SYNTHETIC BIOLOGY

Synthetic genetic circuits to uncover the OCT4 trajectories of successful reprogramming of human fibroblasts

Katherine Ilia[1,2,3]†, Nika Shakiba[1,3,4]†, Trevor Bingham[5,6]†, Ross D. Jones[1,3,4], Michael M. Kaminski[1,3,7,8,9], Eliezer Aravera[1,3,10], Simone Bruno[11], Sebastian Palacios[1,2,3,11,12], Ron Weiss[1,3,12], James J. Collins[1,2,3,13,14], Domitilla Del Vecchio[3,11]*, Thorsten M. Schlaeger[5]*

Reprogramming human fibroblasts to induced pluripotent stem cells (iPSCs) is inefficient, with heterogeneity among transcription factor (TF) trajectories driving divergent cell states. Nevertheless, the impact of TF dynamics on reprogramming efficiency remains uncharted. We develop a system that accurately reports OCT4 protein levels in live cells and use it to reveal the trajectories of OCT4 in successful reprogramming. Our system comprises a synthetic genetic circuit that leverages noise to generate a wide range of OCT4 trajectories and a microRNA targeting endogenous OCT4 to set total cellular OCT4 protein levels. By fusing OCT4 to a fluorescent protein, we are able to track OCT4 trajectories with clonal resolution via live-cell imaging. We discover that a supraphysiological, stable OCT4 level is required, but not sufficient, for efficient iPSC colony formation. Our synthetic genetic circuit design and high-throughput live-imaging pipeline are generalizable for investigating TF dynamics for other cell fate programming applications.

INTRODUCTION

In their landmark papers, Takahashi and Yamanaka demonstrated that murine (*1*) and human (*2*) somatic cells can be reprogrammed to a pluripotent state via simultaneous overexpression of octamer-binding transcription factor 4 (OCT4), SRY (sex determining region Y)-box 2 (SOX2), Krüppel-like factor 4 (KLF4), and cellular myelocytomatosis (c-MYC), collectively referred to as OSKM. Given the immense potential that induced pluripotent stem cells (iPSCs) hold for regenerative medicine and disease modeling, substantial efforts have been directed toward developing robust cell fate reprogramming protocols for the efficient and reproducible generation of scalable quantities of high-quality iPSCs (*3*–*5*). To date, reprogramming of human somatic cells to iPSCs via the ectopic expression of OSKM is still marked by low efficiency (*6*, *7*) and poor-quality iPSCs (*8*–*15*), suggesting a knowledge gap regarding how controllable parameters can be optimized to improve the robustness of reprogramming protocols.

The low efficiency of somatic reprogramming to pluripotency has motivated work to identify both the stoichiometric and the temporal trajectory requirements of key molecular players for successful reprogramming. OCT4, a critical component of the transcriptional circuitry that governs pluripotency (*16*), plays a pivotal role during reprogramming and in the maintenance of pluripotency (*17*–*20*), given that its overexpression alone can generate iPSCs (*21*) and that complete reprogramming cannot be achieved in the total absence of endogenous and ectopic OCT4 (*22*, *23*). Consequently, research has focused on investigating the OCT4 levels and trajectories conducive to reprogramming. It has been demonstrated that fluctuations in OCT4 (*24*, *25*) and SOX2 (*26*) mediate the destabilization of the pluripotent state. Furthermore, the relative expression levels and the duration of expression of OCT4 and SOX2 reportedly influence the efficiency with which somatic cells reach the pluripotent state (*19*, *27*–*31*).

Several mechanistic studies have provided in-depth insight into the molecular changes that cells undergo during reprogramming (*28*, *32*–*40*). Time course reverse transcription quantitative polymerase chain reaction (RT-qPCR) data during reprogramming identified Oct4 levels above those found in embryonic stem cells (ESCs)/iPSCs (*17*, *20*, *38*). These studies, however, only quantified mRNA levels, which do not always correlate with protein levels during reprogramming (*1*, *29*). Furthermore, as the majority of transduced human dermal fibroblasts (HDFs) fail to reprogram, bulk measurements provide limited information about the rare cells that follow successful trajectories. To circumvent this difficulty, the authors in (*38*) used an efficient secondary mouse system in which reprogramming events are more frequent. This study also indicates that high Oct4 levels may be present during reprogramming. Nevertheless, this study did not measure protein level directly, nor could it reconstitute single-cell or single-colony trajectories. In contrast, another study showed that among TRA-1-60$^+$ cells at day 7 of human fibroblast reprogramming, the OCT4 protein is expressed at

[1]Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Institute for Medical Engineering and Science, MIT, Cambridge, MA 02139, USA. [3]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]School of Biomedical Engineering, University of British Columbia, Vancouver, British Columbia V6T 1Z3 Canada. [5]Stem Cell Program, Boston Children's Hospital, Boston, MA 02115, USA. [6]Harvard University, Boston, MA 02115, USA. [7]Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz-Association, Berlin 10115, Germany. [8]Department of Nephrology and Medical Intensive Care, Charité – Universitätsmedizin Berlin, Medizinische Klinik m.S. Nephrologie und Intensivmedizin, Berlin 10117, Germany. [9]Berlin Institute of Health, Berlin 13125, Germany. [10]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY 11794, USA. [11]Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA. [12]Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA. [13]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02215, USA. [14]Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA.
*Corresponding author. Email: ddv@mit.edu (D.D.V.); thorsten.schlaeger@childrens.harvard.edu (T.M.S.)
†These authors contributed equally to this work.

levels close to those found in PSCs (*39*). Nevertheless, as above, this study only provides bulk information for a population in which the cells with pluripotent colony forming potential are vastly outnumbered by those that ultimately fail to reprogram (*11*, *39*).

A few studies have used single-cell mass cytometry to identify relationships between protein-level expression and reprogramming outcomes in mouse cells (*37*, *40*). In particular, Zunder *et al.* (*40*) showed that successful trajectories go through an early mixed-stoichiometry phase and then reach a high Oct4 level. While this approach generates protein expression data at single-cell resolution, it still does not provide information on the specific temporal trajectories that result in reprogramming to pluripotency as the cells are sacrificed at every time measurement. Using yet another approach, the authors in (*19*, *30*) added fluorescent proteins to the reprogramming factor expression cassettes, allowing them to probe the effects of different reprogramming factor levels and stoichiometries. However, these studies were agnostic to the potential contribution of endogenous expression, which, for core pluripotency TFs such as OCT4, can commence as early as 5 to 8 days after transduction (*11*, *22*, *29*, *41*–*44*).

In summary, these prior studies have key limitations. In single-cell RNA sequencing studies (*32*–*36*), it is challenging to determine the total expression level of a gene that is expressed both ectopically and endogenously, and RNA levels are not necessarily reflective of protein levels (*1*, *29*, *29*, *45*). While the quantification of total protein levels is possible with single-cell proteomics (*37*, *40*), both proteomics and transcriptomics approaches can only be used to reconstruct population-level trajectories, as cells must be destroyed for measurements to be taken. Furthermore, because reprogramming events occur with low probability, population-level trajectories are not representative of the few cells that successfully reach pluripotency. In (*36*), the authors were able to leverage optimal-transport analysis to identify single-cell trajectories; this work, however, was done on murine fibroblasts, for which the reprogramming efficiency is higher than for human fibroblasts. Therefore, despite substantial prior work, the expression dynamics of OCT4 remain unknown for the rare subpopulation of cells that become successfully reprogrammed. A more ideal approach would require a means of conducting the live tracking of rare cell subpopulations in a high-throughput fashion using methods that facilitate the precise tracking of TF levels without interfering with their transcriptional regulation.

As synthetic biology technologies for controlling gene expression have come of age, there is an opportunity to apply an engineering lens to the problem of reprogramming. Here, we lay the foundation for such an approach by designing an OCT4 trajectory generator to reverse-engineer the dynamics of the total cellular OCT4 level conducive to reprogramming. To this end, we implemented a microRNA (miRNA)–based strategy that facilitates the setting of the total level of endogenously expressed TFs, such as OCT4 (Fig. 1A, left), inspired by the control theory technique of high-gain negative feedback (*46*, *47*). We delivered the trajectory generator to cells via lentiviral vectors such that it generates a wide range of temporal OCT4 dynamics; we tracked OCT4 levels with clonal resolution over time via high-resolution and high-throughput imaging (Fig. 1A, right). Our imaging data revealed that successfully reprogrammed cells follow trajectories characterized by stable, supraphysiological total OCT4 levels. The approach presented here is broadly applicable to probing the trajectory of TFs, laying the foundation for future work to improve the robustness of cell fate engineering protocols.

RESULTS

A strategy for setting total transcription factor levels

To probe the role of the dynamics of OCT4 on reprogramming success, we developed a strategy for overwriting endogenous TF levels (Fig. 1A, left). Our system consists of two key features: (i) the ectopic expression of the TF of interest and (ii) the suppression of the endogenous contribution to TF levels with a synthetic miRNA that selectively knocks down the endogenous mRNA species encoding the TF of interest. As this knockdown occurs post-transcriptionally in the cytoplasm, it circumvents disrupting transcriptional regulation of the OCT4 gene. To avoid knockdown of the ectopic OCT4 mRNA by the synthetic miRNA targeting the TF coding sequence, we randomized the codons of the ectopic mRNA encoding OCT4.

To implement feature (ii), we assayed the knockdown activity of a panel of published (*19*, *27*) and unpublished miRNAs against OCT4 (fig. S1). Our top candidate was the previously unpublished miR-Oct4e that, when expressed from the U6 promoter (*48*), knocked down OCT4 levels by about 24-fold in transfection experiments, reaching OCT4 levels that are less than 5% of those in cells that do not express the miRNA (fig. S1). To validate that this miRNA could be useful for knocking down physiologically relevant OCT4 expression, we transfected the U6-driven miR-Oct4e into iPSCs. The miRNA eliminated detectable OCT4 expression in over 80% of cells (Fig. 1B).

Recapitulating the role of OCT4 levels in the maintenance of pluripotency

We next evaluated whether we could leverage this system to experimentally demonstrate the known role that OCT4 levels play in the maintenance of pluripotency (*49*). To this end, we constructed a two-lentivirus system (Fig. 2A). One lentivirus encodes U6-driven miR-Oct4e and a strong inducible tetracycline transcriptional regulator (TetR)–responsive element (TRE) promoter (fig. S2), driving the expression of OCT4 linked to the fluorescent reporter mNeonGreen via a self-cleaving 2A peptide (*50*). In this system, doxycycline (DOX) activates expression of ectopic OCT4. miR-Oct4e is encoded on the same virus as OCT4, such that every transduced cell expresses both components. The second virus contains a constitutive promoter driving the expression of the reversed Tet transactivator (rtTA; rTetR-NLS-VP64) and blue fluorescent protein (TagBFP), which serves as a transduction marker. Both lentivirus constructs are flanked by enhancer-blocking insulators (*51*) to mitigate the extent to which the epigenetic context of lentiviral integration affects transgene expression.

