

Direct cell reprogramming is a stochastic process amenable to acceleration

Jacob Hanna^{1*}, Krishanu Saha^{1*}, Bernardo Pando², Jeroen van Zon^{2,3}, Christopher J. Lengner¹, Menno P. Creighton¹, Alexander van Oudenaarden^{2,3} & Rudolf Jaenisch^{1,3}

Direct reprogramming of somatic cells into induced pluripotent stem (iPS) cells can be achieved by overexpression of Oct4, Sox2, Klf4 and c-Myc transcription factors, but only a minority of donor somatic cells can be reprogrammed to pluripotency. Here we demonstrate that reprogramming by these transcription factors is a continuous stochastic process where almost all mouse donor cells eventually give rise to iPS cells on continued growth and transcription factor expression. Additional inhibition of the p53/p21 pathway or overexpression of Lin28 increased the cell division rate and resulted in an accelerated kinetics of iPS cell formation that was directly proportional to the increase in cell proliferation. In contrast, Nanog overexpression accelerated reprogramming in a predominantly cell-division-rate-independent manner. Quantitative analyses define distinct cell-division-rate-dependent and -independent modes for accelerating the stochastic course of reprogramming, and suggest that the number of cell divisions is a key parameter driving epigenetic reprogramming to pluripotency.

Quantifying the efficiency and timescales of crucial events occurring during *in vitro* reprogramming to pluripotency^{1–5} has been problematic due to the cellular and genetic heterogeneity of *de novo* infected somatic cells^{6,7}. To circumvent the need for virus-mediated transduction and reduce the heterogeneity of reprogramming factor expression, a ‘secondary’ reprogramming transgenic system was devised where all somatic cells carry the same integration pattern of drug-inducible *Oct4*, *Sox2*, *Klf4* and *c-Myc* viral transgenes^{8–11}. Although reprogramming of somatic cells from secondary mice was two orders of magnitude higher than in freshly infected somatic cells, only 1–20% of the induced cells generated iPS cells after 3–4 weeks of factor expression^{9–13}. Furthermore, partially reprogrammed ‘intermediate’ cell lines have been derived in different experimental settings, some of which can give rise much later to fully reprogrammed iPS cells either spontaneously or upon additional manipulations^{14,15}. Given that the timescale of several weeks and relatively low efficiencies persist even after controlling for adequate *Oct4*, *Sox2*, *Klf4* and *c-Myc* transgene expression, these studies left important questions relevant to the basic mechanisms of epigenetic reprogramming unresolved: how does the reprogramming process progress over time and what happens to the majority of the cells that do not become reprogrammed upon continued cell growth and expression of the reprogramming factors? Why do some somatic cells that circumvent senescence or apoptosis induced by *Oct4*, *Sox2*, *Klf4* and *c-Myc* convert into iPS cells earlier than others? Do all adult donor cells expressing *Oct4*, *Sox2*, *Klf4* and *c-Myc* reprogramming factors eventually give rise to iPS cells or would this be achieved only upon additional genetic or small molecule manipulation? Is high reprogramming efficiency restricted to non-lineage committed or adult stem cells^{6,7,13,16,17}?

Models to account for the reprogramming process fall into two categories (models i–iv in Fig. 1). ‘Deterministic’ models posit that either ‘all’ (model i) or only a subset of ‘elite’ or ‘stem-like’ cells (model ii) within a donor population have the potential to generate iPS cells and are reprogrammed with a fixed latency. We define latency as the absolute time or the number of cell divisions that an individual donor cell undergoes until it gives rise to a daughter iPS

cell. ‘Stochastic’ models posit that most if not all (model iii) or only a subset of ‘elite’ somatic cells (model iv) within a donor population have the potential to generate iPS cells, albeit with different latencies.

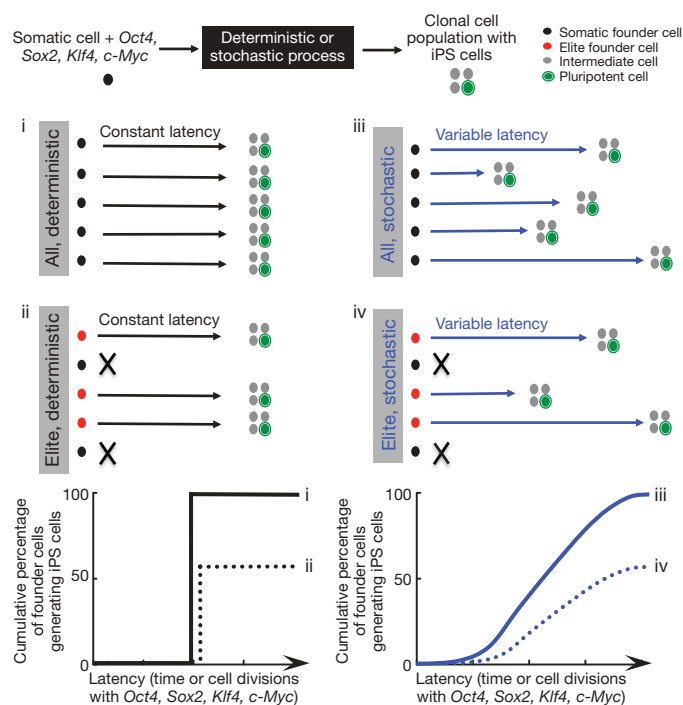


Figure 1 | Models of progressing to a pluripotent state during direct reprogramming. Four different models (i–iv) to account for the latency of donor somatic cells in progressing towards the iPS cell state following the expression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* reprogramming factors. Latency can be measured in units of absolute time or cell divisions until the first iPS cell is generated from a monoclonal population. Graphs display the general shape of the reprogramming kinetics in the different models. Note that elite models do not necessarily reprogram more slowly as shown in the bottom plots.

¹The Whitehead Institute for Biomedical Research, ²Department of Physics, ³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. *These authors contributed equally to this work.

To gain insight into the mechanism of reprogramming and to address some of these questions, we have characterized the reprogramming efficiency and kinetics of over 1,000 somatic-cell-derived monoclonal populations expressing *Oct4*, *Sox2*, *Klf4* and *c-Myc* over an extended period of time and quantitatively defined distinct modes for changing the course of the reprogramming process upon additional genetic perturbations.

Reprogramming of monoclonal populations

We followed the reprogramming of individual somatic donor cells and studied their potential to generate iPS cells. Unlike fibroblasts, B-cell lineage-committed cells at the early pre-B-cell stage can be efficiently cloned as single cells immediately after isolation and were used as a defined, homogenous starting cell population for reprogramming into iPS cells. The NGFP1 iPS cell line was generated by infecting fibroblasts from Nanog–GFP reporter mice with doxycycline-inducible lentiviral vectors encoding the *Oct4*, *Sox2*, *Klf4* and *c-Myc* transcription factors and injected into host blastocysts to generate secondary chimaeras^{9,18,19} (Fig. 2a). NGFP1-derived secondary pre-B cells were single-cell sorted into individual wells and exhibited high cloning efficiency on doxycycline (>80%). Populations were serially passaged and monitored weekly for reactivation of the endogenous Nanog–GFP knock-in reporter (Supplementary Fig. 1), which represents one of the final events during reprogramming^{4,20–22}. We defined reprogramming efficiency as the long-term potential of a cell to generate iPS daughter cells. A detection value of >0.5% for Nanog–GFP⁺ cells per well

reproducibly allowed for stable derivation of Nanog–GFP⁺ iPS cells upon doxycycline withdrawal (Supplementary Fig. 2), and was set as the minimal threshold for defining positive detection of iPS cells in clonal populations.

