes ($\alpha > 0.6$) were overlaid with the cumulative trend lines of the respective behavior of genes in the other 2 cell types and the key ontology category indicated (Database for Annotation, Visualization and Integrated Discovery [DAVID]; *$p < 0.05$, Benjamini-Hochberg FDR).

(Bi–Fi) Correlation analyses for: (B) the union of all genes that have a >4-fold linear change in expression between at least one of the starting cell types and iPSCs; (C and D) the subset of genes that get downregulated (C) or upregulated (D) in at least one of the cell types with a >4-fold linear change in expression between day 0 and the iPSC state; (E) the subset of genes that do not change more than 4-fold linear during reprogramming, but get transiently reactivated in at least one of the cell types at some stage during iPSC generation; and (F) the intersection of genes that have at least a 4-fold linear change in expression in all cell types between day 0 and the iPSC state. Highly correlated (HC) genes were defined to have a Pearson's correlation value >0.7 for at least two pairwise comparisons and are indicated in red.

(Bii–Fii) KeyGO categories associated with the highly correlated genes (*$p < 0.05$, Benjamini-Hochberg FDR).

(Biii–Fiii) Transcriptional kinetics for one example of the highly correlated genes.

See also Figure S4 and Table S2.

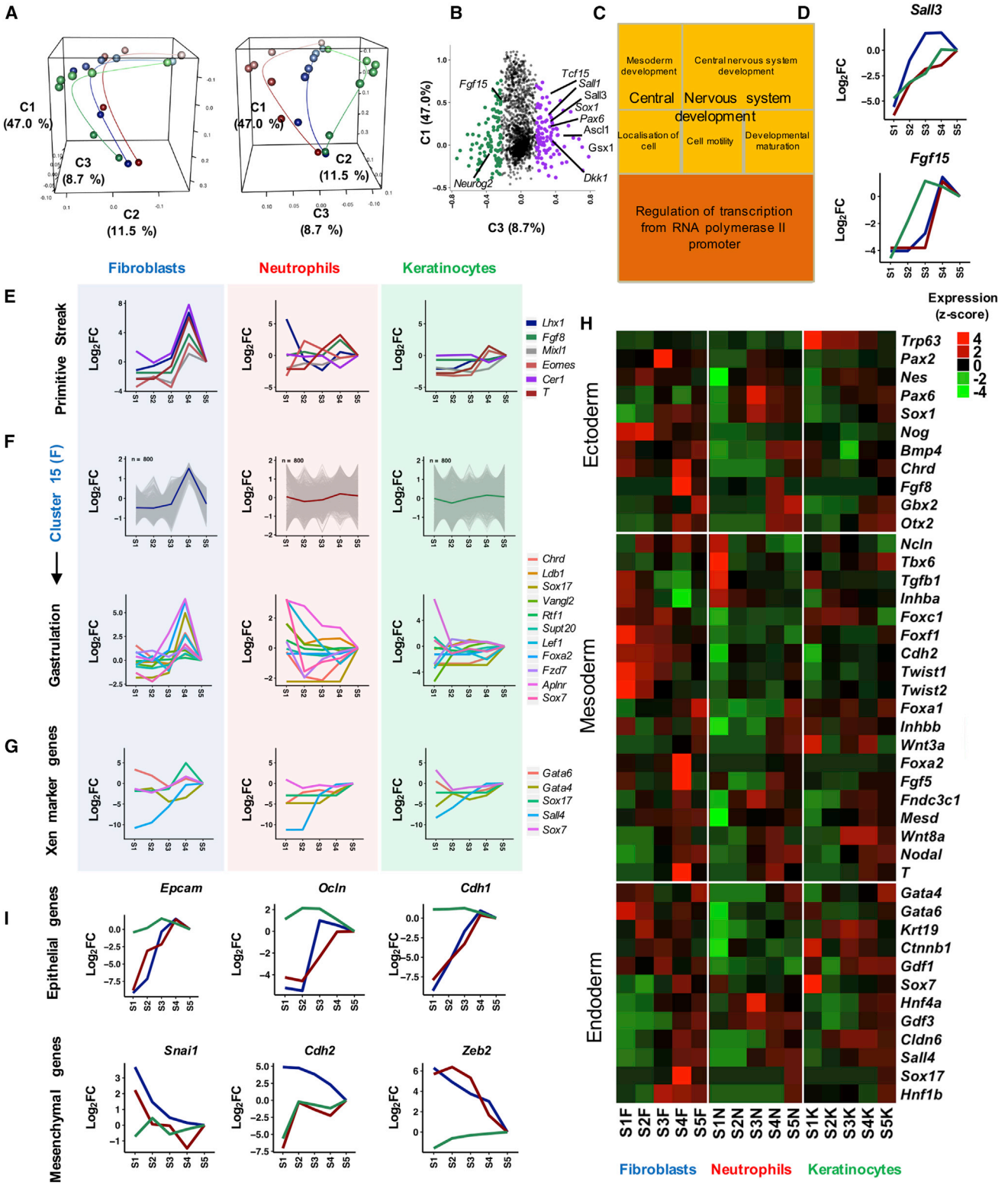


Figure 4. Cell-Type-Specific Changes during Reprogramming

(A) COA of cell-type-specific datasets after discarding genes that are differentially expressed in at least one of the pairwise comparisons between the starting cell types (day 0).

(legend continued on next page)

activation of the pluripotency network, likely modulate other cell-type-specific processes, such as identity loss. In a stage-matched context, neutrophils not only reactivate *Oct4* earlier during reprogramming (Figure 1G) but are also able to reactivate the pluripotency-associated network faster (Figures 5F and 5G), while conversely, fibroblasts, the cell type that is most gradually shedding its identity TF network (Figure 5A), only reactivated the pluripotency network afterward. In conclusion, the loss of somatic identity gene expression occurs in a cell-type-specific manner; however, downregulation of *Egr1* is shared by all three cell types, and its knockdown enhances reprogramming. Reactivation of the pluripotency network, like identity loss, has a strong cell-type-specific component with cell-type-specific transcriptional dynamics during reprogramming.

DISCUSSION

