Cell

# Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase

Yosef Buganim,[1,7] Dina A. Faddah,[1,2,7] Albert W. Cheng,[1,3] Elena Itskovich,[1] Styliani Markoulaki,[1] Kibibi Ganz,[1] Sandy L. Klemm,[5] Alexander van Oudenaarden,[2,4,6] and Rudolf Jaenisch[1,2,*]
[1]The Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology
[3]Department of Computational and Systems Biology
[4]Department of Physics
[5]Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands
[7]These authors contributed equally to this work
*Correspondence: jaenisch@wi.mit.edu
http://dx.doi.org/10.1016/j.cell.2012.08.023

## SUMMARY

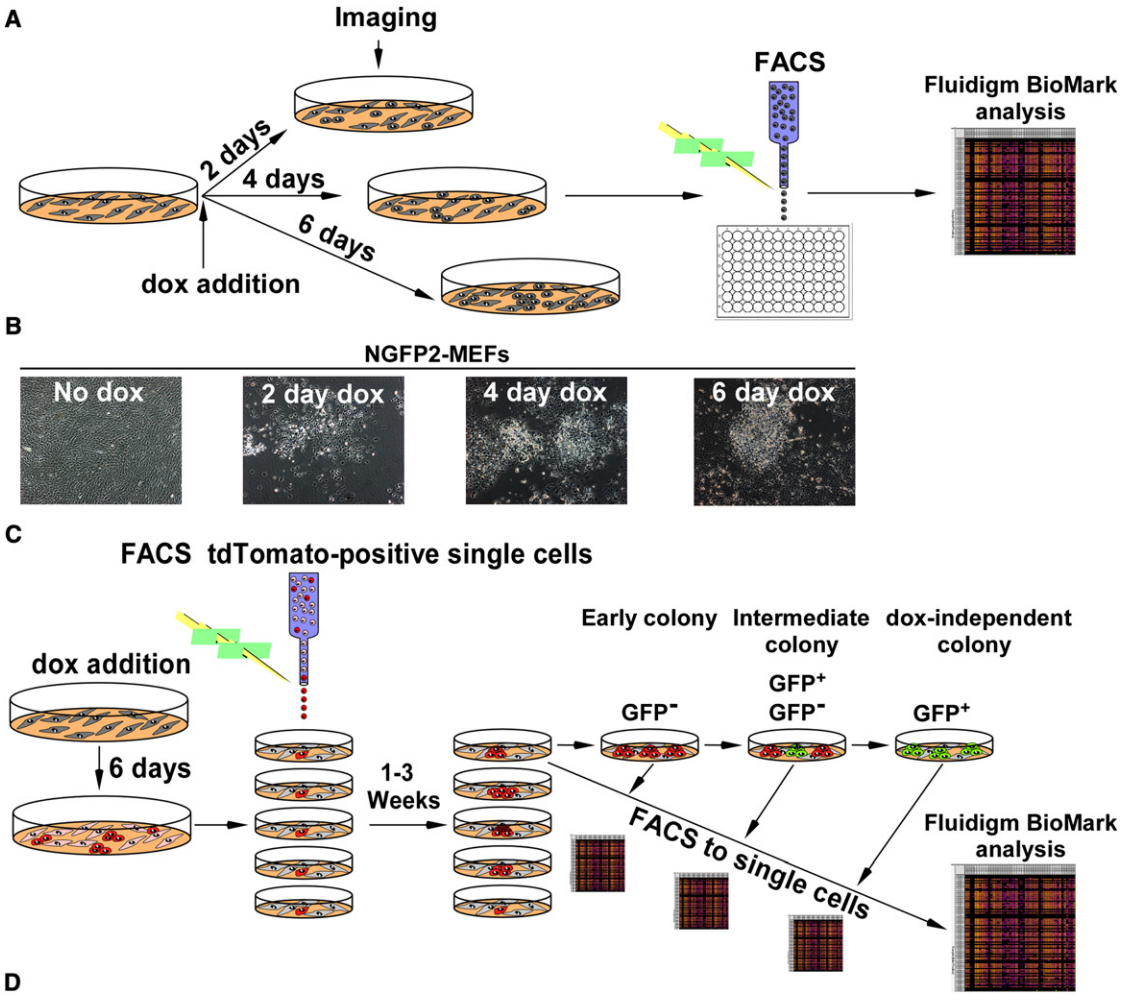
During cellular reprogramming, only a small fraction of cells become induced pluripotent stem cells (iPSCs). Previous analyses of gene expression during reprogramming were based on populations of cells, impeding single-cell level identification of reprogramming events. We utilized two gene expression technologies to profile 48 genes in single cells at various stages during the reprogramming process. Analysis of early stages revealed considerable variation in gene expression between cells in contrast to late stages. Expression of Esrrb, Utf1, Lin28, and Dppa2 is a better predictor for cells to progress into iPSCs than expression of the previously suggested reprogramming markers Fbxo15, Fgf4, and Oct4. Stochastic gene expression early in reprogramming is followed by a late hierarchical phase with Sox2 being the upstream factor in a gene expression hierarchy. Finally, downstream factors derived from the late phase, which do not include Oct4, Sox2, Klf4, c-Myc, and Nanog, can activate the pluripotency circuitry.