Nanog–GFP⁺ cells were detected after 2 weeks of doxycycline induction, with ~3–5% of the wells generating Nanog–GFP⁺ cells at 2 weeks^{10–12} (Fig. 2b). The remaining wells contained viable cells, could be propagated in the presence of doxycycline, and uniformly silenced somatic/haematopoietic surface markers^{15,22} (Supplementary Fig. 3). Upon extended culture, the fraction of wells with iPS cells gradually increased and by 18 weeks after doxycycline addition >92% of the wells had produced Nanog–GFP⁺ cells (Fig. 2b), demonstrating that most if not all donor cells have the potential to generate iPS cells. The reprogramming kinetics were reproducible in independent experiments (Fig. 2b) and were not restricted to B cells, as CD11b⁺ monocytes that showed comparable transgene induction levels generated iPS cells with similar kinetics and efficiency (Supplementary Fig. 4). Nanog–GFP⁺ cell populations selected at random gave rise to stable doxycycline-independent iPS cell lines. Furthermore, all tested iPS cell lines had normal karyotypes and generated teratomas and chimaeras irrespective of whether they were derived at early or late time points during the process (Fig. 2c and Supplementary Fig. 5). The iPS cell lines carried distinct genetic heavy-chain rearrangements verifying their independent clonal origin (Fig. 2d). We assessed whether transgene expression levels or increased proliferation rate could underlie the well-to-well differences seen in the latency of reprogramming. The

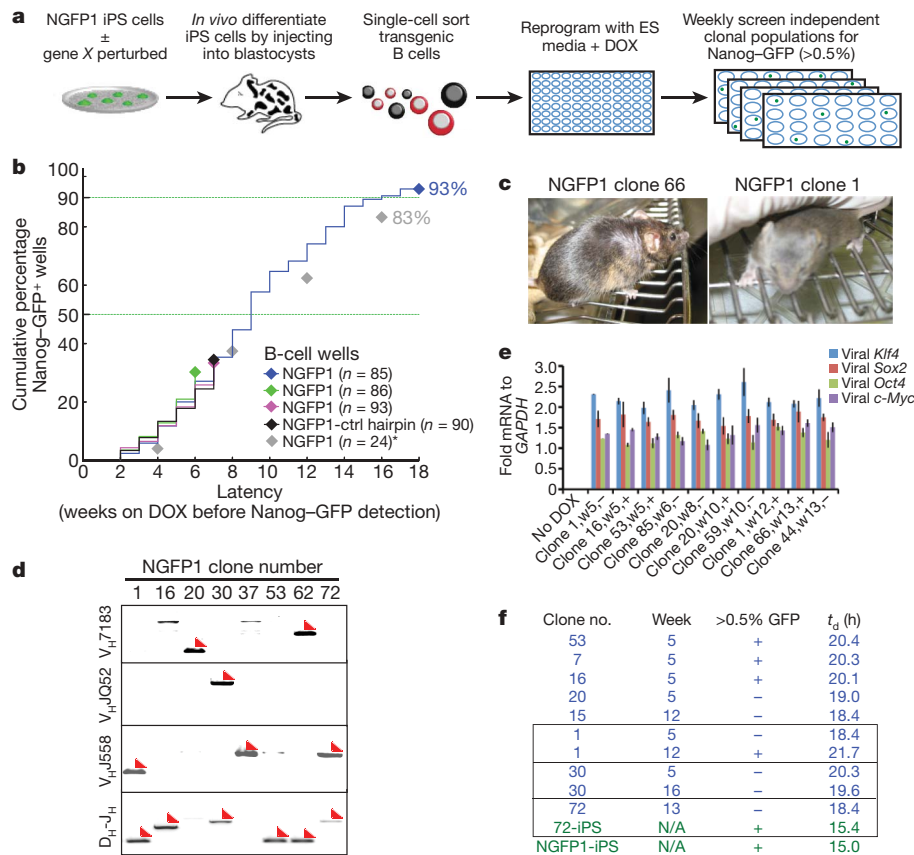


Figure 2 | Long-term analysis of reprogramming monoclonal populations.

a, Schematic of experiments. **b**, Reprogramming of pre-B-cell monoclonal populations measured as the cumulative number of wells that became Nanog–GFP⁺. *n* indicates number of populations monitored. Asterisk indicates flow cytometry for GFP detection was performed every 4 weeks. **c**, Chimaeric mice with agouti coat colour from iPS cells derived after 12–13 weeks of doxycycline (DOX). **d**, Heavy-chain rearrangements in iPS cells. Genetic rearrangements were confirmed by sequencing to distinguish them from background signals, and are highlighted with red triangles.

e, Relative transgene induction levels of monoclonal populations on doxycycline. Time is shown in weeks (w). Error bars indicate standard deviation (s.d.; *n* = 3). **f**, The population-averaged doubling time, t_d , for each clonal population. Boxes delineate cases where the same clonal population was measured at different times during doxycycline induction. The lower two rows (green) represent subcloned iPS cell lines. From left to right, columns show: clone number; weeks on doxycycline; Nanog–GFP >0.5% status (+/–); population-averaged doubling time (h).

population-averaged cell doubling times (t_d) and transgene induction levels during the reprogramming process were similar in NGFP1 clonal populations irrespective of time on doxycycline or whether these populations contained a Nanog-GFP⁺ fraction (Fig. 2e, f).

These results suggest the following: (1) reprogramming of somatic cells is a continuous stochastic process where nearly all somatic donor cells have the ability to give rise to iPS cells upon continuous passaging and expression of *Oct4*, *Sox2*, *Klf4* and *c-Myc*. (2) Although reprogrammed cells do not appear before 8–10 days of *Oct4*, *Sox2*, *Klf4* and *c-Myc* expression^{21,22}, the time of doxycycline exposure or number of cell divisions achieved before a given clonal population generates iPS cells varies widely. (3) Our data are not consistent with an ‘elite component’ in reprogramming, as most if not all lineage-committed B cells or monocytes are able to generate iPS cells rather than only a small fraction of putative somatic stem cells present in the donor cell population^{7,12,13}. (4) Finally, somatic cells reprogram with different latencies that cannot be predicted on the basis of time of doxycycline exposure or proliferation rate, consistent with undefined stochastic events driving the process (model iii in Fig. 1).