Our analysis of the reprogramming pathways of different cell types demonstrates that, although the initial stages are dominated by cell-type-specific transcriptional changes associated with identity loss, these changes mask a conserved transcriptional modulation core shared by all three cell types. Specifically, 54% of the genes that need to change in all three cell types and 48% of pluripotency network TFs are highly correlated throughout the reprogramming process. In agreement with a previous study characterizing the forced expression of *Oct4* alone in three different cell systems (Tiemann et al., 2014), we show that the forced expression of OSKM also resulted in largely cell-type-specific transcriptional changes. However, while *Oct4* overexpression alone resulted in less than 1% of the transcriptional changes being induced in all three cell types (Tiemann et al., 2014), we show that OKSM overexpression results in 24% of all transcriptional changes being highly correlated between our three cells types, including 30% of all OCT4 target genes. Accordingly, coexpression of all four Yamanaka factors synergize to induce a partially conserved transcriptional signature during reprogramming. Furthermore, we cannot exclude the possibility that different TFs upregulated during reprogramming access their targets differentially due to cell-type-specific epigenomic landscapes.

Nevertheless, a significant proportion of the reprogramming process is cell-type specific, and consequently previous studies that provided high-resolution maps for the reprogramming process in fibroblasts have only been able to unveil specific changes to this cell type and in turn convey an incomplete picture. As a consequence, this may impact our understanding of the reprogramming process; for example, the role of the reactivation kinetics of *Oct4* is not consistent during reprogramming of cells of the hematopoietic system and fibroblasts, and, as such, whether we should consider it a late or a definitive marker for iPSC derivation is called into question. In further support of this, we identified a number of genes, which, although expressed at comparable levels in all starting cell types, do behave differently during reprogramming, and our results suggest these differences are related to the germ layer of origin.

Another cell-type-specific feature that we identified in reprogramming mouse cells is the degree to which they undergo a form of developmental reversion during the late stages of reprogramming. While we and others have previously found that fibroblasts undergo a transient upregulation of primitive streak genes, surprisingly neither neutrophils nor keratinocytes appeared to display a transition through a primitive streak-like state. This is in contrast to studies by Yamanaka and colleagues, where they concluded that reprogramming of different human cells entailed a transition through a primitive streak-like state (Takahashi et al., 2014). This seems to indicate that traversing through a primitive streak-like state is not a universal feature of mammalian cell reprogramming. However, because our study was based on the profiling of purified reprogramming intermediates on a bulk level, we cannot exclude that a small subset of neutrophil or keratinocyte intermediates pass through a primitive streak-like state, which is a limitation of this current study.