To confirm our ability to perturb the stability of the pluripotent state using this system, we infected iPSCs with two variants of this two-virus system, one with and one without the U6-miR-Oct4e cassette, and tracked the ability of these cells to maintain pluripotency in a variety of DOX conditions. Whereas the miRNA is constitutively expressed, ectopic OCT4 expression requires the presence of DOX. To simulate the effects of OCT4 above normal pluripotent levels, we analyzed cells that do not express miR-Oct4e while overexpressing ectopic OCT4. In these samples, induction of OCT4 via DOX addition resulted in a general decrease of the fraction of cells

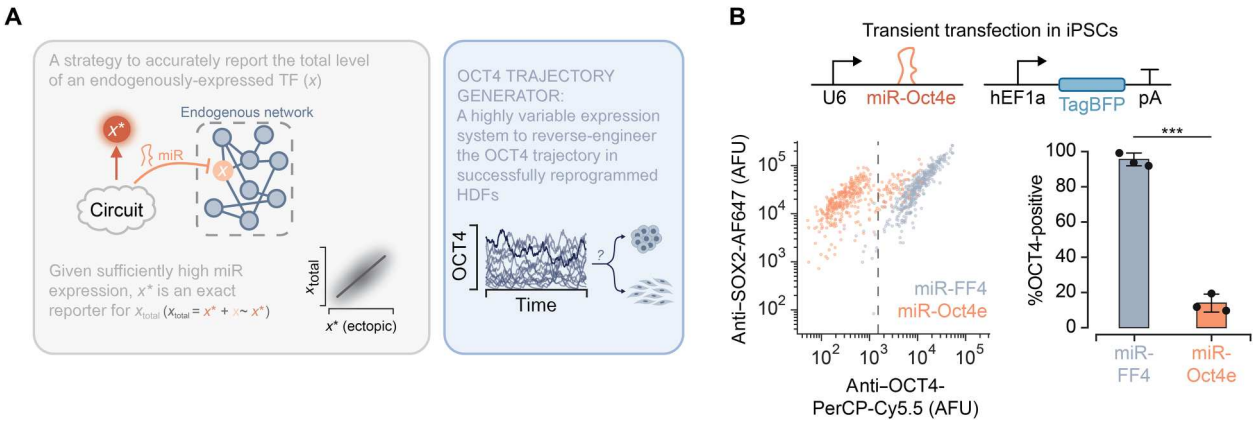

Fig. 1. A miRNA-based strategy for setting total TF levels. (**A**) Left: High-level schematic of a genetic circuit that expresses an ectopic TF (*x**) and a miRNA that knocks down the endogenous TF (*x*) expression. If the miRNA is sufficiently expressed, then the endogenous contribution *x* to TF levels is negligible, and the ectopic and total TF (x_{total}) levels are approximately equivalent. Right: High-level schematic describing the trajectory generator. (**B**) We transiently transfected iPSCs with the two illustrated plasmids and stained for OCT4 and SOX2 48 hours later. The scatter plot contains flow cytometry data for TagBFP-positive cells. The dashed line denotes the OCT4-positive threshold used to calculate the values in the bar graph. Error bars represent the SDs for $n = 3$ biological replicates, shown as dots. Asterisks indicate significant differences (***$P < 0.001$) according to an unpaired two-tailed Student's *t* test. AFU, arbitrary fluorescence units; pA, poly(A) tail.

staining positive for the pluripotent stem cell marker TRA-1-60 (*11*) (Fig. 1C, bottom left), with statistically significant changes observed in cases when we added DOX 2, 4, and 6 days after transduction (fig. S3, A and B). To simulate the effect of OCT4 below normal pluripotent levels, we considered cells after transduction with the system expressing both miR-Oct4e and ectopic OCT4. We found that, unless ectopic OCT4 is expressed (i.e., in conditions without DOX induction), a decreased fraction of iPSCs stain positive for TRA-1-60 (Fig. 1C, bottom right), with statistically significant differences between the samples maintained in the absence and presence of DOX observed 4, 6, and 8 days after transduction (fig. S3, C and D). This indicates that the miRNA effectively destabilizes the pluripotent state and that DOX induction can mitigate this effect. We then leveraged this experimental system to perform a preliminary investigation of the role of OCT4 dosage on the maintenance of pluripotency. We maintained the iPSCs transduced with the viruses in Fig. 2A in DOX-containing medium and found that no statistically significant dosage dependence trend emerged (Fig. 2, B and C). Nevertheless, all cells infected exhibited a substantial loss of TRA-1-60 signal 14 days after infection (Fig. 2B). In addition, the net reduction of the number of cells highly expressing ectopic OCT4 (bin 5 in Fig. 2B) suggests that very high expression may be detrimental to the iPSC state; we also observed toxicity associated with transgene expression (Fig. 2D). Together, these results suggest that, with our system, we may be able to disturb the pluripotency gene regulatory network by both increasing and decreasing total OCT4 levels relative to baseline endogenous expression levels and that these perturbations can be of sufficient magnitude to trigger the loss of pluripotency.

A trajectory generator to produce wide OCT4 distributions

These results prompted us to investigate in greater detail how temporal OCT4 dynamics shape the reprogramming process. In particular, we developed a trajectory generator (Fig. 3A) that achieves wide OCT4 distributions and facilitates dynamic tracking of total OCT4 levels over time. As the reprogramming efficiency of HDFs is low, we found that leveraging lineage tracking approaches based on barcoding was not feasible for this application (fig. S4). We hypothesized that a live imaging-based approach, which does not require sacrificing cells for measurements and allows us to track the levels of a fluorescent reporter in cells over time, would likely not face the same limitations. This led us to fuse the OCT4 and mNeonGreen proteins to obtain accurate temporal tracking of ectopic OCT4 levels (*52*). In the absence of endogenous OCT4 production, mNeonGreen is an exact reporter for total OCT4 levels. In addition, as OCT4 is a TF, the OCT4-mNeonGreen protein is localized in the nucleus; this simplifies high-throughput image processing and cell tracking, as it allows for the spindle-shaped or stellate fibroblasts to be identified by their round nuclei. Previous studies have demonstrated that such a fusion protein does not negatively affect the reprogramming process (*53*).

We reasoned that, for this type of approach to be viable, we needed a genetic circuit that promotes gene expression variability. As "burst-like" transcriptional activation is known to lead to wide variance in expression levels (*54*–*57*), we reasoned that a switch-like inducible gene expression system, such as with rtTA and DOX, set to intermediate activator expression levels could generate a wide variety of OCT4 trajectories. Via deterministic modeling, we considered the impact of DNA copy number, set by the multiplicity of infection (MOI), on the variability of the OCT4 distribution generated by the inducible trajectory generator in Fig. 3A and an analogous system in which the inducible system is replaced with a constitutive promoter. We found that an inducible system at a low MOI yielded the highest coefficient of variation of OCT4 expression levels in a cell population (Fig. 3B). Stochastic simulations of these circuit topologies confirmed that the inducible system leads to greater temporal variability in OCT4 expression, allowing us to capture a wide range of trajectories (Fig. 3B). To experimentally validate that an inducible system expressed at a low copy number can be used to generate a wide output distribution, we performed polytransfection experiments, which enable a de-correlation in the delivery of rtTA and rtTA-responsive transcriptional units to cells (*58*), and found that we can achieve a broader output distribution at low plasmid dosages of each component (Fig. 3C). These results

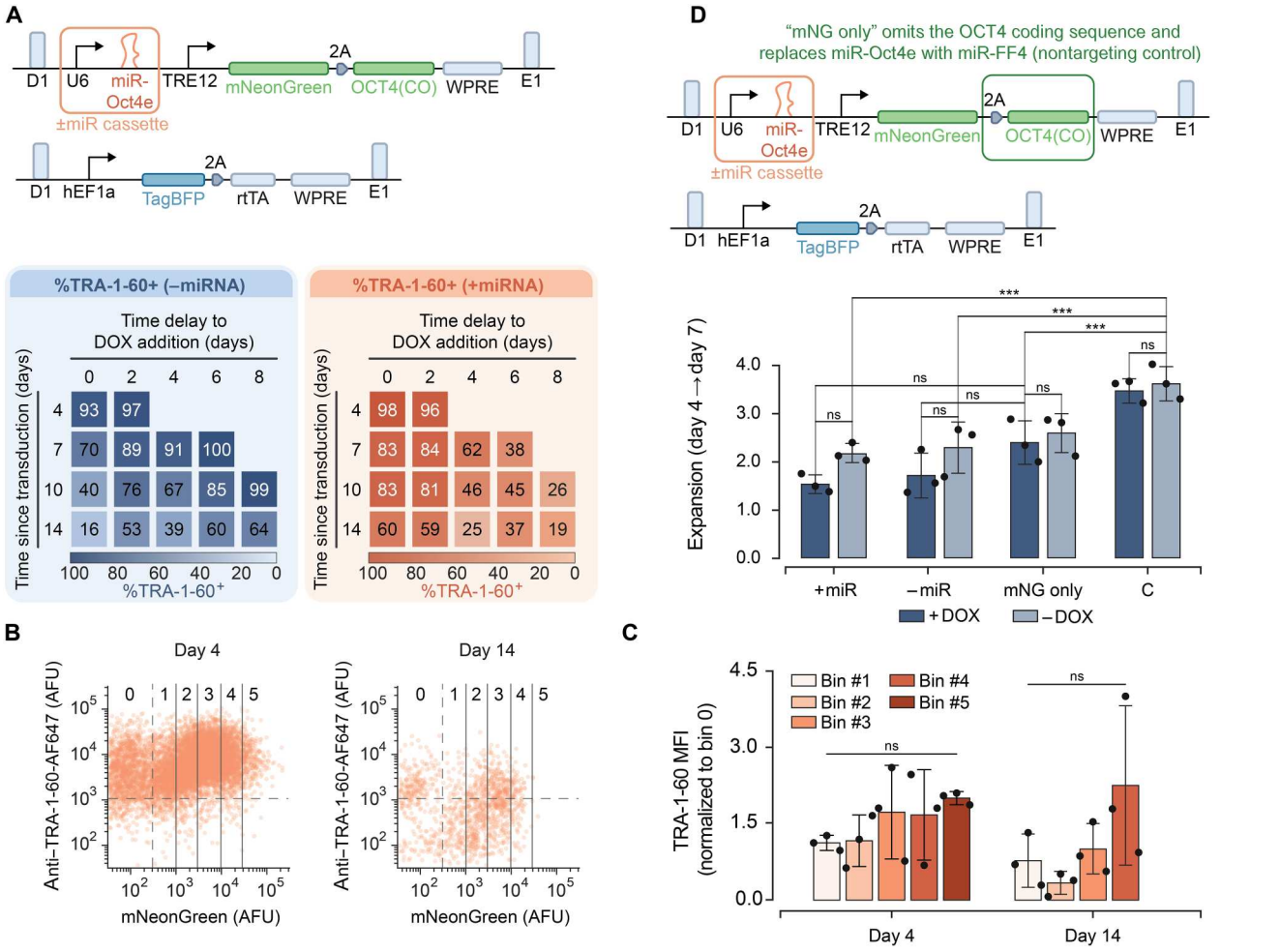

Fig. 2. The importance of total OCT4 dosage for the maintenance of pluripotency. (**A**) "Time delay to DOX addition" is the day after transduction when DOX was added to the medium. The heatmaps summarize the proportion of TRA-1-60–positive iPSCs transduced with variants of this two-lentivirus system (with and without U6-miR-Oct4e). See fig. S3 for replicates and statistical analysis. CO, codon-optimized; WPRE, woodchuck hepatitis virus posttranscriptional regulatory element; TRE12, TRE with 12 Tet-binding sites. (**B**) We transduced iPSCs with the miR-Oct4e–containing system and maintained them in DOX-containing medium. We stained the cells for TRA-1-60 and analyzed them via flow cytometry. We partitioned the mNeonGreen-positive population and (**C**) calculated the bin-wise TRA-1-60 intensity normalized to the TRA-1-60 intensity in bin 0. The dashed lines correspond to the TRA-1-60 and mNeonGreen thresholds. The bars represent the means and SDs of $n = 3$ biological replicates, shown as dots. We omitted bin 5 on day 14 because of low cell counts. A one-way analysis of variance (ANOVA) to compare the effect of mNeonGreen-2A-OCT4 levels on TRA-1-60 signal revealed no significant difference. See fig. S12 for replicates. MFI, median fluorescence intensity; ns, not significant. (**D**) We infected iPSCs with variants of the illustrated circuit. Three days after transduction, we sorted for mNeonGreen-positive, TagBFP-positive iPSCs. We plated 3000 sorted iPSCs per well and tracked their expansion. The bars represent the means and SDs of the expansion values (the ratios of cell counts at days 7 and 4 in $\log_2$ space) for $n = 3$ biological replicates, shown as dots. Asterisks indicate significant differences (***$P < 0.001$) according to an unpaired two-tailed Student's t test. Cells were cultured +DOX and –DOX for days 4 through 7. C, control.

suggest that transducing cells with the inducible trajectory generator system at low MOI would yield highly variable OCT4 levels.