Cell division rate and reprogramming

We next characterized parameters of the reprogramming process by introducing defined genetic perturbations. Recently, *p53* (also called *Trp53*) inhibition has been shown to enhance the efficiency of iPS cell formation from fibroblasts by direct viral infection or transient transfection protocols by reducing apoptosis after initial transgene induction^{16,23–27}. We tested whether and how *p53* inhibition would influence the reprogramming of secondary transgenic NGFP1 iPS-cell-derived pre-B cells that have a high single-cell cloning efficiency and stably grow in the presence of doxycycline without requiring additional immortalization and with only background apoptosis levels (Supplementary Fig. 6). NGFP1 iPS cells were infected with a constitutively expressed lentiviral vector encoding a short interfering RNA (siRNA) hairpin for *p53* (Supplementary Fig. 7)²⁸. Infected cells were injected into host blastocysts and NGFP1-*p53*^{KD} B cells were single-cell sorted and cultured in doxycycline. *p53* inhibition did not alter transgene expression levels or affect the already residual levels of

apoptosis (Fig. 3a and Supplementary Figs 6 and 8), but rather shortened the cell-population-averaged doubling time of *p53*^{KD} cells in the presence of doxycycline by ~2-fold as compared to control NGFP1-derived cells (Fig. 3b and Supplementary Fig. 9). The kinetics of iPS cell formation proceeded with a significantly accelerated rate, with 93% of the wells producing Nanog-GFP⁺ cells within 8 weeks of doxycycline, as compared to 17 weeks for the control cells (Fig. 3c). The iPS cell lines were doxycycline independent, expressed pluripotency markers and generated teratomas and mouse chimaeras (Supplementary Fig. 10).

To assess whether the enhanced reprogramming could be attributed to the effect of *p53* inhibition on proliferation rate, we estimated, based on the population-averaged doubling times measured throughout the process, how many cell divisions have occurred for each NGFP1 and NGFP1-*p53*^{KD} clonal population during the latency period. Upon rescaling of latency by the doubling time (Fig. 3c, d), the cumulative fraction of wells generating iPS cells collapsed to the same statistically significant distribution with cell division number for both the NGFP1 and NGFP1-*p53*^{KD} wells (log-rank test for dissimilarity, $P = 0.518$). A similar analysis on NGFP1 and NGFP1-*p53*^{KD} CD11b⁺ cell-derived clonal populations showed a comparable distribution (log-rank test for dissimilarity, $P = 0.209$; Supplementary Fig. 11). Latencies were not normally distributed about the mean latency, but rather were better fit with a gamma distribution (Supplementary Figs 12 and 13). Knock-down of the *p21* gene, a downstream effector of *p53* which regulates cell-cycle progression²⁹, recapitulated the change in cell division rate and acceleration of reprogramming dynamics upon *Oct4*, *Sox2*, *Klf4* and *c-Myc* expression (Fig. 3a–d and Supplementary Figs 6–9 and 14), further substantiating that change in cell division rate directly rescales the kinetics of reprogramming to distributions similar to *Oct4*, *Sox2*, *Klf4* and *c-Myc* alone. Finally, we studied the effect of Lin28 overexpression on the reprogramming dynamics, as it enhances the reprogramming of human fibroblasts and acts as an oncogene by modulating the expression of cell-cycle regulators^{5,30}. NGFP1-Lin28^{OE} (Lin28 overexpressor)-derived B-cell populations demonstrated accelerated reprogramming kinetics that directly correlated with the observed increase in cell division rate on doxycycline ($P = 0.327$; Fig. 3a–d). In summary, our data demonstrate that both *p53/p21*

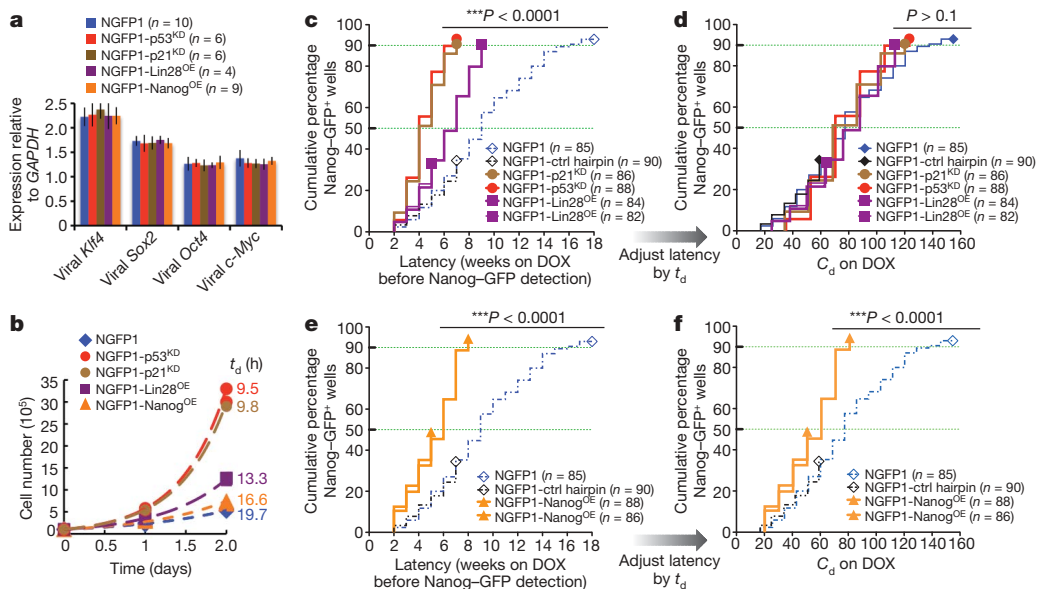


Figure 3 | Cell-division-rate-dependent and -independent acceleration of reprogramming. **a**, Average induction levels for transgenes in different NGFP1 cell populations. n indicates number of populations sampled per group, presented as mean \pm s.d. **b**, Growth curves for cells on doxycycline. Exponential growth (dashed line) described the data well ($R^2 = 0.97$ – 1.0), and the population-averaged doubling times (t_d) were calculated from these fits (Supplementary Fig. 9). **c**, As in Fig. 2b, latencies for reprogramming various clonal B-cell-derived populations. NGFP1-*p53*^{KD}, NGFP1-*p21*^{KD}

and NGFP1-Lin28^{OE} wells were statistically distinct from the NGFP1 and NGFP1-control hairpin wells ($P < 0.0001$, log-rank test for dissimilarity). **d**, Rescaling time by t_d provides an estimate for the number of cell divisions occurring during latency. C_d , population-averaged number of cell divisions on doxycycline before Nanog-GFP detection. No statistical difference between groups was observed after rescaling time by t_d ($P > 0.1$). **e**, **f**, As in **c**, **d**, but for NGFP1-Nanog^{OE} wells. n indicates number of populations monitored.

inhibition or Lin28 overexpression accelerate the reprogramming process rather than enhancing overall efficiency, as the cells divide more rapidly, resulting in an increased cumulative probability for the stochastic events to occur earlier in time (Fig. 3a–d).