In addition to cell-type-specific features, we also identified seemingly universal aspects of the reprogramming process of mouse cells, namely the presence of two transcriptional waves. The first wave is largely associated with identity loss, and the second with reactivation of the pluripotency network. TF network analysis of these waves allowed us to identify factors, like *Egr1*, that are downregulated in all cell types and appear crucial for reprogramming and whose modulation facilitated enhanced reprogramming efficiency and the removal of refractory cells. Overall, our study unveils a large cell-type-specific component to the reprogramming process, which is partially the consequence of a restricted developmental reversion determined by the cell type of origin. Accordingly, early reprogramming intermediates have distinct cell-type-specific molecular signatures, a finding with possible implications for direct reprogramming strategies using transient OKSM expression. By revealing universal and cell-type-specific components of the reprogramming process, our findings show limitations for the use of only fibroblasts as a sole model for its study.

(B) Gene projections over components 2 and 3 for genes in (A); the top 200 genes, ranked according to their contribution to the variance of component 3, are highlighted.
(C) Treemap plot depicting GO super-categories associated with the most variable genes in C3.
(D) Example genes that were not found to be differentially expressed in the starting populations, but have different kinetics during reprogramming.
(E) Depiction of primitive streak genes (*Lhx1*, *Fgf8*, *Mixl1*, *Eomes*, *Cer1*, and *T*) during reprogramming of all three cell types.
(F) Visualization of fibroblast Cluster 15 genes (Figure S5) in all three cell types and visualization of the subset of genes associated with GO category "gastrulation."
(G) Expression kinetics of extraembryonic endoderm (XEN) genes during reprogramming.
(H) Heatmap depicting average expression of early ecto-, endo-, and mesodermal genes during reprogramming (stages 1–5) of the three cell types.
(I) Expression of key epithelial and mesenchymal marker genes during reprogramming in all three cell types.
See also Figure S4 and Table S2.