Because mNeonGreen is an exact reporter for total OCT4 levels when endogenous OCT4 is strongly knocked down, we next confirmed the ability of low doses of the miR-Oct4e expression cassette to suppress endogenous OCT4 levels. To this end, we transduced iPSCs with the OCT4-containing trajectory generator lentivirus and another lentivirus that constitutively expresses rtTA (without the other reprogramming factors) at low MOIs; we found a strong correlation between mNeonGreen and OCT4 levels only when the OCT4-containing lentivirus also encoded the U6-driven miR-Oct4e (Fig. 3D). These observations demonstrate that, even at low MOI, the U6 and the TRE promoters drive sufficiently high expression levels to substantially knock down endogenous OCT4 levels and highly express the transgenic OCT4-mNeonGreen fusion protein, respectively; therefore, we can use the signal intensity of mNeonGreen as an accurate proxy for the total level of OCT4.

As previous work has shown that endogenous OCT4 can become activated at an early stage of the reprogramming process (*11*, *22*, *29*, *41*–*44*), we next verified that the trajectory generator is capable of overwriting OCT4 levels during reprogramming by eliminating the endogenous OCT4 contribution throughout a 21-day reprogramming time course. To this end, we reprogrammed HDFs according to the pipeline in Fig. 3E with two variants of

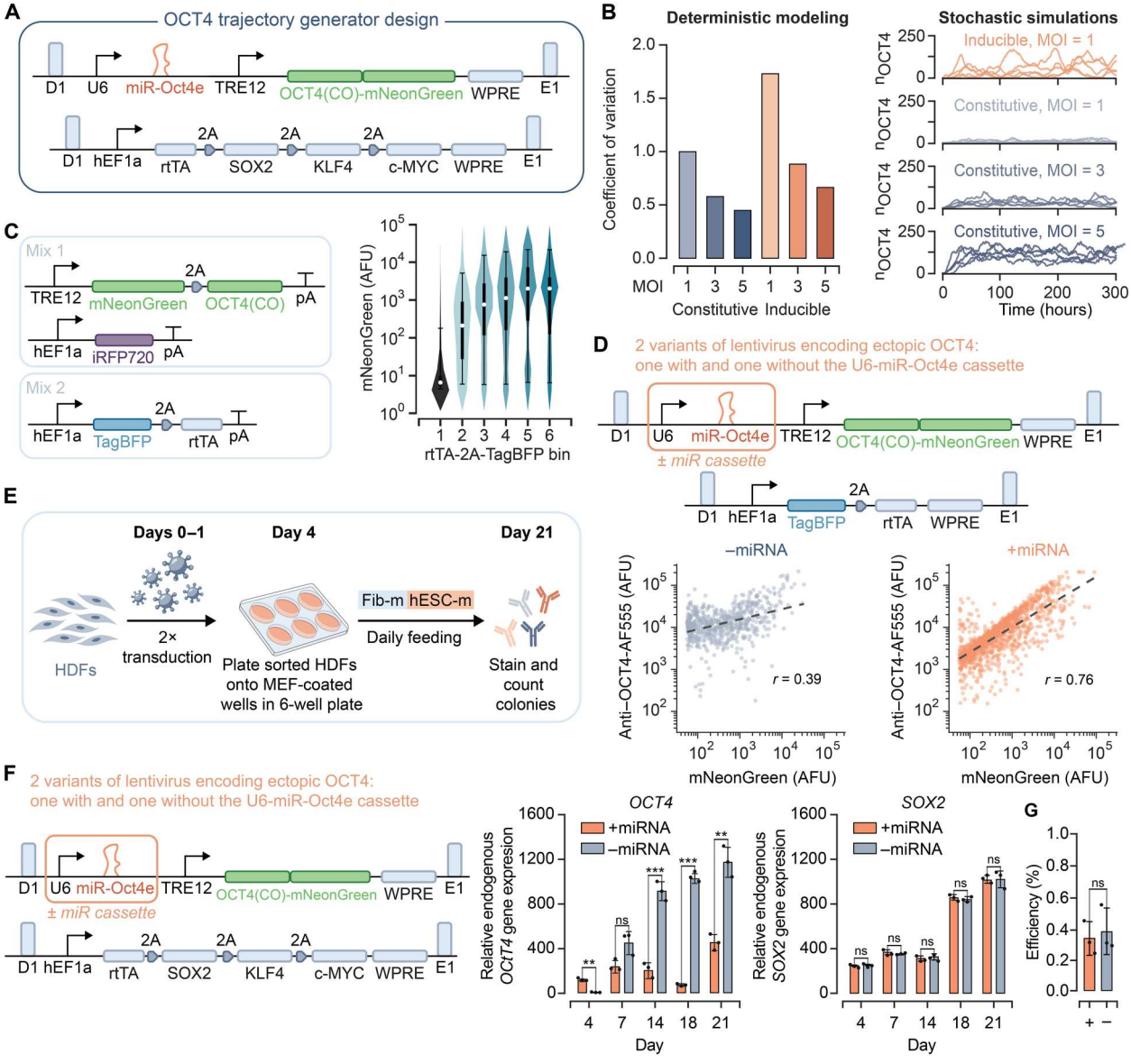

Fig. 3. The trajectory generator: A synthetic genetic circuit for generating variable total OCT4 protein levels during reprogramming. (**A**) Two-lentivirus trajectory generator system. (**B**) Left: We created a deterministic model of the trajectory generator ("inducible") and an analogous system in which a constitutive promoter replaces TRE12 ("constitutive"). We determined the coefficient of variation of their output distribution for different MOIs. Right: We simulated the single-cell OCT4 temporal trajectories for these systems using Gillespie's stochastic simulation algorithm (*102*). (**C**) We performed poly-transfection experiments in HEK293FTs with the illustrated mixes. Violin plots for a low iRFP720 bin illustrate the output distribution as a function of rtTA dosage. (**D**) The scatter plots show flow cytometry data for the double-positive (OCT4, mNeonGreen) for iPSCs transduced with the illustrated two-lentivirus systems. The dashed lines represent the corresponding linear regression fit, with the *r* value shown in each plot. See fig. S13 for replicates. (**E**) Reprogramming pipeline overview. Fib-m, fibroblast medium; hESC-m, human ESC medium. See the Materials and Methods section for recipes. (**F**) We measured endogenous OCT4 and SOX2 in HDFs transduced with these two versions of the trajectory generator via RT-qPCR. We calculated OCT4 and SOX2 levels relative to GAPDH, shown in the bar graph. The error bars represent the SDs for $n = 3$ biological replicates, shown as dots. Asterisks indicate significant differences (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$) according to an unpaired two-tailed Student's *t* test. (**G**) For both systems, we observed comparable reprogramming efficiency, as determined via staining for TRA-1-60 21 days after transduction. The bars represent the means and SDs of the reprogramming efficiency for $n = 3$ biological replicates, shown as dots. No significant difference was found on the basis of an unpaired two-tailed Student's *t* test.

the trajectory generator, one with and one without the U6-driven miR-Oct4e (Fig. 3F). We sampled the HDFs at various time points during reprogramming and measured endogenous OCT4 and SOX2 levels via RT-qPCR. Although the reprogramming efficiencies of the systems with and without miR-Oct4e were comparable, we observed strong activation of endogenous OCT4 in HDFs reprogrammed without miR-Oct4e and found that the miRNA minimizes the endogenous contribution to total OCT4 levels (Fig. 3, F and G). In contrast, the levels of endogenous SOX2, for which we did not implement miRNA-based control, rose during the reprogramming process in both conditions, as expected (Fig. 3F).

A high-throughput live imaging pipeline to track OCT4 trajectories during reprogramming

We next used the trajectory generator (Fig. 3A) to reprogram HDFs in a high-throughput imaging setup to track OCT4 levels over time during the reprogramming process (Fig. 4A). We used a CellVoyager7000 confocal imaging system for automated imaging of the transduced HDFs between one and nine times a day. We verified that mNeonGreen remains an accurate reporter of total OCT4 at the protein level at the end of the time course (Fig. 4, B to E, and fig. S5). We stained the cells for SSEA4 and TRA-1-60, which are surface markers of pluripotent stem cells that have been previously used to gauge colony quality (Fig. 4, F and G) (*11*). As in our previous work, we used the following classifications of colony quality: Type I colonies are SSEA4$^-$, TRA-1-60$^-$. Type II colonies are SSEA4$^+$, TRA-1-60$^-$, and type III colonies are SSEA4$^+$, TRA-1-60$^+$ (Fig. 4, F and G). Type I and II colonies represent cells in