Cell-division-rate-independent acceleration

We investigated whether reprogramming could be accelerated by mechanisms that are independent of cell proliferation rate. Nanog is a pluripotency factor expressed in the inner cell mass, and deriving embryonic stem cells and iPS cells requires the presence of functional endogenous *Nanog* alleles³¹. Moreover, Nanog enhances the transfer of pluripotency by cell fusion and facilitates direct reprogramming of human cells^{5,32–35}. Therefore, we aimed to investigate the effect of additional ectopic Nanog expression on the reprogramming kinetics by Oct4, Sox2, Klf4 and *c-Myc* proteins. B cells were derived from an NGFP1-Nanog^{OE} (Nanog overexpressor) iPS cell line carrying a doxycycline-inducible Nanog-encoding transgene (Supplementary Fig. 7). The presence of the transgene did not alter the expression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (Fig. 3a). Overall efficiency of iPS cell formation was similar to that of the control cells but reprogramming occurred with a significantly accelerated kinetics, with 94% of the wells producing pluripotent Nanog-GFP⁺ cells within 8 weeks of doxycycline (Fig. 3e and Supplementary Figs 15 and 16). Nanog overexpression slightly increased the cell-population-averaged doubling time of doxycycline-induced NGFP1-Nanog^{OE} cells compared to control NGFP1-derived reprogramming cells (Fig. 3b)^{36,37}. However, even after rescaling time on doxycycline by the change in doubling time (Fig. 3f), the cumulative fraction of wells generating iPS cells did not collapse to the same distribution with cell division number (log-rank test for dissimilarity, $P < 0.0001$). The median number of cell division divisions before wells produced iPS cells was significantly reduced from about 70

cell divisions in the NGFP1, NGFP1-p21^{KD} and NGFP1-p53^{KD} wells to 50 cell divisions in the NGFP1-Nanog^{OE} wells (Fig. 3d, f and Supplementary Fig. 12d). These results suggest that Nanog overexpression accelerates the reprogramming kinetics by cell-intrinsic mechanisms that are independent of an altered cell proliferation rate.

Numerical modelling of reprogramming

We used our data sets to conduct simulations of the stochastic reprogramming process occurring in individual cells. To assess the intrinsic rate of reprogramming per cell, we tested whether a simple stochastic model would describe our observations. The model considers the reprogramming of B cells to occur as a one-step process with a constant cell-intrinsic rate k (Fig. 4a). Because iPS cells were adherent and selectively retained throughout the culturing procedures, the latency measured for each well was representative of the timing at which the first B cell in the population reprogrammed plus a time delay, t_p , during which the daughter cells needed to grow to reach the detection threshold. At any given time t , the number of cells in each well, $N(t)$, scales the rate at which the first reprogramming event takes place, and the cumulative probability distribution of reprogramming times is $P(t + t_p) \approx 1 - e^{-k\tau}$, where τ is the ‘population-rescaled time’, $\tau = \int_0^t N(t') dt'$. Interpreting the experimental results in terms of this rescaled version of time allowed one to separate the contributions of population size and cell-intrinsic reprogramming rate (Fig. 4). Cell division rate can control the population size, N , and thus re-scales time, affecting the observed rate of reprogramming. For example, if the cell division rate and culturing procedures were modified to an effective population size of 1,000 cells in each well rather than $\sim 10^6$ cells per well (Fig. 4b, c), longer times would have been necessary to reach >90% reprogrammed wells.

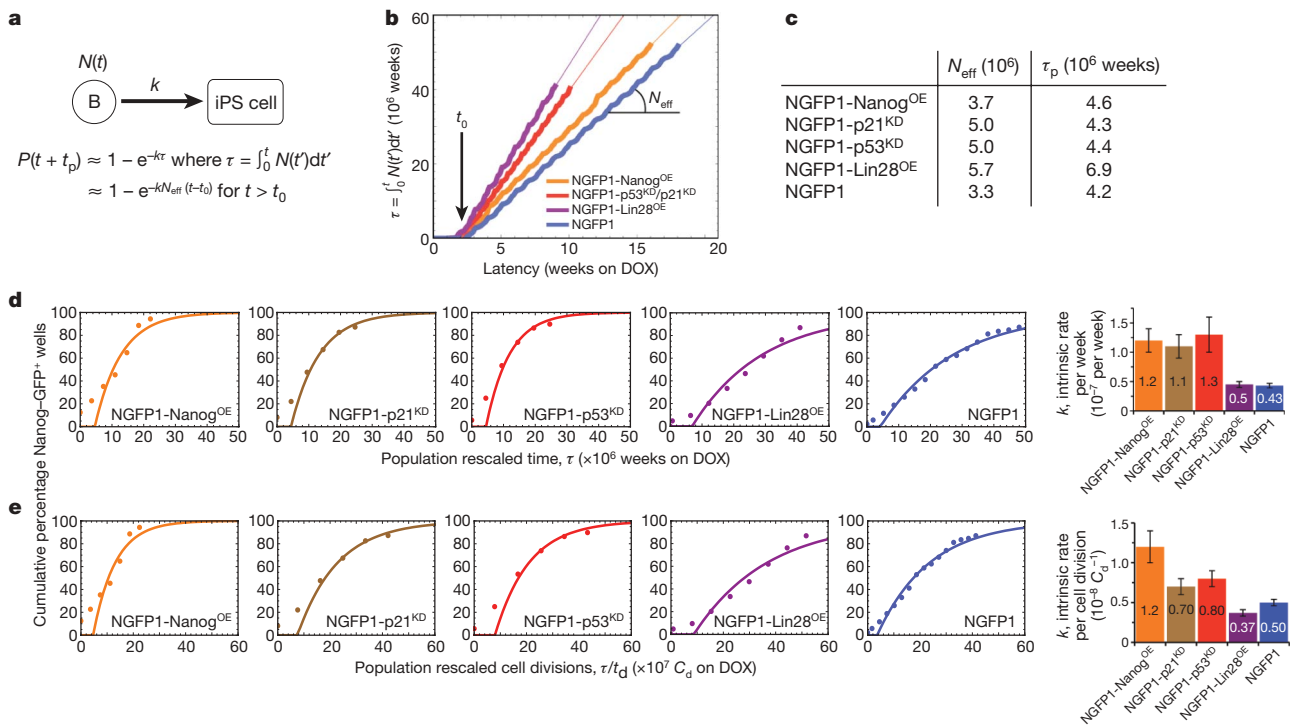


Figure 4 | Quantitative analysis of cell reprogramming. **a**, Stochastic model summary. Sequential re-plating of individual wells during each experiment establishes that, after a time t_0 representative of the time at which the re-platings started, each experiment can be described in terms of a population of an effective size, N_{eff} . **b**, Estimate of the population rescaled time, τ , throughout each experiment. After t_0 , population dynamics are effectively described by a fixed population of size, N_{eff} . NGFP1-p53^{KD} and NGFP1-p21^{KD} have similar dynamics. **c**, N_{eff} and the population-rescaled average proliferation times, τ_p , estimated as the population-rescaled time necessary

for one iPS cell to reach the detection threshold ($\tau_p = t_{d,i} N_{\text{eff}} \log_2(\rho N_{\text{eff}})$, where ρ is the detection threshold and $t_{d,i}$ is the doubling time of iPS cells). **d**, Cumulative percentage of Nanog-GFP⁺ wells as a function of τ , and best fits according to the proposed model. Modelling results are shown as thick lines; experiments are shown as dots. The far-right graph indicates best-fit estimates of the cell-intrinsic rate k expressed in terms of weeks. **e**, As in **d**, but per population-rescaled cell divisions, τ/t_d , instead of per τ units. t_d is the doubling time of the populations. Error bars indicate 95% confidence intervals.