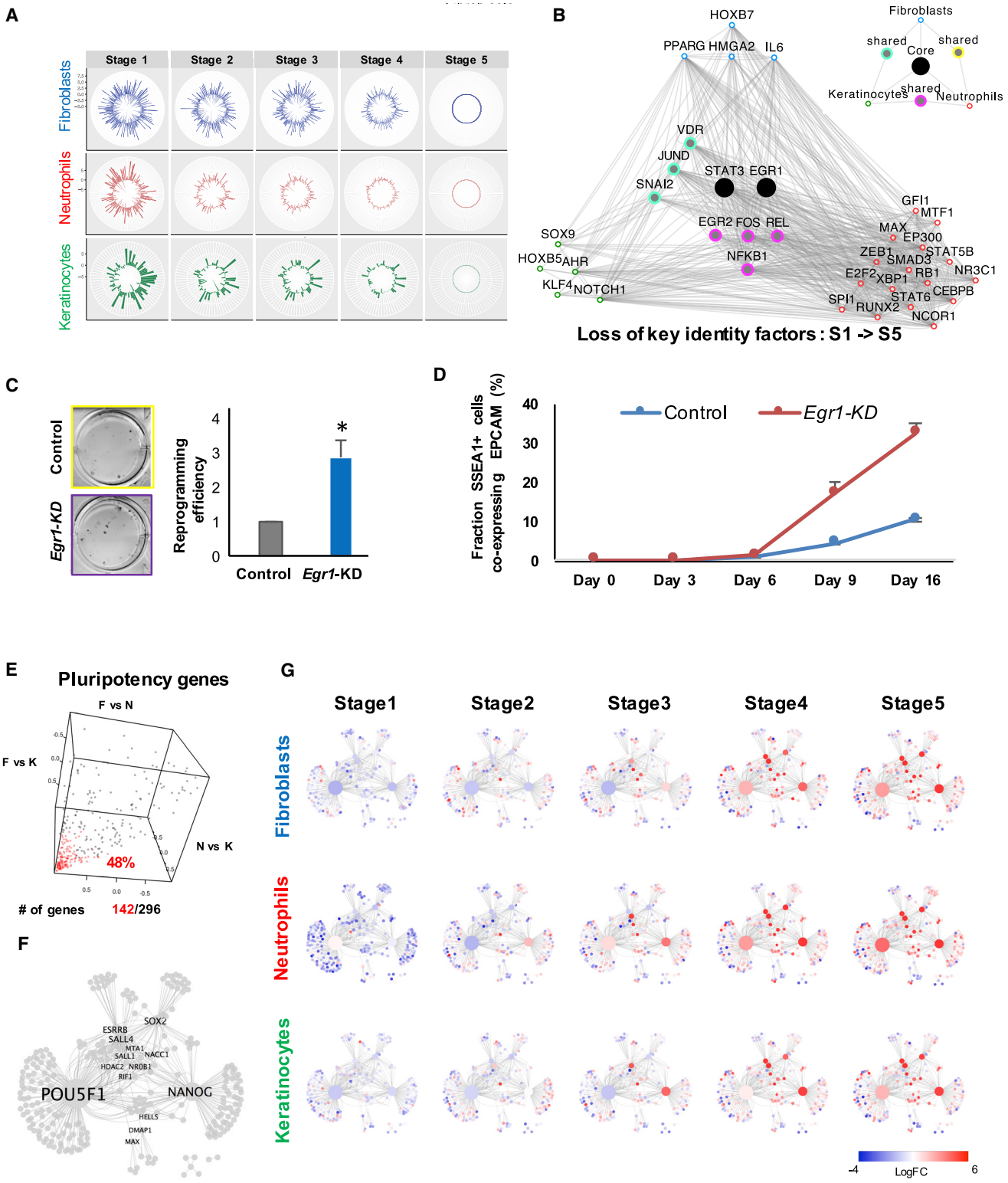


Figure 5. Loss of Original Cell Identity and Reactivation of the Pluripotency Network

(A) Polar plots depicting stage-wise loss of key identity genes identified by the Mogrify algorithm (Table S3).

(B) Mogrify network analysis depicting key identity TFs for each of the initial cell types. Cell-type-specific loss of TFs, as well as TFs that are shared by two or more cell types, are displayed.

(legend continued on next page)

EXPERIMENTAL PROCEDURES

Animals were housed in specific pathogen-free animal house conditions at the animal facility (Monash Animal Services) in strict accordance with good animal practice as defined by the National Health and Medical Research Council (Australia) Code of Practice for the Care and Use of Animals for Experimental Purposes. Experimental procedures were approved by the Monash Animal Research Platform Animal Ethics Committee.

Statistical Analyses

All experiments were performed as experimental and/or biological replicates of three, with the exception of RNA-seq experiments, which were performed as biological replicates of two. In detail, the following experiments were performed as experimental and biological replicates: Figures 1A, 1B, and 1E; or biological replicates alone: Figures 1C, 1D, 1F, 5C, 5D, S1–S3, and S5B–S5D. Multidimensional scaling analysis and visualization were performed using limma's plotMDS function (Ritchie et al., 2015). Correspondence analysis and visualization were performed using made4 (Culhane et al., 2005) and the plot3d function of the RGL package (Murdoch, 2001). Unsupervised hierarchical clustering (Euclidean distance) were performed using bioDist (Ding et al., 2017) and hclust (R Core Team, 2017). Unless otherwise specified, charts and plots were produced using gplots (Warnes et al., 2016) RGL's plot3d, and ggplot2 (http://ggplot2.org/).

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, Jose M. Polo (jose.polo@monash.edu).

For additional information, see Supplemental Experimental Procedures.

DATA AND SOFTWARE AVAILABILITY

The accession number for whole transcriptome sequencing experiments reported in this paper is SRA: SRP119979.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and three tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.11.029.