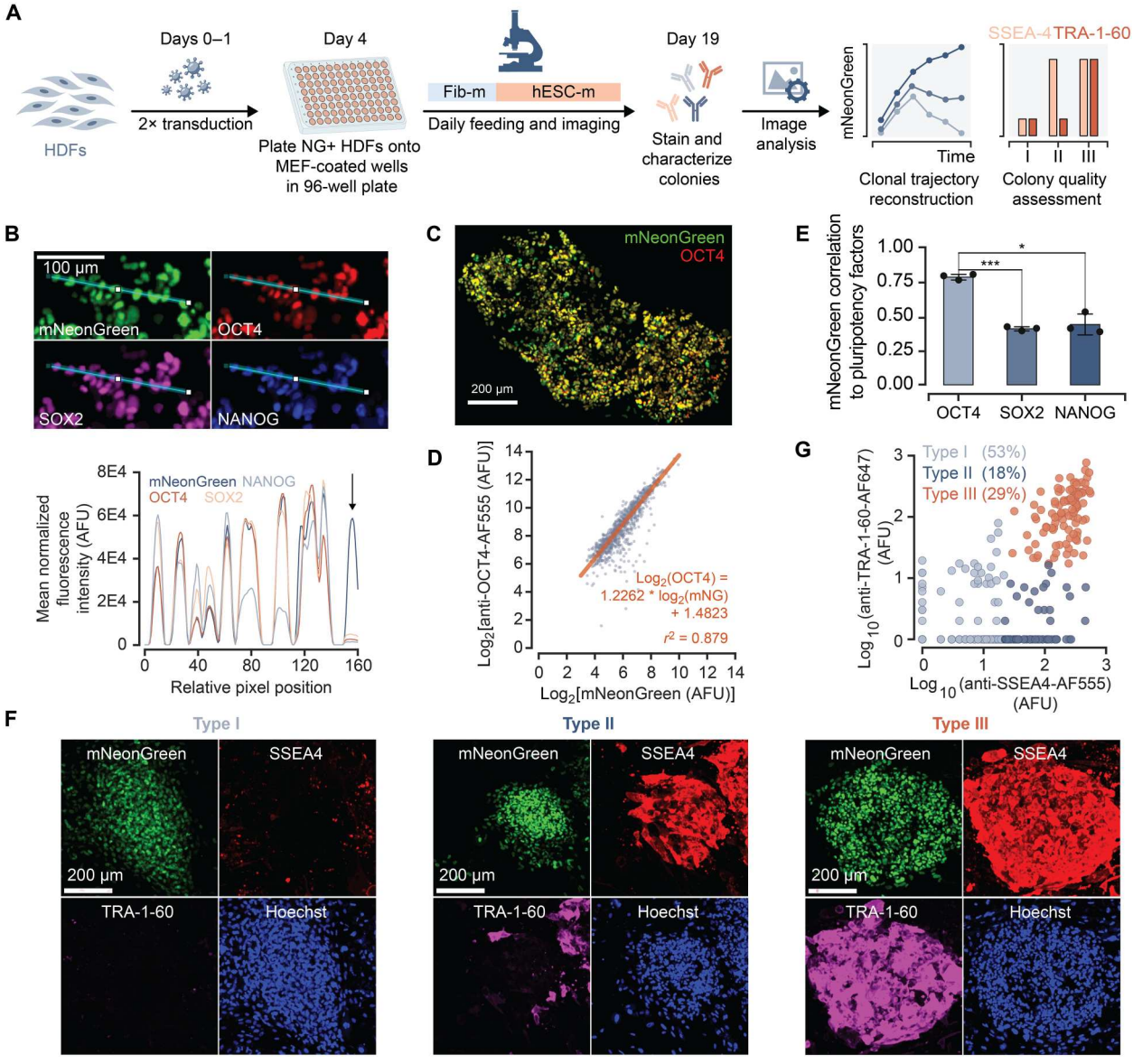

Fig. 4. A high-throughput live-imaging pipeline facilitates the dynamic tracking of OCT4 levels during reprogramming. (**A**) Overview of the high-throughput pipeline used for OCT4 trajectory tracking. (**B**) Stained colonies derived via the trajectory generator reprogramming system in Fig. 3A demonstrate that mNeonGreen and total OCT4 are correlated at the experimental end point. We extracted the mNeonGreen, OCT4, SOX2, and NANOG values along the cyan lines in the images and plotted them in the line graph. The cell corresponding to the position marked by the arrow was lost after fixation, resulting in an apparent lack of consistency between mNeonGreen and OCT4 in the line plots. Scale bar, 100 µm. (**C**) The mNeonGreen and AF555 anti-OCT4 antibody fluorescence microscopy images are overlaid for a representative colony. In this image, we set the values of the areas without cells to 0. Scale bar, 200 µm. (**D**) Each point represents an HDF expressing OCT4-mNeonGreen and stained for OCT4. (**E**) We calculated the Pearson correlation between the mNeonGreen and OCT4, SOX2, and NANOG values of all pixels for three representative colonies. The bars represent the means and SDs of the correlation coefficient for $n = 3$ replicates, shown as dots. Asterisks indicate significant differences ($^*P < 0.05$ and $^{***}P < 0.001$) according to a paired two-tailed Student's t test. (**F**) We classified the reprogrammed cells via SSEA4 and TRA-1-60 staining (*11*). Representative examples are shown in the microscopy images. Scale bar, 200 µm. (**G**) The scatter plot summarizes the frequency of each colony type obtained in the imaging reprogramming runs.

incompletely reprogrammed states, while type III colonies have been shown to be bona fide iPSCs (*11*).

We confirmed with imaging that we could achieve a wide distribution of OCT4-mNeonGreen levels among transduced cells and observed a decline in the number of highly expressing cells 6 days after transduction compared to 4 days after transduction (Fig. 5A), likely because of toxicity associated with high transgene expression levels (Fig. 5B). Virtually all identified colonies, regardless of their degree of SSEA4 and TRA-1-60 staining, arose from a narrow subpopulation of relatively highly expressing cells (Fig. 5C). We found that these cells had a median doubling time of 1.83 days (fig. S6). This result suggests that, in our system, high proliferation does not correlate with reprogramming efficiency; this contrasts with studies of murine reprogramming, in which hyperproliferative cells have been shown to be more likely to successfully reprogram (*59*, *60*). To further probe the effect of OCT4 dosage on reprogramming efficiency, we split the data describing the broad OCT4-mNeonGreen distribution from the first imaging time point into three bins, each containing the same number of cells, and calculated the relative efficiency of reprogramming to each type of colony (Fig. 5D). We found that higher levels of OCT4 were enriched in trajectories for all three colony types and, in particular, type III colonies were produced at the highest efficiency from cells in the highest OCT4 bin.

Next, we reconstructed the trajectories for each colony present at the last time point (fig. S7 and movie S1). Our data show that the clonal trajectories for type I, II, and III colonies are virtually indistinguishable: The vast majority of trajectories are marked by constant, high–OCT4-mNeonGreen levels throughout reprogramming, corresponding to an OCT4 level that is substantially higher than what is found in pluripotent stem cells ($P < 0.0001$, unpaired two-tailed Student's t test; Fig. 5E). In contrast, non–colony-forming cells (HDFs that were transduced with the same virus system but did not produce reprogramming-induced colonies) have trajectories that drop to OCT4-mNeonGreen levels at and below the OCT4 level found in pluripotent stem cells ($P < 0.0001$, unpaired two-tailed Student's t test). The only exception to this otherwise strict correlation was a set of incompletely reprogrammed colonies formed by cells that expressed OCT4-mNeonGreen at barely detectable levels (about 100-fold lower; fig. S8). These findings indicate that colony formation requires a sustained OCT4 level that is substantially higher than what is found in pluripotent stem cells (Fig. 5, E and F). In particular, when OCT4 levels are too low, the cells are unlikely to reach the pluripotent state (Fig. 5D). However, when OCT4 levels are high, the cells expand at a markedly lower rate, perhaps because of either cell death or senescence (Fig. 5B), suggesting that to undergo successful reprogramming, cells must be able to sustain (and tolerate) potentially toxic OCT4 expression levels. Together, these data indicate that a supraphysiological OCT4 level is enriched in the trajectories of successfully reprogrammed cells and is required for type III colony formation. However, the maintenance of this supraphysiological OCT4 level is insufficient, as incompletely reprogrammed colonies also follow the same high OCT4 trajectory.

DISCUSSION

Here, we sought to use synthetic genetic circuits to uncover the temporal trajectories of OCT4 during reprogramming in the rare cells that achieve pluripotency. Reprogramming of HDFs to iPSCs served as a test bed to demonstrate the utility of a genetic circuit design that accurately reports TF levels and allows for the dynamic tracking in live cells of the TF trajectories associated with successful cell fate transitions. The genetic circuit sets total cellular OCT4 level by overwriting the levels of the endogenously expressed TF via a miRNA-based strategy (Fig. 1), inspired by the control theory concept of high-gain negative feedback (*47*). Our miRNA-based approach contrasts with previous work, which relied on either tuning ectopic expression levels without considering the contribution from endogenous expression (*61*) or directly perturbing the endogenous genes at the transcriptional level (*62*, *63*). The former does not allow for full control over TF levels, and the latter requires that the cellular machinery involved in the transcriptional regulation of key genes be co-opted, a phenomenon that may disrupt the ability of cells to stabilize the pluripotent gene regulatory network required for reprogramming. We developed an inherently noisy DOX-inducible system (the trajectory generator; Fig. 3) to generate a variety of total OCT4 trajectories and implemented it to evaluate the effect of temporal OCT4 dynamics on reprogramming. Our approach complements the work in (*64*), in which it was reported that transcriptional noise can promote cell fate transitions. We opted to deliver the trajectory generator via lentiviruses, which are a flexible and well-characterized medium for designing and delivering genetic circuits that are more sophisticated than just simple overexpression systems. Within the context of our high-throughput live imaging approach, lentiviruses are advantageous, as they are less prone to silencing than gammaretroviruses (*65*) and do not face the same technical challenges associated with temporal variability in gene expression from mRNA transfections (*66*).

We applied this tool to hiPSC reprogramming, a well-studied yet inefficient process with potential clinical applications. Our experiments revealed that supraphysiological OCT4 expression is a requirement for successful reprogramming. In particular, we found that the rare cells that reach pluripotency follow a stable level of OCT4 that is substantially higher than that found in iPSCs throughout the entire reprogramming process (Fig. 5). This contrasts with studies that found OCT4 levels comparable to those in iPSCs at the beginning (*39*) and at the end of reprogramming (*67*). However, it is possible that the successfully reprogrammed cells expressed high OCT4 levels even in those experiments and that these low-probability trajectories were "masked" by the bulk measurements of the OCT4 level. With our live reporter of total cellular OCT4, we unambiguously identified the OCT4 expression trajectories of those few cells that successfully reprogram. In addition, a number of papers have investigated the optimal stoichiometry of the OSKM factors (*19*, *28*, *30*), indicating that higher Oct4 levels improve reprogramming efficiency. However, these studies did not provide quantitative information about the specific OCT4 level that is conducive to reprogramming. Via clonal tracking of OCT4 in colony-forming cells, we identified an OCT4 trajectory conducive to the generation of type III hiPSCs.

As stable, high-OCT4 trajectories also result in the generation of incompletely reprogrammed type I and type II colonies, our results indicate that a supraphysiological level of OCT4 expression is necessary, but not sufficient, for successful reprogramming using the Yamanaka factors. The efficiencies observed in this study with our system (up to 0.45%; Fig. 3G) are consistent with those reported by others for primary human fibroblast reprogramming with

