

## STEM CELLS

## Revealing the reprogramming program

“ reprogramming can broadly be divided into ... an early phase that has a high degree of variability among cells ... and a later phase that is more hierarchical ”

In cellular reprogramming, the molecular steps between the starting somatic cells and the final induced pluripotent stem cells (iPSCs) have remained something of a ‘black box’. Some insights have been gained through studies of heterogeneous populations of cells undergoing reprogramming, but as only rare individual cells complete the journey to an iPSC, such studies have limitations. A recent paper presents the first single-cell analyses of gene expression during reprogramming, leading to a new model of the phases of reprogramming.

Buganim *et al.* used a previously developed reprogramming system in which mouse embryonic fibroblasts carry integrated doxycycline-inducible copies of the OSKM (that is, *Oct4*, *Sox2*, *Klf4* and *Myc*) reprogramming factors; a GFP reporter knocked into the endogenous *Nanog* locus also allows identification of iPSCs. After exposure to doxycycline for 6 days, single cells were seeded into separate wells, and at different times during the next 1–3 weeks the colonies derived from these single cells were split, and single cells were analysed. This strategy meant that gene expression could be studied in clonally related sister cells at different time points. The authors used two approaches to analyse gene expression: a Fluidigm microfluidics platform that allows quantitative real-time PCR analysis of 48 genes in 96 single cells per run; and single-molecule mRNA fluorescent *in situ* hybridization, which allows transcripts from up to three genes to be quantified in thousands of cells.

The authors carried out principal-component analysis on data from Fluidigm analysis of 1,864 cells sampled at different time points; the 48 genes studied were selected on the basis of processes that are implicated in reprogramming and pluripotency, as well as a number of controls. Principle components are patterns in high-dimensional data that explain a large percentage of the variation. Projecting the expression data onto the first two principal components revealed three cell groupings: unreprogrammed fibroblasts, iPSCs and a more heterogeneous group of cells representing an intermediate stage. The authors were also able to use these data to assess the most information-rich genes in

each cluster and the degree of variability among cells.

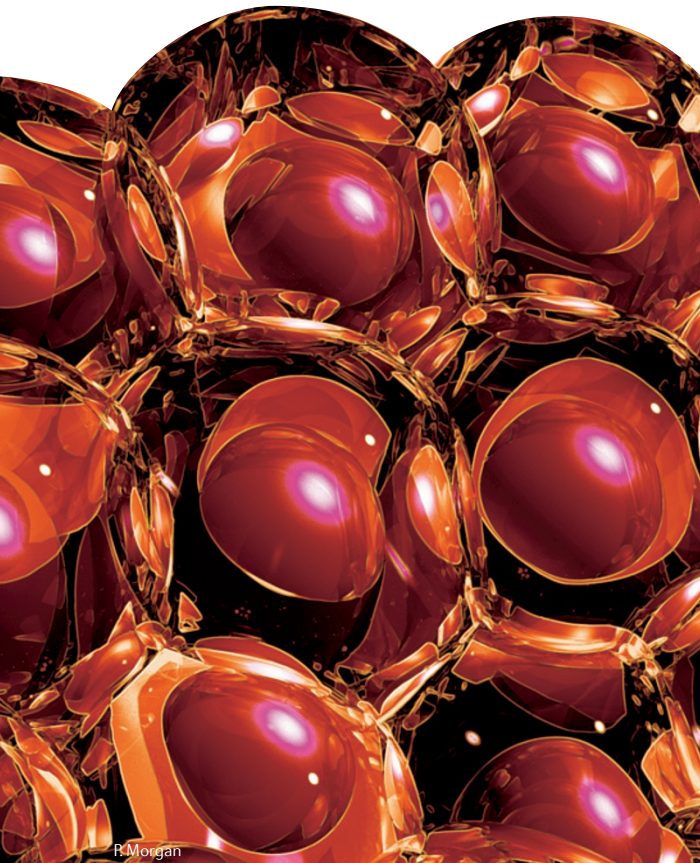
The wealth of data obtained from these analyses and follow-up experiments enabled the authors to identify genes that represent early, predictive markers of successful reprogramming to the iPSC state. In addition, they were able to conclude that reprogramming can broadly be divided into two phases: an early phase that has a high degree of variability among cells and that is consistent with the previously proposed ‘stochastic’ model for reprogramming, and a later phase that is more hierarchical (that is, it has a more ordered series of gene activation steps).

The authors used Bayesian network analysis to work out the hierarchical relationships among genes in this later phase, and they applied these findings to predict combinations of transcription factors that can achieve reprogramming. For example, they found that SOX2 is the top node of the network that drives a series of events allowing cells to become pluripotent. Factors further down the hierarchy, such as OCT4, could be replaced in reprogramming cocktails by certain factors that are higher up.

The authors comment that the stochasticity of gene expression in the early phase may reflect stochastic epigenetic changes; future single-cell analyses of such events will help to uncover the molecular underpinnings of the route to pluripotency.

Mary Muers

**ORIGINAL RESEARCH PAPER** Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012)  
**FURTHER READING** Plath, K. & Lowry, W. E. Progress in understanding reprogramming to the induced pluripotent state. *Nature Rev. Genet.* **12**, 253–265 (2011)



R. Morgan