## INTRODUCTION

Differentiated cells can be reprogrammed to a pluripotent state by overexpression of Oct4, Sox2, Klf4, and c-Myc (OSKM) (Takahashi and Yamanaka, 2006). Fully reprogrammed induced pluripotent stem cells (iPSCs) can contribute to the three germ layers and give rise to fertile mice by tetraploid complementation (Okita et al., 2007; Zhao et al., 2009). The reprogramming process is characterized by widespread epigenetic changes (Koche et al., 2011; Maherali et al., 2007; Mikkelsen et al., 2008) that generate iPSCs that functionally and molecularly resemble embryonic stem cells (ESCs).

To further understand the reprogramming process, transcriptional and epigenetic changes in cell populations were analyzed at different time points after factor induction. For example, microarray data showed that the immediate response to the reprogramming factors was characterized by dedifferentiation of mouse embryonic fibroblasts (MEFs) and upregulation of proliferative genes, consistent with c-Myc expression (Mikkelsen et al., 2008). It has been shown that the endogenous pluripotency markers Sox2 and Nanog are activated after early markers such as alkaline phosphatase (AP) and SSEA1 (Stadtfeld et al., 2008). Recently, gene expression profiling and RNAi screening in fibroblasts revealed three phases of reprogramming termed initiation, maturation, and stabilization, with the initiation phase marked by a mesenchymal-to-epithelial transition (MET) (Li et al., 2010; Samavarchi-Tehrani et al., 2010).

Given these data, a stochastic model has emerged to explain how forced expression of the transcription factors initiates the process that eventually leads to the pluripotent state in only a small fraction of the transduced cells (Hanna et al., 2009; Yamanaka, 2009). Most data have been interpreted to support a stochastic model (Hanna et al., 2009) posing that the reprogramming factors initiate a sequence of probabilistic events that eventually lead to the small and unpredictable fraction of iPSCs. Clonal analyses support the stochastic model, demonstrating that activation of pluripotency markers occurs at different times after infection in individual daughters of the same fibroblast (Meissner et al., 2007). However, because the molecular changes occurring at the different stages during the reprogramming process were based upon the analysis of heterogeneous cell populations, it has not been possible to clarify the events that occur in the rare single cells that eventually form iPSCs. Moreover, there has been little insight into the sequence of events that drive the process.




E

| Colony | Reprogramming characteristics | Day of cell collection (across 94 days) | | | |
|---|---|---|---|---|---|
| | | tdTomato+ | GFP+ | GFP+ dox independent | Dox withdrawal |
| 20 | Early reprogramming | 32 | 32 | 66 | 36 |
| 34 | | 32 | 32 | 61 | 36 |
| 43 | | 12, 45 | 45 | 61 | 41 |
| 16 | Late reprogramming | 81 | 81 | 94 | 81 |
| 23 | Induced cells that did not give rise to iPSCs | 12, 81 | NA | NA | NA |
| 44 | | 12, 61 | 61* | NA | NA |