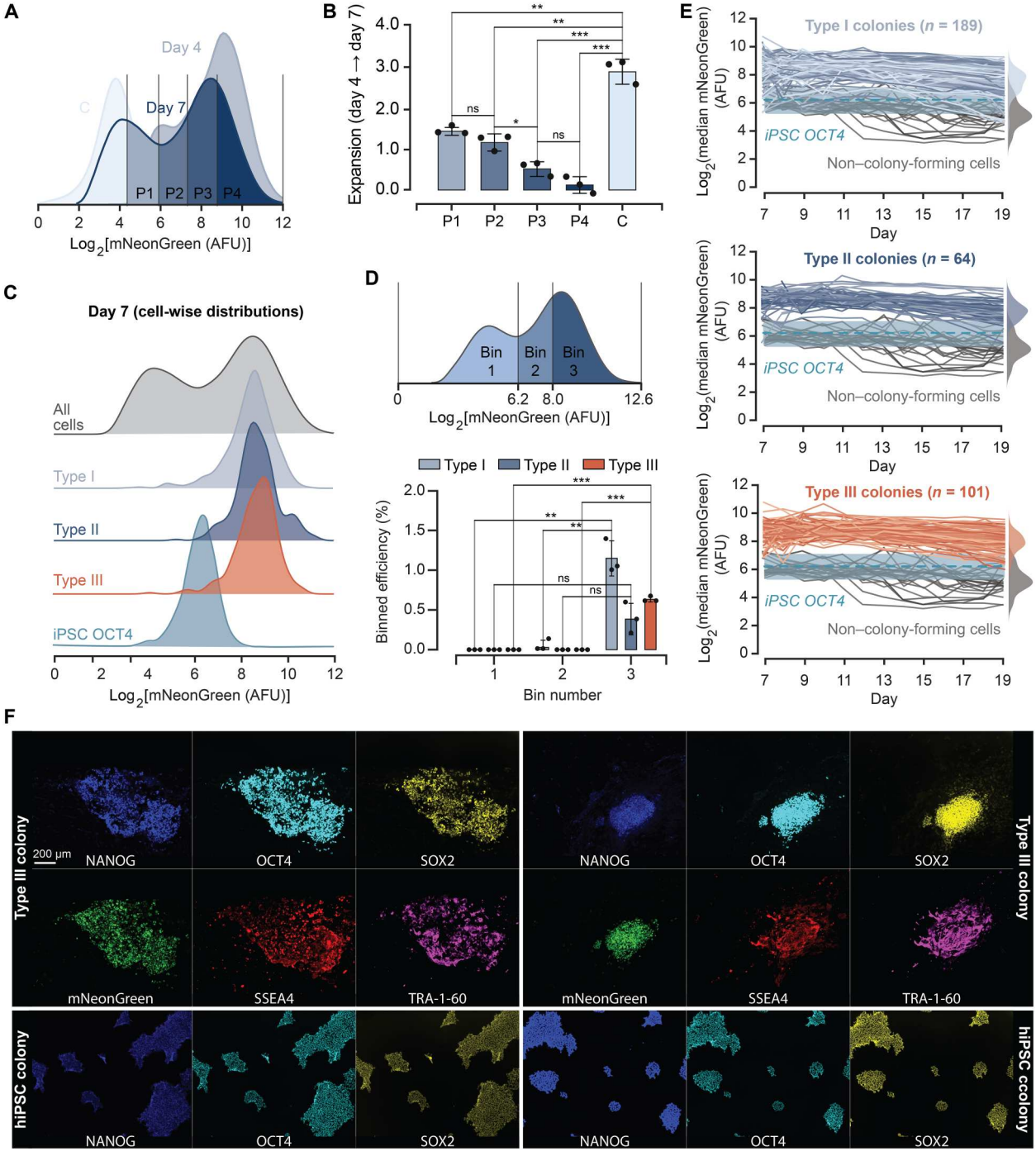

Fig. 5. Trajectory reconstruction of reprogrammed HDFs reveals an "ideal" OCT4 trajectory. (**A**) Histograms show the mNeonGreen distribution of HDFs infected with the trajectory generator 4 and 7 days after transduction. We sorted mNeonGreen-positive HDFs into four populations (P1 to P4) 3 days after transduction. (**B**) We plated 3000 P1 to P4 HDFs per well and tracked their expansion. The bars represent the means and SDs of the expansion values for $n = 3$ biological replicates, shown as dots. Asterisks indicate significant differences (**$P < 0.01$ and ***$P < 0.001$) according to an unpaired two-tailed Student's t test. (**C**) The gray histogram shows the mNeonGreen distribution at the first imaging time point. The distributions labeled by the colony type summarize the cells at this time point that gave rise to each colony type. The teal distribution corresponds to the OCT4 distribution in iPSCs. (**D**) We partitioned the initial mNeonGreen distribution into three sections and calculated the binned efficiency of reprogramming to each colony type. The error bars represent the SDs for $n = 3$ biological replicates, shown as dots. Asterisks indicate significant differences (**$P < 0.01$ and ***$P < 0.001$) according to an unpaired two-tailed Student's t test. (**E**) Each line summarizes the reconstructed trajectory of the median mNeonGreen level for an individual colony. The dashed line and shaded box correspond to the median ± 1 SD of the iPSC OCT4 distribution. The colored histograms summarize the "colony-wise" distributions, in which each point represents the median OCT4-mNeonGreen level for a colony of each type. The gray histograms summarize the "well-wise" non–colony-forming cell distributions, in which each point represents the median OCT4-mNeonGreen level of all cells in a well without colonies. (**F**) Examples of stained type III colonies derived using the trajectory generator and established hiPSC colonies. Scale bar, 200 μm.

lentiviruses [0.002% (*68*), 0.27% (*7*), and 0.1 to 1.5% (*69*)] and not far below those of "high efficiency" secondary reprogramming systems [1 to 3% (*68*) and 0.26 to 2% (*70*)]. Given that we observed a reduction in the expansion of transduced HDFs as a result of high transgene expression, our results suggest that a critical next step for improving reprogramming efficiency is to determine the nature of the toxicity of high OCT4 and develop strategies for specifically mitigating the toxic effects without affecting the reprogramming activity of OCT4. This will entail methods to prevent growth rate reduction in high OCT4 expressors, some of which may be due to the metabolic burden caused by ectopic gene overexpression (*71*).

We showed that high levels of OCT4 enable the highest reprogramming efficiencies with our lentiviral system, in which the other three reprogramming TFs are expressed at fixed levels. Given that these factors are part of an interconnected network (*72*) and that the stoichiometry of these factors has been shown to be relevant for pluripotency acquisition and exit (*19*, *27*, *28*, *73*–*75*), our pipeline can serve as a platform to identify optimal TF expression profiles across other contexts, including other reprogramming systems, in which the levels of the TFs may differ as a result of changes to the OSKM expression cassette design, insertion sites, and copy numbers, as well as across somatic cell populations in which the endogenous contributions of these factors may vary. It is possible that changes made to these and other parameters may result in different "optimal" OCT4 trajectories. In particular, although we focused solely on the OSKM cocktail for hiPSC reprogramming as a case study to explore the utility of a synthetic biology approach for investigating determinants of the efficiency of somatic reprogramming to pluripotency, our approach can be used synergistically with other reprogramming methods that rely on different combinations of TFs, leverage small molecules and various culture conditions, and incorporate epigenetic modifiers (*22*, *44*, *76*–*87*). An in-depth study to investigate the optimal TF dynamics in cells reprogrammed with these different methods could result in the development of optimized protocols that further improve reprogramming efficiency.

Moreover, to achieve a substantial improvement in efficiency, the concurrent steering of multiple TFs following optimized expression profiles may be required, as has been shown in early modeling work (*47*). This would entail identifying which trajectories are optimal for each TF by multiplexing the trajectory generator for the simultaneous tracking of multiple endogenous genes with different sets of miRNAs. For cases where this is required, our approach can be extended to leverage inducible promoters controlled by orthogonal synthetic TFs, leveraging recent technical advancements in CRISPR-mediated transcriptional activation and tunable protein destabilization, among other technologies (*88*–*90*).

We also envision that our high-throughput live imaging pipeline and genetic circuit design can be leveraged to study the roles of the temporal dynamics of key TFs and improve the efficiency of other cell fate conversion applications, such as differentiation and transdifferentiation. This is because the circuit architectures described herein can be easily extended to control any endogenous TF of interest. For instance, our pipeline could be used to study ETV2-driven transdifferentiation of human fibroblasts to endothelial cells (*91*), C/EBP-mediated reprogramming of B cell to macrophages (*92*), or MyoD-induced reprogramming of human fibroblasts to muscle cells (*93*), all of which are dependent on the dosages of key TFs.

Overall, we expect that our strategy will have broad utility in identifying optimal TF trajectories for cell programming applications in general. Hence, this paper establishes a system that can serve both as a tool for scientific discovery and as a stepping stone to designing efficient and robust genetic programming systems that can enforce optimal TF trajectories for regenerative medicine and cell therapy.

MATERIALS AND METHODS

Design of de novo miRNA sequences

We designed the sequences of miR-Oct4e, miR-Oct4f, and miR-Oct4g by inputting the full mRNA transcript sequence of human OCT4 into the Integrated DNA Technologies (IDT) siRNA design tool. We manually scanned sequences to find candidates with the fewest off-target genes. As there are several OCT4 pseudogenes with high sequence identity for OCT4 itself (*94*), it is nearly impossible to find a sequence that selectively targets only OCT4. It is unknown whether miRNA knockdown of the pseudogenes would have an effect on the reprogramming process specifically or the cell generally, although OCT4 pseudogenes appear to be incapable of transcriptional activation (*95*). Of these manually curated sequences targeting the OCT4 open reading frame, we selected three and cloned them as miR-Oct4e, miR-Oct4f, and miR-Oct4g. miRNA sequences are available in table S1.

Cloning

We constructed plasmids using a modular Golden Gate strategy similar to previously reported workflows (*58*, *96*). At a high level, our strategy consisted of the following hierarchy: we assembled parts (referred to as "level 0 s" or "pL0s"), such as insulators, promoters, 5′ untranslated regions (5′UTRs), coding sequences, 3′UTRs, and terminators, to generate transcription units without ("level 1s" or "pL1s") or with ("pVVs") a lentivirus-compatible entry backbone. We constructed pL0s from PCR fragments generated using high-fidelity Q5 polymerase (NEB, #M0494); we then assembled these fragments using In-Fusion (Takara Bio, #638948). Oligonucleotides for the PCRs were synthesized by IDT. We generated pL1s by assembling pL0s and an entry vector through Bsa I Golden Gate reactions [10 to 50 cycles between 16° and 37°C; T4 DNA ligase (NEB, #M0202L); Bsa I (NEB, #R3733L)]. To generate pVVs, we used Bsa I Golden Gate reactions to assemble the pL0s and subsequently ligated the reaction products to a Bsm BI–digested lentivirus transfer plasmid backbone. We propagated all the plasmids in *Escherichia coli* Stellar (Takara Bio), Turbo (NEB, #C2984), or Stable (NEB, #C3040) strains. We plated transformed cells on LB agar (VWR) and cultured in TB (Sigma-Aldrich), with spectinomycin (100 μg/ml), carbenicillin (100 l μg/ml), or kanamycin (50 μg/ml) for selection. We used the QIAprep Spin Miniprep and QIAGEN Plasmid Plus Midiprep Kits for plasmid extraction and subsequently verified the regions of interest (ROIs) in individual clones by Sanger sequencing (Quintara Biosciences).

Mammalian cell culture

We maintained human embryonic kidney (HEK) 293FT cells (Thermo Fisher Scientific) in Dulbecco's modification of Eagle's medium (DMEM; Gibco, #10569010) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco, #16000044), and we split the HEK293FTs using trypsin-EDTA (Gibco, #25300054) every 2 to 3

days before they reached 80% confluence. We used cells at low passage numbers (<15) for all experiments. We cultured primary HDFs (ScienCell, #2300, lot #4382) in fibroblast medium (Fib-m), consisting of DMEM supplemented with 10% (v/v) FBS (Gemini Bio, #100-106), L-glutamine (Life Technologies, #11350-070), 1× MEM non-essential amino acids (Gibco, #11140050), and 1 mM sodium pyruvate (Thermo Fisher Scientific, #11360070). We plated iPSCs on tissue culture plates coated with Matrigel (Sigma-Aldrich, CLS354277) diluted in DMEM/F-12 (Thermo Fisher Scientific, #11320033) according to the manufacturer's suggestions and maintained them in mTeSR$^+$ (STEMCELL Technologies, #100-0276). We passaged the colonies using ReLeSR (STEMCELL Technologies, #5872) every 2 to 3 days to prevent differentiation. Following single-cell passaging, we added ROCK inhibitor (Y-27632 dihydrochloride; Sigma-Aldrich, Y0503) to a final concentration of 10 μM for 24 hours. We grew all the cells in a humidified atmosphere at 37°C with 5% CO_2.