Using cell population dynamics to account for the growth and culturing procedures used across the different NGFP1 lines (Fig. 4b, c), we calculated the population-rescaled time and fit the cell-intrinsic rate k using a likelihood maximization approach (Fig. 4d, e and Supplementary Fig. 17). Reasonable agreement was seen between best fits and the experiments. Furthermore, to take into account fluctuations in cell division times, number of cells and potential loss of iPS cells during culturing, in addition to the stochasticity in the cell-intrinsic reprogramming process, we implemented a detailed computer simulation of each experiment (Supplementary Fig. 18). Simulation results were consistent with those achieved using the analytical approach (Fig. 4d, e). The inferred intrinsic reprogramming rates per cell division for NGFP1-p53^{KD} and NGFP1-p21^{KD} cells were similar to NGFP1 cells (Fig. 4e and Supplementary Figs 17 and 18), suggesting that p53/p21 pathway inhibition in our system accelerates reprogramming through a predominately cell-division-rate-dependent mechanism. The slightly higher cell-intrinsic rates in some of our modelling results for p53/p21 inhibition (Fig. 4d, e and Supplementary Figs 17 and 18) may reflect a modest acceleration due to cell-division-rate-independent mechanisms and we cannot exclude that inhibition of the p53/p21 pathway slightly enhances reprogramming by additional mechanisms as well as in different experimental systems. The acceleration observed for NGFP1-Lin28^{OE} cells occurred predominantly because of a larger effective population size resulting from a faster cell division rate on doxycycline when following a culturing procedure similar to that used for NGFP1 cells (Fig. 4b, c). Only in the case of NGFP1-Nanog^{OE} cells, higher than 1.75–2 fold intrinsic reprogramming rates per cell division compared to NGFP1 cells was robustly obtained in all modelling approaches (Fig. 4e and Supplementary Figs 17 and 18), supporting the hypothesis that Nanog overexpression accelerates reprogramming through a predominately cell-division-rate-independent mechanism. Future characterization of how Nanog molecularly orchestrates the re-establishment of the core pluripotency circuitry is of great interest.

Discussion

We used a highly defined quantitative system for analysing the reprogramming process and suggest that yet-to-be-defined rate-limiting stochastic events occur as a function of cell division before the fully

reprogrammed pluripotent state is attained. Upon *Oct4*, *Sox2*, *Klf4* and *c-Myc* induction in our system, most, if not all, monocytes and lineage-committed B cells harbouring genetic rearrangements have the potential to generate iPS cells, albeit with very different latencies (model iii, Fig. 1). These results are inconsistent with models in which iPS cells preferentially arise from a particular epigenetic state in the donor cell population, such as a progenitor or an adult stem cell (models ii and iv, Fig. 1). The dynamics of direct *in vitro* reprogramming to pluripotency by *Oct4*, *Sox2*, *Klf4* and *c-Myc* are consistent with a continuous stochastic process, in which the conversion from a somatic cell to an iPS cell can be thought of as a drift in cell state. This cell state can be defined by a gene expression or epigenetic pattern, the fluctuations of which drive the conversion to an iPS cell. These fluctuations are probably due to cell-to-cell extrinsic heterogeneity or the inherent stochastic nature of gene expression or regulatory signalling processes^{38–40}. In the simplest scenario, the process could be modelled as a one-step rate-limiting transition characterized by a cell-intrinsic rate, and we found that this model describes reasonably the observed dynamics across all of the different cell lines. Our model relies on inferring a single cell-intrinsic rate averaged over the entire course of reprogramming, although the model does not optimally describe the reprogramming behaviour at early times after transgene induction (Fig. 4). This deviation between modelling and experimental results may indicate that the cell-intrinsic rate changes over the course of reprogramming within clonal populations. Intracolon monitoring of reprogramming dynamics at various points throughout the process could be used to add further mechanistic detail to the model, and to decipher the identity and sequence of rate-limiting step(s) and epigenetic changes that occur during reprogramming.

Quantitative analysis of the differences in reprogramming latency upon different genetic perturbations defined two distinct modes of accelerating the reprogramming process (Fig. 5 and Supplementary Fig. 19). In the ‘cell-division-rate-dependent’ mode, the cumulative probability for successful reprogramming is higher and can be achieved earlier in time and directly proportional to the augmentation in cell division rate, whereas in the ‘cell-division-rate-independent’ mode, reprogramming acceleration occurs over a lower average number of cell divisions. Notably, the different modes need not be mutually

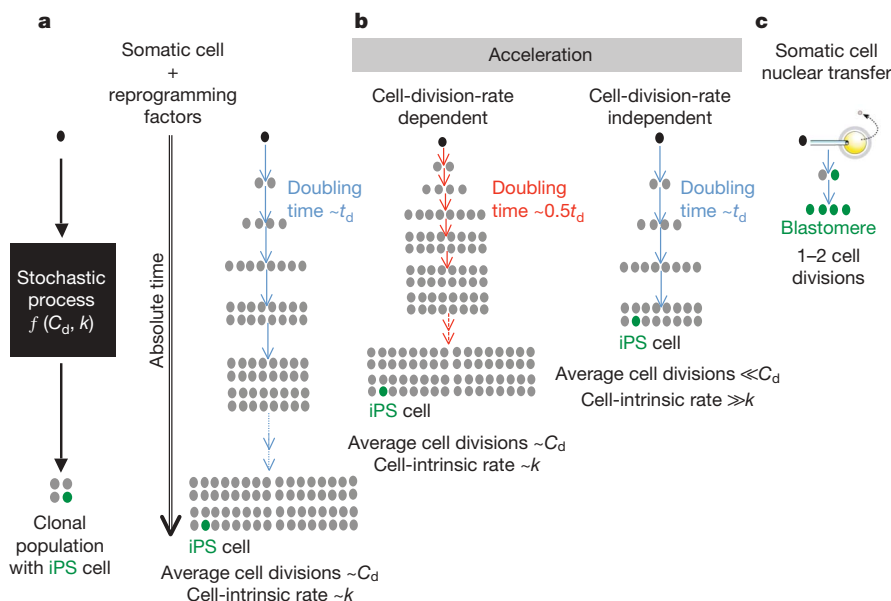


Figure 5 | Distinct modes for accelerating reprogramming to pluripotency. **a**, Nearly all donor cells can give rise to iPS cells via a stochastic process. Two parameters characterize the kinetics of the process: the average number of cell divisions required, C_d , and the cell-intrinsic reprogramming rate per cell division, k . **b**, Accelerating reprogramming in a cell-division-rate-dependent manner still requires as many divisions as the unaccelerated reference

scenario (that is, still C_d on average) but occurs earlier in time because cells divide faster, whereas in the cell-division-rate-independent mode, the cell-intrinsic rate reflecting the occurrence of an unknown stochastic event(s) is enhanced ($\gg k$) and reprogramming is achieved within a lower average number of divisions ($\ll C_d$). **c**, In comparison, somatic cell nuclear transfer can reprogram within 1–2 cell divisions.