AUTHOR CONTRIBUTIONS

C.M.N., F.J.R., O.J.L.R., and J.M. Polo conceived the study and designed the experiments; C.M.N. performed somatic cell reprogramming and molecular experiments with support from J.C., X.L., A.S.K., J.F., S.A., K.F.-C., and J.M. Paynter; S.K.N. and B.W. guided and supported the hematopoietic reprogramming experiments; S.M.L performed the teratoma experiments; C.M.N. performed the FACS experiments and SPADE analyses; F.J.R. analyzed the RNA sequencing data under the guidance of J.M. Polo, O.J.L.R., D.R.P., E.P., and M.R.; R.L., J.P., S.B., and J.F. provided reagents and technical assistance; C.M.N., F.J.R, O.J.L.R., and J.M. Polo wrote the manuscript; and all authors approved and contributed to the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the high-quality cell sorting service and technical input provided by the Monash University FlowCore facility. Furthermore, the authors thank the ACRF Centre for Cancer Genomic Medicine at the MHTP Medical Genomics Facility for assistance with next generation library preparation and Illumina sequencing. This work was supported by National Health and Medical Research Council project grant APP1051309 to J.M. Polo. A.S.K. was supported by National Health and Medical Research Council Early Career Fellowship APP1092280. X.L. and S.A. were supported by a Monash International Postgraduate Research Scholarship and a Monash Graduate Scholarship. J.M. Polo and R.L. were supported by Silvia and Charles Viertel Senior Medical Research Fellowships. O.R. was supported by National Medical Research Council (Singapore) grant OFYIRG16may050. Owen Rackham and Jose Polo are founders and directors of Cell Mogrify. The other authors do not declare any conflict of interest.

Received: June 20, 2017
Revised: October 2, 2017
Accepted: November 8, 2017
Published: December 5, 2017

REFERENCES

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687–696.

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

Bryder, D., Rossi, D.J., and Weissman, I.L. (2006). Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. Am. J. Pathol. *169*, 338–346.

Chantzoura, E., Skylaki, S., Menendez, S., Kim, S.-I., Johnsson, A., Linnarsson, S., Woltjen, K., Chambers, I., and Kaji, K. (2015). Reprogramming Roadblocks Are System Dependent. Stem Cell Reports *5*, 350–364.

Culhane, A.C., Thioulouse, J., Perrière, G., and Higgins, D.G. (2005). MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics *21*, 2789–2790.

Ding, B., Gentleman, R., and Carey, V. (2017). bioDist: Different distance measures. R package version 1.50.0.

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., et al. (2014). Genome-wide characterization of the routes to pluripotency. Nature *516*, 198–206.

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

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.

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

(C) Alkaline phosphatase labeling (six-well format) and relative changes in numbers of alkaline-positive colonies after infection with *Egr1* short hairpin RNA (shRNA) at MOI = 1 at the start of reprogramming (n = 3 biological replicates; error bars represent SD).
(D) Graphical representation of the fraction of SSEA1$^+$ cells that coexpress EPCAM at the indicated time points during reprogramming in the presence or absence of *Egr1* shRNA-mediated knockdown (n = 3 biological replicates; Student's t test, *p > 0.05; error bars represent SD).
(E) Correlation scatterplot depicting the number and percentage of pluripotency network genes that are correlated during reprogramming. Highly correlated (HC) genes were defined to have Pearson's correlation values of >0.7 for at least two pairwise comparisons and are indicated in red.
(F and G) Visualization of the reactivation of key pluripotency network nodes during reprogramming of the three cell types; (F) depicts the network in detail and (G) depicts the change over time.
See also Figure S5 and Table S3.

Murdoch, D (2001). RGL: An R Interface to OpenGL.

Nefzger, C.M., Jardé, T., Rossello, F.J., Horvay, K., Knaupp, A.S., Powell, D.R., Chen, J., Abud, H.E., and Polo, J.M. (2016). A Versatile Strategy for Isolating a Highly Enriched Population of Intestinal Stem Cells. Stem Cell Reports *6*, 321–329.

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

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

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Jr., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat. Biotechnol. *29*, 886–891.

R Core Team (2017). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

Rackham, O.J.L., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., Suzuki, H., Nefzger, C.M., Daub, C.O., Shin, J.W., et al.; FANTOM Consortium (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

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

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

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

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat. Commun. *5*, 3678.

Tiemann, U., Marthaler, A.G., Adachi, K., Wu, G., Fischedick, G.U.L., Araúzo-Bravo, M.J., Schöler, H.R., and Tapia, N. (2014). Counteracting activities of OCT4 and KLF4 during reprogramming to pluripotency. Stem Cell Reports *2*, 351–365.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., et al. (2016). gplots: Various R programming tools for plotting data. R package version 3.0.1. https://cran.r-project.org/web/packages/gplots/index.html.

Weissman, I.L., and Shizuru, J.A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. Blood *112*, 3543–3553.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Worringer, K.A., Rand, T.A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., and Yamanaka, S. (2014). The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. Cell Stem Cell *14*, 40–52.

Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M., and 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.