Transfections

For transient transfection of HEK293FTs and iPSCs, we used the Lipofectamine 3000 (Invitrogen, #L3000015) and Lipofectamine Stem (Invitrogen, STEM00015) reagents, respectively. We performed cotransfections in 96-well and poly-transfections in 24-well tissue culture–treated plates (Costar). For each cotransfection experiment, we transfected a total of 120 ng of plasmid DNA per well. We diluted the DNA into 10 μl of Opti-MEM (Gibco, #31985062), added the transfection reagent, and then vortexed the mixture. After a 10 to 30 min of incubation of the DNA-reagent mixtures, we added the transfection mixes to each well first and then plated 40,000 cells. We used Lipofectamine 3000 at a ratio of 2 μl of P3000 reagent and 2 μl of Lipofectamine 300 per 1 μg of DNA and Lipofectamine Stem at a ratio of 2 μl of transfection reagent per 1 μg of DNA. For all transfections with DOX (Sigma-Aldrich, D9891), we added DOX (1000× stock: 1 μg/ml diluted in water) to the medium immediately after transfection. For poly-transfections, we scaled up the total amount of DNA and reagents based on the surface area difference between 96- and 24-well plates and used equal amounts of DNA for each transfection complex. In each transfection complex, we included a hEF1a-driven transfection marker to indicate the dosage of DNA delivered to each cell and to facilitate consistent gating of transfected cells.

Fluorescence analyses

We analyzed fluorescent protein expression and staining intensity by flow cytometry. To do so, we harvested cells 48 hours after transfection or after the amount of time described in the main text using trypsin-EDTA for HEK293FTs and HDFs or ReLeSR for iPSCs. We washed the cells three times with flow cytometry buffer, made of phosphate-buffered saline (PBS) without calcium or magnesium (Corning, #21031CV) supplemented with 1% FBS and 5 mM EDTA. We kept the cells on ice until analysis with the HTS module of a BD LSR II flow cytometer (Koch Institute flow cytometry core). We analyzed the data using MATLAB scripts (based on https://github.com/jonesr18/MATLAB_Flow_Analysis).

Lentivirus production

We seeded Lenti-X 293T cells (Takara, #632180) in 15-cm tissue culture dishes (CellTreat, #229652) 24 hours before transfection such that they would be 90% confluent the next day. One hour before transfection, we replaced the medium. We prepared the transfection mix consisting of 2.5 ml of Opti-MEM, 31 μg of psPax.2, 13 μg of pMD2.G, 38 μg of the lentivirus construct, and 82 μl of PEIpro (PolyPlus, #101000017). psPAX2 and pMD2.G were a gift from D. Trono (Addgene plasmid 12260 and 12259, respectively). After a 15-min incubation period, we gently added the transfection mix dropwise to each plate. Twenty-four hours later, we replaced the medium, which we supplemented with Hepes buffer at a final concentration of 25 mM (Sigma-Aldrich, SRE0065). Two days later, we harvested the medium, centrifuged it at 3000 rcf for 5 min, filtered it with 0.22-μm filter (VWR, #97066-200), added Lenti-X concentrator (Takara, 631232) to the supernatant according to the manufacturer's instructions, and kept the mixture at 4°C on a gyratory rocker for at least 1 hour. Then, we centrifuged the mixture at 4°C and 1500*g* for 45 min. We carefully decanted the supernatant and resuspended the pellet containing the vector particles with 800 μl of DPBS (Gibco, #14190144) + 0.001% Pluronic F68 (Gibco, #24040032).

Transduction experiments

Before the lentiviral transduction experiments, we measured the functional titers of each lentivirus batch. We aliquoted newly produced viruses into small volumes (5 to 50 μl) and flash-froze them in liquid nitrogen. At least 1 day after freezing, we thawed one vial per batch to estimate the titer of each batch. We diluted the thawed virus in the appropriate medium at $1:1 \times 10^2$, $1:1 \times 10^3$, $1:1 \times 10^4$, $1:1 \times 10^5$, $1:1 \times 10^6$, and $1:1 \times 10^7$ ratios, and supplemented the medium with polybrene (Millipore-Sigma), which facilitates viral uptake, at a final concentration of 8 μg/ml. We replaced the diluted virus solutions with fresh medium 12 hours after transduction.

If needed, we cotransduced at high MOI (5 to 10) a "helper" virus expressing rtTA or other necessary genes for expression of a given virus. Afterward, we changed the medium daily. Four days after infection, we measured via flow cytometry the fluorescence of either a fluorescent reporter protein or a fluorophore-conjugated antibody-stained factor. We assumed that the lentiviral infections follow a Poisson distribution (*97*–*99*) and estimated the titers by determining the linear range of the dilution expected on the basis of the Poisson distribution (i.e., where a 10× dilution of the virus leads to a 10× reduction in fluorescence-positive cells).

For lentiviral transduction experiments, we plated cells of interest 24 hours before infection. On the day of transduction, we sacrificed one well to determine viable cell counts by Trypan Blue exclusion on a hemocytometer. We calculated the amount of virus to add per well on the basis of the number of cells and the MOI (MOI = 1 for the OCT4-containing trajectory generator virus, MOI = 10 for the trajectory controller viruses, and MOI = 1 for all SKM-containing viruses). We added the appropriate volume of virus solution to medium containing polybrene at a final concentration of 8 μg/ml. We replaced the diluted virus solutions with fresh medium 12 hours after transduction. For experiments requiring double transductions, we cultured the cells in fresh medium for 12 hours before reinfection. We harvested cells for assays at the time points stated here.

Reprogramming of HDFs to iPSCs

We thawed low passage (passage 3) primary HDFs (cryopreserved in DMEM + 10% FBS) and seeded 10,000 cells per well in a 24-well plate 24 hours before transduction (i.e., we plated the cells on "day

−1″). The next day, we trypsinized two wells with briefly warmed Trypsin EDTA 0.05% and obtained a cell count per well using Trypan Blue exclusion on a hemocytometer. We calculated the volume of each virus necessary to achieve the desired MOI and then proceeded with transduction as previously described. In all reprogramming experiments, we infected twice using the timing described above. On the third day after transduction, we prepared either 6-well plates (Costar) or 96-well plates (ibidi, #89626) with irradiated feeder mouse embryonic fibroblasts (MEFs; Gibco, A34180). To do this, we coated each well with 0.1% gelatin (dissolved in sterile water; Millipore Sigma, G1890) and incubated the plates at 37°C. We then thawed, counted, and plated MEFs at a density of 250,000 MEFs per well in a 6-well plate (scaled on the basis of surface area for a 96-well plate). On the fourth day after transduction, we trypsinized and sorted the mNeonGreen-positive transduced cells in experiments with the trajectory generator or the mNeonGreen-positive, tagBFP-positive transduced cells in experiments with the trajectory controllers using a BD FACSAria III cell sorter or a Sony MA900 Multi-Application Cell Sorter. The sorted fibroblasts were plated at a density of 150 cells per well in a 96-well plate. Until the seventh day after transduction, we fed the cells daily with Fib-m (fibroblast medium) supplemented with DOX, after which we fed the cells daily with hESC-m (human ESC medium). We made hESC-m using 195 ml of DMEM/F-12 with 15 mM Hepes (STEMCELL Technologies, #36254), 50 ml of KnockOut Serum Replacement (Gibco, #10828028), 1× MEM non-essential amino acids (Gibco, #11140050), 1× β-mercaptoethanol (Gibco, #31350010), and DOX. Until the end point, we maintained the reprogramming cells in this medium, to which fresh fibroblast growth factor 2 (Thermo Fisher Scientific, PHG0021) was added to a final concentration of 10 ng/ml daily. Throughout this period, we monitored the density of the MEFs in the wells and replated MEFs as they dropped to approximately 50% of the initial plating density.

Quantification of gene expression

For the time course RT-qPCR experiment in Fig. 3F, we sorted for the mNeonGreen-positive cells using a BD FACSAria III Cell Sorter. We collected the cells in microfuge tubes containing 200 µl of RLT buffer (QIAGEN, #79216). We kept the cells at −80°C until we collected samples for each time point. We then thawed the samples and extracted RNA from the sorted HDFs using an RNeasy Micro Kit (QIAGEN, #74004) according to the manufacturer′s instructions. We used about 100 ng of total RNA in each RT-qPCR reaction using the Luna Universal One-Step RT-qPCR Kit (NEB, #E3005), set up according to the manufacturer′s instructions and analyzed on a CFX Opus 96 instrument (Bio-Rad) in the SYBR-green channel. We normalized gene expression within each biological sample to the levels of the housekeeping gene GAPDH. Primer sequences are available in table S2.

Staining

For the experiments in Figs. 1, 2, and 3, we stained for the specified proteins and quantified the immunofluorescence signal via flow cytometry. To do this, we collected the cells using either trypsin for HDFs or ReLeSR for iPSCs followed by centrifugation. After centrifugation, we resuspended the cells in 150 µl of PBS and transferred them to a 96-well plate. We washed the cells twice with PBS and fixed them using 50 µl of a fixation solution (BioLegend, #420801). After a 10-min incubation at room temperature and being protected from light, we washed the cells twice using fluorescence-activated cell sorting (FACS) buffer and resuspended them in a staining solution containing (i) a permeabilization buffer (BioLegend, #420801) diluted 10-fold in PBS and (ii) the relevant antibodies diluted to yield a total volume of 50 µl of buffer per well. For staining of surface markers such as TRA-1-60, we used FACS buffer in place of the permeabilization buffer. We used the following antibodies and dilutions: mouse anti-human TRA-1-60 AF647 antibody (BD Biosciences, #560122) at a 1:100 dilution, mouse anti-human OCT4 AF555 antibody (BD Biosciences, #560306) at a 1:25 dilution, mouse anti-human OCT4 PerCP-Cy5.5 antibody (BD Biosciences, #560794) at a 1:10 dilution, and mouse anti-human SOX2 AF647 antibody (BD Biosciences, #562139) at a 1:100 dilution.

For the experiments in Figs. 4 and 5, we stained for TRA-1-60 and SSEA4 and quantified the immunofluorescence signal via imaging. At the end of the reprogramming experiment (time points are specified here), we fixed the cells in the 96-well plates with a 4% paraformaldehyde solution (Sigma-Aldrich, #1004960700) for 15 min at room temperature and then washed the wells twice with DPBS (Gibco, #14190144). We stored the plates for up to 1 week at 4°C. After fixation, we prepared a blocking buffer solution containing 5.19 ml of PBS, 3.31 ml of bovine serum albumin fraction V 7.5% (Gibco, #15260037), 746 µl of goat serum (Sigma-Aldrich, G9023), and 746 µl of horse serum (Sigma-Aldrich, H0146) and added it to the plates. Following a 20-min incubation period at room temperature, we washed the cells twice with a wash/stain buffer, prepared by diluting the blocking buffer 10-fold in PBS. We stained the cells for 2 hours at room temperature using a staining solution containing 1:2000 mouse immunoglobulin G (IgG) anti-human SSEA4 antibody (BioLegend, #330402) and 1.2:2000 mouse IgM anti-human TRA1-60 antibody (BioLegend, #330602). Then, we washed the cells twice and incubated the wells in a secondary solution containing 1:400 goat AF555 anti-mouse IgG antibody (Thermo Fisher Scientific, #A280180) and 1:2000 goat anti-mouse IgM AF647 antibody (Thermo Fisher Scientific, #A21238) for 45 min at room temperature. Last, we washed the wells twice with wash buffer and added 450 µl of PBS before imaging. If necessary, we removed the fluorescent signal from mNeonGreen or previous immunostaining by photochemical bleaching as described in (*100*).