exclusive as certain perturbations could enhance or inhibit reprogramming via both cell-proliferation-dependent and -independent effects. Recent studies failed to attain correlation between cell division rate and reprogramming efficiency; however, these studies relied on measuring proliferation in the absence of *Oct4*, *Sox2*, *Klf4* and *c-Myc* transgenes¹⁶ or by inducing proliferation with haematopoietic cytokines for which the receptors are rapidly silenced after transgene induction¹³. Our experimental and modelling data suggest that close monitoring of transgene induction, plating efficiency, cell proliferation and changes in population size throughout the experiments are needed to gain insight into the stochastic process of reprogramming. We consider two possibilities for the ability of increased cell division rate to accelerate and drive the kinetics of the reprogramming process: (1) cell division could amplify the number of daughter cells from partially reprogrammed cells where each resulting individual cell has an independent probability of progressing towards becoming an iPS cell; and (2) nuclear changes during cell division may facilitate the acquisition of epigenetic marks such as DNA and histone modifications that allow the re-establishment of the core transcriptional circuitry that stabilizes pluripotency^{6,15,41,42}.

iPS cells can be derived by various combinations of transcription factors and/or small molecules^{3,5,43,44}. Our study identifies the average number of cell divisions required to give rise to an iPS cell at a particular efficiency as a key parameter that may be affected by different reprogramming strategies. After nuclear transfer, the pluripotency gene *Oct4* (also called *Pou5f1*) of the somatic nucleus is reactivated in the cloned embryo within 1–2 cell divisions^{42,45,46} (Fig. 5c), indicating that the egg cytoplasm carries as-of-yet undefined determinants that accomplish robust reprogramming within very few cell divisions. These observations raise questions of whether direct reprogramming strategies devised so far can be robustly enhanced (for example, by optimizing stoichiometry, small molecule treatment or utilizing oocyte-derived transcript libraries) to enable complete *in vitro* reprogramming within only a few cell divisions.

METHODS SUMMARY

The NGFP1 iPS cell line was derived after culturing the infected mouse embryonic fibroblasts (MEFs) carrying ROSA26-M2rtTA mice and Nanog-GFP alleles with doxycycline-inducible lentiviruses encoding *Oct4*, *Klf4*, *c-Myc* and *Sox2* cDNA driven by the TetO/CMV promoter⁹. To generate the different genetically perturbed NGFP1 subclones, 50,000 NGFP1 cells were infected with the pSicoR vector²⁸ encoding a hairpin for the *p53* or *p21* tumour suppressor genes or with doxycycline-inducible lentiviruses encoding *Nanog* or *Lin28* cDNAs. To generate mouse chimaeras iPS cells were injected into diploid blastocysts that were initially harvested 94–98 h after hormone injection and placed in a drop of DMEM with 15% FCS under mineral oil. Pre-BCR⁺ (Igu⁺IgK⁻IgL⁻) early-pre B cells were single-cell sorted from 3–5-week-old chimaeras and were plated in gelatinized and irradiated OP9-coated wells with ES media supplemented with doxycycline (4 µg ml⁻¹). IL-7 and SCF (10 ng ml⁻¹ each, Peprotech) were supplemented to the media for the first 1–2 weeks to optimize cloning efficiency in the presence of doxycycline. After 2 weeks of doxycycline induction, cultures were trypsinized every week and passaged on gelatinized plates in ES cell medium plus doxycycline. Populations were defined positive for iPS cell generation when >0.5% of the 10,000 adherent/semi-adherent screened cells by FACS were GFP⁺. NGFP1-p53^{KD} and p21^{KD} derived cells were passaged twice a week (versus once a week at the time of GFP assays for the other groups) to avoid over-confluence in the culture wells due to their accelerated proliferation rate. The reprogramming experiments were stopped at different time points and all non-reprogrammed populations were viable and demonstrated robust growth on doxycycline upon ending each experiment. For teratoma generation, 2 × 10⁶ iPS cells were injected subcutaneously into both flanks of recipient SCID mice, and tumours were harvested for sectioning 3–6 weeks after initial injection.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 June; accepted 21 October 2009.

Published online 8 November 2009.

- Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920–1923 (2007).

- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
- Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
- Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
- Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49–52 (2009).
- Hanna, J., Carey, B. W. & Jaenisch, R. Reprogramming of somatic cell identity. *Cold Spring Harb. Symp. Quant. Biol.* doi:10.1101/sqb.2008.73.025 (2008).
- Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
- Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnol.* **26**, 916–924 (2008).
- Markoulaki, S. *et al.* Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nature Biotechnol.* **27**, 169–171 (2009).
- Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513–524 (2009).
- Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41**, 968–976 (2009).
- Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364–377 (2009).
- Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
- Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460**, 1145–1148 (2009).
- Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnol.* **26**, 795–797 (2008).
- Signer, R. A., Montecino-Rodriguez, E., Witte, O. N. & Dorshkind, K. Aging and cancer resistance in lymphoid progenitors are linked processes conferred by p16^{Ink4a} and Arf. *Genes Dev.* **22**, 3115–3120 (2008).
- Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).
- Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
- Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151–159 (2008).
- Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230–240 (2008).
- Banito, A. *et al.* Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.* **23**, 2134–2139 (2009).
- Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460**, 1132–1135 (2009).
- Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**, 1140–1144 (2009).
- Li, H. *et al.* The *Ink4/Arf* locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).
- Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).
- Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101**, 10380–10385 (2004).
- Rowland, B. D., Bernards, R. & Peepker, D. S. The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nature Cell Biol.* **7**, 1074–1082 (2005).
- Xu, B., Zhang, K. & Huang, Y. Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA* **15**, 357–361 (2009).
- Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009).
- Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Navarro, P. *et al.* Molecular coupling of Xist regulation and pluripotency. *Science* **321**, 1693–1695 (2008).
- Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997–1001 (2006).
- Darr, H., Maysar, Y. & Benvenisty, N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development* **133**, 1193–1201 (2006).
- Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118–122 (2009).
- Huang, S. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31**, 546–560 (2009).
- Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).

40. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
41. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
42. Egli, D., Birkhoff, G. & Eggan, K. Mediators of reprogramming: transcription factors and transitions through mitosis. *Nature Rev. Mol. Cell Biol.* **9**, 505–516 (2008).
43. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**, 411–419 (2009).
44. Shi, Y. *et al.* Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3**, 568–574 (2008).
45. Boiani, M. *et al.* Variable reprogramming of the pluripotent stem cell marker Oct4 in mouse clones: distinct developmental potentials in different culture environments. *Stem Cells* **23**, 1089–1104 (2005).
46. Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447**, 679–685 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank N. Maheshri, K. Wood, Y. Bugnanim, B. Carey, J. Cassidy and F. Soldner for commenting on the manuscript;

J. Dausmann, S. Markoulaki, P. Schorderet and M. Pawlak for conducting blastocyst injections; and members of the Jaenisch laboratory for discussions. R.J. is supported by grants from the NIH (RO1-HDO45022, R37-CA084198, RO1-CA087869). J.H. is supported by a postdoctoral fellowship from Helen Hay Whitney Foundation. K.S. is supported by the Society in Science: The Brano-Weiss fellowship. J.v.Z. is supported by the Human Frontiers Science Program. A.v.O. is supported by a National Institutes of Health (NIH) Director's Pioneer award and an award from the National Cancer Institute (U54CA143874).