Time course microscopy imaging

We performed live imaging of the transduced HDFs plated onto MEFs in square-well 96-well plates (ibidi, 89626) one to nine times per day from reprogramming days 7 to 21 or 22 on a CellVoyager7000 confocal imaging system using a 10× lens, 488-nm laser excitation (at 50% power), and environmental control [37°C, 5% CO_2; MicroClime lids (Beckman, #001-5718)]. We acquired the entire area of each well (0.56 cm^2) using 30 (5 × 6) fields per well, 8 *Z*-levels (10 µm apart), 2 × 2 binning, and 150-ms exposure per image for channel 1 (mNeonGreen, 525/20-nm emission filter) and an "empty" channel 3 (used to capture background and autofluorescence, as well as mNeonGreen signal "bleed through" at 600/37 nm). In total, we acquired over 12 million images throughout three time course imaging experiments.

To compare live channel 1 fluorescence intensity measurements acquired before the first and after completion of the last time course imaging run, we imaged an hiPSC line with highly consistent enhanced green fluorescent protein (EGFP) expression (line 1701

with monoallelic mEGFP-tagged HIST1H2BJAICS-0061-036, Coriell Institute). We determined that the medians of the integrated mEGFP signals (M-phase 2N chromatin) varied by less than 0.5% (137,798 versus 138,386), indicating that imager system performance did not deteriorate over the time span of the experiments.

Time course image processing and analysis

We carried out the following image preprocessing steps on the acquired channel 1 and 3 images: *Z*-projection, background subtraction, flat-field correction, and montaging. To facilitate image alignment throughout the time course and identification of nuclear ROIs, we calculated a ratio image ("C13") by dividing the signal intensity from the preprocessed channel 1 ("C01") and channel 3 ("C03") images (fig. S9A). We identified the locations of candidate nuclear ROIs by finding the local maxima after removing noise and outlier pixels and applying a Gaussian blur followed by dynamic thresholding, watershedding, and application of nuclear size and shape selection criteria (fig. S8, A to E). For each nuclear ROI, we extracted the mNeonGreen signal on the basis of the fluorescence intensity in the C01 images. All image processing and analysis was performed using FIJI (2.0.0-rc-69/1.52p) on a MacPro (64GB RAM, OS10.15.6); all scripts are available upon request.

We used fluorescent beads (TetraSpeck 4 μm, Thermo Fisher Scientific, T7283, lot #55790A; Inspeck Green, Thermo Fisher Scientific, I14785, lot #2407893) to determine the required *Z*-depth to confirm signal linearity and consistency of the fluorescence intensity measurements over time and to enable comparison of signal intensities between imaging and flow cytometry data (fig. S8).

Imaging-based comparison of SOX2, OCT4:mNeonGreen, and OCT4 levels

We plated the transduced HDFs at low density on MEFs and multiple hiPSC lines (723, 1157.2, 603, 1698, 1566, 280, 1400, 1480, and PGP1) on MEFs using hESC medium in parallel wells of square-well 96-well plates. We imaged the live HDFs to capture mNeonGreen levels using the same conditions and procedures as described above for the time-course imaging. After fixing and staining the HDFs and iPSCs for OCT4 and SOX2 as described above, we imaged them to capture immunofluorescence levels. We used our image analysis pipeline to identify the nuclei on processed (flat field–corrected, background-subtracted, montaged, and aligned) OCT4 images for hiPSC colonies. To identify transduced HDF nuclei, we created masks by merging the thresholded OCT4 and SOX2 images to capture OCT4- and/or SOX2-positive cells (fig. S10). We used the C13 images to identify the nuclear ROI of the same cells during live imaging. We rejected from further analysis cells for which no matching C13 nucleus could be found (if greater or fewer than one nucleus was found within 50 pixels or if the nuclear shape or size did not match that of the OCT4/SOX2 nuclear region). We defined OCT4 and SOX2 antibody background staining levels as the 40th percentile of the mean fluorescence intensities observed in the nuclei of MEFs (which do not express human OCT4 or SOX2) present in the same well and subtracted the obtained values from OCT4 and SOX2 images before fluorescence intensity measurements.

Trajectory reconstruction

To identify the colonies present at the end of the imaging time course, we took a *Z* projection of each well (*Z*: time axis). We then applied a Gaussian blur and threshold to these images to location regions likely to contain colony formation events (fig. S9B). For each of these regions, we selected those with greater than 50 cells at the final time point for further analysis. We used "reverse" and "forward" trajectory reconstruction strategies. For the reverse trajectory reconstruction, we created a polygon for the last time point based on the region identified in the previous step. For each subsequent time step, we expanded and contracted the polygon to identify cells likely to belong to the same colony formation event. Given that this strategy is more successful after the colony has begun to form, we coupled this analysis with a forward algorithm, in which we predicted cellular movement and division among cells between two consecutive early time points. This way, we generated trajectory predictions for each cell present at the first time point. As this strategy is more accurate at early time points, we determined a cutoff, set as the first time point at which the colony contains at least 50 cells. Last, we manually validated each trajectory to improve the accuracy of our predictions. We discarded from analysis colonies that formed but were not present at the staining time point and colonies for which we could not confidently reconstruct a trajectory. We report the median mNeonGreen signal intensity for a given nuclear ROI as the expression level for each corresponding cell and found that the integrated density and the median of all the pixels within the nuclear ROI are correlated (fig. S11), demonstrating that our results also hold for integrated density measurements.

Lineage tracing via barcoding

This protocol (fig. S4A) was developed in collaboration with the Wrana Lab (Lunenfeld-Tanenbaum Research Institute, Toronto). The barcode library was generated by adapting a previously reported barcoding strategy (*101*) by introducing the barcode region into the pLVX DsRed N1 plasmid (Clontech Laboratories). This barcoding technique involved lentiviral integration of a variable DNA sequence ligated to an RFP reporter. The DNA barcode region is 27 base pairs long and composed of 12 interspersed variable bases (NNATCNNGATSSAAANNGGTNNAACNN). The constant base pair region enables sequencing read error frequencies to be quantified. The library has a theoretical estimated complexity of about 36,000 unique possible barcodes, but experimentally, we note a frequency of around 18,000 unique barcodes. We transduced HDFs with barcoding virus at MOI 0.1 (at a low MOI such that each cell only receives one barcode) on day −3. On day 0, we counted the cells and transduced them with the reprogramming lentiviruses in Fig. 3A. We sampled one-third of the cells for sequencing to determine the number of unique barcodes. On day 6, we sorted the cells on the basis of low, medium, and high OCT4 (measured via mNeonGreen fluorescence) and plated them separately in three parallel wells per condition. We sampled the cells for sorting and barcode sequencing on days 10 and 14. On day 22, we stained the cells for TRA-1-60 to isolate successfully reprogrammed cells for sequencing.

Modeling

We compared two expression systems: (i) a two-virus system encoding a transcriptional activator (*A*) and its cognate strong inducible promoter, which drives expression of the output protein (*Y*), on

separate constructs, and (ii) a one-virus system encoding a constitutive promoter driving expression of the output protein (Y). In the first system, the transcriptional activator A binds to the promoter only if it is bound to DOX.

We analyzed ordinary differential equations (ODEs) describing each system to compare the distributions of Y for each system transduced at different MOIs. Assuming that the binding and unbinding reactions occur on a much faster time scale than the protein production and decay reactions (*46*) and that DOX is at a saturating concentration, we wrote the ODEs for system (i) as

$$\begin{aligned}\dot{m}_A &= \alpha_A n_2 - \delta_A m_A, \dot{A} = \beta_A m_A - \gamma_A A, \dot{m}_Y \\ &= \alpha_Y H(n_2) n_1 - \delta_Y m_Y, \dot{Y} = \beta_Y m_Y - \gamma_Y Y \end{aligned} \quad (1)$$

in which α_A and α_Y represent the transcription rate constants for A and Y, β_A and β_Y represent the translation rate constants for A and Y, δ_A and δ_Y represent the mRNA dilution rate constants for A and Y, $\gamma_A$and γ_Y represent the protein dilution rate constants for A and Y, and n_2 and n_1 represent the DNA copy number related to A and Y, respectively. $\dot{m}_A$ and $\dot{m}_Y$ represent the dynamics of each mRNA species, and $\dot{A}$and $\dot{Y}$ represent the dynamics of each protein species. Furthermore, $H(n_2)$ represents a Hill function (*46*), which can be written as $H(n_2) = (n_2\overline{A}/K_{dT})/(1 + n_2\overline{A}/K_{dT})$, where $\overline{A}$ represents the concentration of A expressed from one DNA molecule ($n_2 = 1$) and K_{dT} corresponds to the dissociation constant of A from the inducible promoter.

Using the same assumptions described above, we wrote the ODEs for system (ii) as

$$\dot{m} = \alpha_Y n_1 - \delta_Y m_Y, \dot{Y} = \beta_Y m_Y - \gamma_Y Y \quad (2)$$

From Eqs. 1 and 2, we obtained the expression for the steady-state concentration of Y for systems (i, left) and (ii, right)

$$\begin{aligned} Y_{ss} &= \frac{\beta_Y \alpha_Y}{\gamma_Y \delta_Y} H(n_2) n_1 = \frac{\beta_Y \alpha_Y}{\gamma_Y \delta_Y} \frac{\frac{\beta_A \alpha_A}{\gamma_A \delta_A} \frac{n_2}{K_{dT}}}{1 + \frac{\beta_A \alpha_A}{\gamma_A \delta_A} \frac{n_2}{K_{dT}}} n_1 \\ &= \frac{\beta_Y \alpha_Y}{\gamma_Y \delta_Y} \frac{\frac{\overline{A_{ss}}}{K_{dT}} n_2}{1 + \frac{\overline{A_{ss}}}{K_{dT}} n_2} n_1, Y_{ss} = \frac{\beta_Y \alpha_Y}{\gamma_Y \delta_Y} H(n_2) n_1 \end{aligned} \quad (3)$$

in which we introduced the parameter $\overline{A_{ss}} = (\beta_A \alpha_A)/(\gamma_A \delta_A)$.

To evaluate how the copy number variation affects levels of Y, we modeled n_1 and n_2 as random variables following a Poisson distribution for which the MOI is the Poisson parameter: $P(n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$ with $(n, \lambda) = \{(n_1, \lambda_1), (n_2, \lambda_2)\}$, in which $\lambda_1 = \mathrm{MOI}_1$ and $\lambda_2 = \mathrm{MOI}_2$.

We compared the coefficient of variation (CV) of the distribution of Y_{ss} of the two systems at different MOIs. The CV formula is given by $\mathrm{CV}(X) = \frac{\sqrt{E[X^2] - E[X]^2}}{E[X]}$.