Author Contributions J.H., K.S. and R.J. conceived the ideas for this study and designed the experiments. J.H. conducted the reprogramming experiments and data collection. M.P.C. and C.J.L. assisted in construct preparation, transgene induction and blastocyst injections. K.S., B.P., J.v.Z., J.H., and A.v.O. performed statistical and numerical analyses, and also constructed the stochastic models. J.H., K.S. and R.J. wrote the manuscript with contributions from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J. (Jaenisch@wi.mit.edu) or J.H. (Hanna@wi.mit.edu).

METHODS

Pluripotent lines and viral infections. Mouse embryonic fibroblasts (MEFs) used to derive primary induced pluripotent stem (iPS) cell lines by infections with inducible lentiviruses were harvested at 13.5 days post coitum from F₁ mating between ROSA26-M2rtTA and Nanog-GFP mice⁹. Lentiviral preparation and infection with doxycycline-inducible lentiviruses encoding *Oct4*, *Klf4*, *c-Myc* and *Sox2* cDNAs driven by the TetO/CMV promoter were previously described²¹. The NGFP1 iPS cell line used in this study was derived after culturing the infected MEFs with doxycycline, and grew stably in culture independent of doxycycline. To generate NGFP1-p53^{KD} and p21^{KD} clonal cell lines, 50,000 NGFP1 cells were infected with the pSicoR-PGK-puro vector encoding a hairpin for mouse *p21* (targeting sequence: GCAGATTGGTCTTCTGCAA) and a previously described specific hairpin for the mouse *p53* tumour suppressor gene²⁸. Hairpin against the CD8 sequence was used as control where indicated. A total of 10 µg of lentiviral vector and packaging vectors was co-transfected in 293T cells by using the FuGENE 6 reagent (Roche Diagnostics). Supernatants were collected 36–48 h after transfection, filtered through a 0.45-µm filter, and used to infect the iPS cells for 48 h. Afterwards, the cells were trypsinized, plated at low densities and subcloned to test for knockdown specificity (NGFP1 already contains a puromycin selection cassette in the ROSA26 locus restricting our ability to use puromycin as a marker for infected cells). To verify integration of PsicoR lentiviral vectors, the following oligonucleotides were used: forward, CCCGGTTAATTTGCATATAATATTTTC; reverse, CATGATACAAAGGCATTAAAGCAG. cDNAs encoding *Nanog* and *Lin28* transcripts were cloned into the EcoRI site of TetO-FUW lentiviral vectors and were used to infect NGFP1 iPS cells. Integration for TetO-Nanog and TetO-Lin28 lentiviral vectors was verified by PCR and Southern blot analysis on genomic DNA using the EcoRI-digested insert as a probe (data not shown). A NGFP1 subclone that demonstrated >95% constitutive knockdown of *p53* (*Trp53*) or *p21* gene products, termed NGFP1-p53^{KD} or NGFP1-p21^{KD}, was used in the subsequent experiments. iPS cells were cultured in DME containing 15% fetal calf serum (FCS), leukaemia inhibiting factor (LIF), penicillin/streptomycin, L-glutamine, β-mercaptoethanol and non-essential amino acids. Chromosomal karyotyping of iPS cell lines was performed by Cell Line Genetics on 20 G-banded metaphase cells from each line tested.

Chimaera and teratoma formation. iPS cells were injected in C57B6 × 129Sv F₁ *Rag2*^{-/-} or BDF2 diploid blastocysts that were initially harvested 94–98 h after hormone injection and placed in a drop of DMEM with 15% FCS under mineral oil. A flat-tip microinjection pipette with an internal diameter of 1.2–1.5 mm was used for iPS cell injection (using a Piezo micromanipulator). A controlled number of cells was injected into the blastocyst cavity. After injection, blastocysts were returned to KSOM media (Invitrogen) and placed at 37 °C until transferred to recipient females. Ten to fifteen injected blastocysts were transferred to each uterine horn of 2.5 days post coitum pseudo-pregnant females. To recover full-term pups, recipient mothers were killed at 19.5 days post coitum. For teratoma generation, 2 × 10⁶ iPS cells were injected subcutaneously into both flanks of recipient SCID mice, and tumours were harvested for sectioning 3–6 weeks after initial injection.

Reprogramming into iPS cells. Transgenic pre-BCR⁺ (Igu⁺IgK⁻IgL⁻) early pre-B cells, which are not generated in *Rag2*^{-/-} mice owing to their inability to undergo heavy- and light-chain rearrangements, were isolated from bone marrow of 3–5-week-old *Rag2*^{-/-} NGFP1 chimaeras and single-cell sorted into 96-well plates. This choice of host ensured that any isolated B cells were derived from the injected iPS cells, and not from the host blastocysts. In certain experiments, we labelled the NGFP1-iPS cell clone with a lentivirus constitutively expressing dTomato (a gift from K. Hochedlinger) and injected cells into BDF2 blastocysts to produce chimaeric mice, and used the dTomato as a marker of transgenic cells. One of the experimental replicates was performed on B220⁺CD25⁺ pro-B cells isolated from reprogrammable mice carrying the same set of transgenes¹¹. Cells were plated in gelatinized and irradiated (3,000 rad) OP9-coated wells in ES medium containing doxycycline (4 µg ml⁻¹). IL-7 and SCF (10 ng ml⁻¹ each, Peprotech) were added to the medium for the first 1–2 weeks to optimize cloning efficiency in the presence of doxycycline (the plating efficiency of pre-B cells to grow as single cells was >80%). Cells on doxycycline did not exhibit contact inhibition and grew both on gelatin and in suspension. After 2 weeks of doxycycline induction, cultures were trypsinized every week and passaged on gelatinized plates in ES medium plus doxycycline. Populations were defined as being positive for iPS cell generation when >0.5% of the adherent/semi-adherent screened cells were GFP⁺. Upon detection of a GFP⁺ fraction, further flow cytometric (FACS) assaying of the clonal population was discontinued and, in all cases, stable doxycycline-independent iPS cell lines were derived by growing the cells in the absence of doxycycline (which allows the iPS cells to overgrow transgene-dependent, partially reprogrammed GFP⁻ cells in the population). A fraction of ~250,000 cells, including one-quarter of all fully adherent cells, was replated at the end of each

week (beginning at week 2) for further follow-up analysis. By including adherent cells in the culture plate upon passaging, we consistently retained all iPS cells generated during a given culture period, as no iPS cells were detected by FACS in the non-adherent fraction (Supplementary Fig. 1). NGFP1-p53^{KD} and NGFP1-p21^{KD} derived cells were passaged and replated twice a week (versus once a week at the time of GFP assays for other cells) to avoid over-confluence in the culture wells due to their higher proliferation rate. Population-averaged doubling time for reprogramming and iPS cell populations was determined by plating 10⁵ cells, and counting total cell number after 24 and 48 h in duplicates. Numbers of cells as a function of time were fit using exponential growth regression in Excel (*R*² ranging from 0.97 to 1.00). The CD11b⁺ myeloid fraction was sorted from spleen and cells were plated in gelatinized and irradiated OP9-coated 96-well plates with ES medium supplemented with doxycycline (4 µg ml⁻¹). M-CSF, Flt3L, LPS and SCF (10 ng ml⁻¹, Peprotech) was added to the medium for the first 2 weeks to boost plating efficiency. Wild-type CD11b⁺ cells were obtained from 'reprogrammable' transgenic mice carrying identical doxycycline-inducible copies of the reprogramming factors *Oct4*, *Sox2*, *Klf4* and *c-Myc*, Rosa26-M2rtTA and a Nanog-GFP knock-in reporter. Unlike iPS cell chimaeras, all cells in these reprogrammable mice carried the same set of transgenes¹¹. However, unlike for pre-B cells, which are not found in *Rag2*^{-/-} hosts and thus all isolated cells were of transgenic origin, CD11b⁺ cells were isolated from reprogrammable mice carrying the same set of transgenes¹¹ or from dTomato-labelled chimaeric donor mice. V(D)J-IgH rearrangements were amplified from genomic DNA samples by PCR using degenerate primer sets as described⁹.

Immunofluorescence. Cells were fixed in 4% paraformaldehyde for 20 min at 25 °C, washed three times with PBS and blocked for 15 min with 5% FBS in PBS containing 0.1% Triton X-100. After incubation with primary antibodies against Oct4 (Santa Cruz), Nanog (polyclonal rabbit, Bethyl) and SSEA1 (monoclonal mouse, Developmental Studies Hybridoma Bank) for 1 h in 1% FBS in PBS containing 0.1% Triton X-100, cells were washed three times with PBS and incubated with fluorophore-labelled appropriate secondary antibodies purchased from Jackson ImmunoResearch. Specimens were analysed on an Olympus fluorescence microscope, and images were acquired with a Zeiss Axiocam camera.

Quantitative RT-PCR. Total RNA was isolated using the RNeasy Kit (Qiagen). Three micrograms of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). One microgram of DNase-I-treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately re-suspended in 100 µl of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction in an ABI Prism 7000 (Applied Biosystems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen). Error bars indicate standard deviation of triplicate measurements for each clone. Primers used for transgene-encoded amplification were as follows: *c-Myc*, forward 5'-ACCTAACTCGAGGAGGAGTGG-3', reverse 5'-TCCACATAGCGTAAAAGGAGC-3'; *Klf4*, forward 5'-ACACTGTCTTCCCACGAGGG-3', reverse 5'-GGCATTAAAGCAGCGTATCCA-3'; *Sox2*, forward 5'-CATTAACGGCAGCACTGCC-3', reverse 5'-GGCATTAAAGCAGCGTATCCA-3'; *Oct4*, forward 5'-AGCCTGGCCTGTCTGTCACTC-3', reverse 5'-GGCATTAAAGCAGCGTATCCA-3'; *Nanog*, forward 5'-ACATGCAACCTGAAGACGTG-3', reverse 5'-CACATAGCGTAAAAGGAGCAA-3'. To ensure equal loading of cDNA into RT reactions, *GAPDH* mRNA was amplified using the following primers: forward 5'-TTCACCACATGGAGAAGGC-3', reverse 5'-CCCTTTTGGCTCCACCCT-3'. Data were extracted from the linear range of amplification.

Antibodies and apoptosis measurement assays. Fluorescently conjugated antibodies (phycoerythrin, FITC, Cy-Chrome or APC-labelled) were used for flow cytometric analysis, and for cell sorting: anti-CD11b⁺, pre-BCR, IgK, IgL, CD19, B220, CD45.2 and IL-7R antibodies (BD-Biosciences) were used. Enrichment for CD11b^{high} cells by using a CD11b magnetic bead isolation kit was carried out before sorting. Cell sorting was performed by using FACS-Aria (BD-Biosciences), and consistently achieved cell sorting purity of >99%. Antibodies for western blot analysis: anti-mouse p53 (BAF1355, R&D systems), anti-β actin (ab8226, Abcam) and anti-p21^{cip} (clone C-19, Santa Cruz Biotechnology). To determine the rate of apoptosis, samples were stained with the annexin V staining kit (BD Biosciences) and propidium iodide according to the manufacturer's recommendations and analysed by flow cytometry. In addition, samples were subjected to the Tdt-mediated dUTP-biotin nick end labelling (TUNEL) test for apoptotic cells by using the flow-cytometry-based MEBSTAIN apoptosis kit (MBL).

Non-parametric statistical analysis. The latency of iPS cell appearance of each well was used to generate a survival curve for each experiment. Latency was considered either as a function of absolute time (that is, the weeks on doxycycline until GFP was detected) or as a function of the population-averaged number of cell divisions during latency (that is, latency in time divided by doubling time). Any wells that failed to generate Nanog-GFP⁺ cells at the end of each experiment were considered to be censored, and are shown with solid geometric shapes in Figs 2 and 3. Histograms of the fraction of cells at initial GFP detection were performed using

Prism 5 (version 5.0b; Graphpad Software). Using the censored latency data sets, the log-rank (Mantel–Cox) test, a non-parametric statistical test appropriate for right-censored data, was used to test the null hypothesis that survival functions do not differ across groups. In contrast to parametric analyses, this test does not require knowledge about the shape of the survival curve or the distribution of survival times. Analysis was performed using Prism 5 (version 5.0b; Graphpad Software) and two-tailed *P*-values. *P*-values above 0.05 indicate that the latencies between the two groups were similar (that is, the treatment did not change survival; accept the null hypothesis at a 95% confidence level), whereas *P*-values less than 0.05 indicate that latencies between groups were not similar (that is, survival functions differ; reject the null hypothesis at a 95% confidence level).

Parametric statistical analysis. The censored latency data sets were also fit to several univariate probability distributions using maximum likelihood estimation

via the 'dfittool' in Matlab (The Mathworks). For optimal univariate distribution fits, the chi-squared goodness-of-fit statistic, χ^2 , was used to assess the quality of each fit. Gamma distributions had the lowest χ^2 of any fit distribution (see Supplementary Fig. 12).

Computational simulation. Simulations of the proposed model were generated using a hybrid scheme in which the size of each subpopulation (either B-cells or iPS cells) was evolved by considering (1) stochastic Gillespie-like evolution of Poissonian growth dynamics for small population sizes; (2) deterministic evolution of population sizes using a time step of $\Delta t = 0.001$ weeks whenever the probability of generating a new cell in the time Δt exceeded 0.1; (3) Gillespie-like evolution of the reprogramming transition; (4) periodic re-plating, selection of iPS cells and observation of the fraction of reprogrammed cells according to the details of the experimental protocol followed in each experiment.