We also performed stochastic simulations of the two systems to study their temporal dynamics. To this end, we constructed the following chemical reactions for system (i)

$$\begin{gathered} n_2 \xrightarrow{\alpha_A} n_2 + m_A, \quad m_A \xrightarrow{\beta_A} m_A + A, \quad m_A \xrightarrow{\delta_A} \emptyset, \quad A \xrightarrow{\gamma_A} \emptyset, \\ A + D \xrightarrow{a} C, \quad C \xrightarrow{d} A + D, \quad C + P_0 \xrightarrow{a} P_1, \quad P_1 \xrightarrow{d} C + P_0, \\ P_1 \xrightarrow{d} C + P_0, \quad C \xrightarrow{\gamma_A} \emptyset, P_1 \xrightarrow{\gamma_A} P_0, \quad P_1 \xrightarrow{\alpha_Y} P_1 + m_Y, \\ m_Y \xrightarrow{\beta_Y} m_Y + Y, \quad m_Y \xrightarrow{\delta_Y} \emptyset, \quad Y \xrightarrow{\gamma_Y} \emptyset \end{gathered} \quad (4)$$

in which D represents the DOX concentration, C represents the complex between D and A, P_1 represents the activated promoter, that is, the complex between the Y promoter (P_0) and the complex C, and a and d represent the binding and unbinding reaction rate constants, respectively. Furthermore, because of the DNA conservation law, we know that $n_{P_0} + n_{P_1} = n_1$.

For system (ii), we derived the following chemical reactions

$$n_1 \xrightarrow{\alpha_Y} n_1 + m_Y, \quad m_Y \xrightarrow{\beta_Y} m_Y + Y, \quad m_Y \xrightarrow{\delta_Y} \emptyset, \quad Y \xrightarrow{\gamma_Y} \emptyset \quad (5)$$

We simulated the full set of reactions written above with Gillespie's stochastic simulation algorithm (*102*).

Supplementary Materials

This PDF file includes:

Figs. S1 to S13
Tables S1 and S2
Legend for movie S1
Legend for data S1
References

Other Supplementary Material for this manuscript includes the following:

Movie S1
Data S1

REFERENCES AND NOTES

1. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
2. K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
3. A. E. Omole, A. O. J. Fakoya, Ten years of progress and promise of induced pluripotent stem cells: Historical origins, characteristics, mechanisms, limitations, and potential applications. *PeerJ* **2018**, e4370 (2018).
4. Y. Y. Lipsitz, N. E. Timmins, P. W. Zandstra, Quality cell therapy manufacturing by design. *Nat. Biotechnol.* **34**, 393–400 (2016).
5. Y. Y. Lipsitz, P. Bedford, A. H. Davies, N. E. Timmins, P. W. Zandstra, Achieving efficient manufacturing and quality assurance through synthetic cell therapy design. *Cell* **20**, 13–17 (2017).
6. K. Takahashi, S. Yamanaka, A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183–193 (2016).
7. T. M. Schlaeger, L. Daheron, T. R. Brickler, S. Entwisle, K. Chan, A. Cianci, A. DeVine, A. Ettenger, K. Fitzgerald, M. Godfrey, D. Gupta, J. McPherson, P. Malwadkar, M. Gupta, B. Bell, A. Doi, N. Jung, X. Li, M. S. Lynes, E. Brookes, A. B. Cherry, D. Demirbas, A. M. Tsankov, L. I. Zon, L. L. Rubin, A. P. Feinberg, A. Meissner, C. A. Cowan, G. Q. Daley, A comparison of non-integrating reprogramming methods. *Nat. Biotechnol.* **33**, 58–63 (2015).
8. Y. Buganim, D. A. Faddah, R. Jaenisch, Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* **14**, 427–439 (2013).
9. K. Hochedlinger, R. Jaenisch, Induced pluripotency and epigenetic reprogramming. *Cold Spring Harb. Perspect. Biol.* **7**, a019448 (2015).

10. A. H. Lang, H. Li, J. J. Collins, P. Mehta, Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. *PLOS Comput. Biol.* **10**, e1003734 (2014).
11. E. M. Chan, S. Ratanasirintrawoot, I. H. Park, P. D. Manos, Y. H. Loh, H. Huo, J. D. Miller, O. Hartung, J. Rho, T. A. Ince, G. Q. Daley, T. M. Schlaeger, Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat. Biotechnol.* **27**, 1033–1037 (2009).
12. T. S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander, A. Meissner, Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
13. J. S. Kim, H. W. Choi, S. Choi, H. G. Seo, S. H. Moon, H. M. Chung, J. T. Do, Conversion of partially reprogrammed cells to fully pluripotent stem cells is associated with further activation of stem cell maintenance- and gamete generation-related genes. *Stem Cells Dev.* **23**, 2637–2648 (2014).
14. S. M. Hussein, M. C. Puri, P. D. Tonge, M. Benevento, A. J. Corso, J. L. Clancy, R. Mosbergen, M. Li, D. S. Lee, N. Cloonan, D. L. Wood, J. Munoz, R. Middleton, O. Korn, H. R. Patel, C. A. White, J. Y. Shin, M. E. Gauthier, K. A. L. Cao, J. I. Kim, J. C. Mar, N. Shakiba, W. Ritchie, J. E. Rasko, S. M. Grimmond, P. W. Zandstra, C. A. Wells, T. Preiss, J. S. Seo, A. J. Heck, I. M. Rogers, A. Nagy, Genome-wide characterization of the routes to pluripotency. *Nature* **516**, 198–206 (2014).
15. S. Sebban, Y. Buganim, Nuclear reprogramming by defined factors: Quantity versus quality. *Trends Cell Biol.* **26**, 65–75 (2016).
16. L. A. Boyer, I. L. Tong, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, R. A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
17. A. Radzisheuskaya, J. C. Silva, Do all roads lead to Oct4? The emerging concepts of induced pluripotency. *Trends Cell Biol.* **24**, 275–284 (2014).
18. G. Shi, Y. Jin, Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res. Ther.* **1**, 1–9 (2010).
19. E. P. Papapetrou, M. J. Tomishima, S. M. Chambers, Y. Mica, E. Reed, J. Menon, V. Tabar, Q. Mo, L. Studer, M. Sadelain, Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12759–12764 (2009).
20. A. Radzisheuskaya, G. Le Bin Chia, R. L. Dos Santos, T. W. Theunissen, L. F. C. Castro, J. Nichols, J. C. Silva, A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* **15**, 579–590 (2013).
21. F. Hammachi, G. M. Morrison, A. A. Sharov, A. Livigni, S. Narayan, E. P. Papapetrou, J. O'Malley, K. Kaji, M. S. Ko, M. Ptashne, J. M. Brickman, Transcriptional activation by oct4 is sufficient for the maintenance and induction of pluripotency. *Cell Rep.* **1**, 99–109 (2012).
22. S. Velychko, K. Adachi, K. P. Kim, Y. Hou, C. M. MacCarthy, G. Wu, H. R. Schöler, Excluding Oct4 from Yamanaka cocktail unleashes the developmental potential of iPSCs. *Cell Stem Cell* **25**, 737–753.e4 (2019).
23. G. Wu, D. Han, Y. Gong, V. Sebastiano, L. Gentile, N. Singhal, K. Adachi, G. Fischedick, C. Ortmeier, M. Sinn, M. Radstaak, A. Tomilin, H. R. Schöler, Establishment of totipotency does not depend on Oct4A. *Nat. Cell Biol.* **15**, 1089–1097 (2013).
24. D. C. Hay, L. Sutherland, J. Clark, T. Burdon, Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* **22**, 225–235 (2004).
25. H. Niwa, J.-I. Miyazaki, A. G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376 (2000).
26. J. L. Kopp, B. D. Ormsbee, M. Desler, A. Rizzino, Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26**, 903–911 (2008).
27. B. W. Carey, S. Markoulaki, J. H. Hanna, D. A. Faddah, Y. Buganim, J. Kim, K. Ganz, E. J. Steine, J. P. Cassady, M. P. Creyghton, G. G. Welstead, Q. Gao, R. Jaenisch, Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell* **9**, 588–598 (2011).
28. G. Nagamatsu, S. Saito, T. Kosaka, K. Takubo, T. Kinoshita, M. Oya, K. Horimoto, T. Suda, Optimal ratio of transcription factors for somatic cell reprogramming. *J. Biol. Chem.* **287**, 36273–36282 (2012).
29. D. Sui, Z. Sun, C. Xu, Y. Wu, M. R. Capecchi, S. Wu, N. Li, Fine-tuning of iPSC derivation by an inducible reprogramming system at the protein level. *Stem Cell Rep.* **2**, 721–733 (2014).
30. U. Tiemann, M. Sgodda, E. Warlich, M. Ballmaier, H. R. Schöler, A. Schambach, T. Cantz, Optimal reprogramming factor stoichiometry increases colony numbers and affects molecular characteristics of murine induced pluripotent stem cells. *Cytometry A* **79**, 426–435 (2011).
31. T. Brambrink, R. Foreman, G. G. Welstead, C. J. Lengner, M. Wernig, H. Suh, R. Jaenisch, Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151–159 (2008).
32. P. D. Tonge, A. J. Corso, C. Monetti, S. M. Hussein, M. C. Puri, I. P. Michael, M. Li, D. S. Lee, J. C. Mar, N. Cloonan, D. L. Wood, M. E. Gauthier, O. Korn, J. L. Clancy, T. Preiss, S. M. Grimmond, J. Y. Shin, J. S. Seo, C. A. Wells, I. M. Rogers, A. Nagy, Divergent reprogramming routes lead to alternative stem-cell states. *Nature* **516**, 192–197 (2014).
33. M. Francesconi, B. Di Stefano, C. Berenguer, L. de André-Aguayo, M. Plana-Carmona, M. Mendez-Lago, A. Guillaumet-Adkins, G. Rodriguez-Esteban, M. Gut, I. G. Gut, H. Heyn, B. Lehner, T. Graf, Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming. *eLife* **8**, e41627 (2019).
34. K. A. Tran, S. J. Pietrzak, N. Z. Zaidan, A. F. Siahpirani, S. G. McCalla, A. S. Zhou, G. Iyer, S. Roy, R. Sridharan, Defining reprogramming checkpoints from single-cell analyses of induced pluripotency. *Cell Rep.* **27**, 1726–1741.e5 (2019).
35. V. Malik, L. V. Glaser, D. Zimmer, S. Velychko, M. Weng, M. Holzner, M. Arend, Y. Chen, Y. Srivastava, V. Veerapandian, Z. Shah, M. A. Esteban, H. Wang, J. Chen, H. R. Schöler, A. P. Hutchins, S. H. Meijsing, S. Pott, R. Jauch, Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. *Nat. Commun.* **10**, 3477 (2019).
36. G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, E. S. Lander, Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943.e22 (2019).
37. E. Lujan, E. R. Zunder, Y. H. Ng, I. N. Goronzy, G. P. Nolan, M. Wernig, Early reprogramming regulators identified by prospective isolation and mass cytometry. *Nature* **521**, 352–356 (2015).
38. J. Elbaz, M. C. Puri, M. Faiz, K. W. Bang, L. Nguyen, B. Makovoz, M. Gertsenstein, S. M. Hussein, P. W. Zandstra, L. Briollais, N. Shakiba, A. Nagy, Highly efficient reprogrammable mouse lines with integrated reporters to track the route to pluripotency. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2207824119 (2022).
39. K. Tanabe, M. Nakamura, M. Narita, K. Takahashi, S. Yamanaka, Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12172–12179 (2013).
40. E. R. Zunder, E. Lujan, Y. Goltsev, M. Wernig, G. P. Nolan, A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* **16**, 323–337 (2015).
41. M. Stadtfeld, N. Maherali, D. T. Breault, K. Hochedlinger, Defining molecular cornerstones during fibroblast to ips cell reprogramming in mouse. *Cell Stem Cell* **2**, 230–240 (2008).
42. Y. Buganim, D. A. Faddah, A. W. Cheng, E. Itskovich, S. Markoulaki, K. Ganz, S. L. Klemm, A. V. Oudenaarden, R. Jaenisch, Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
43. L. V. Greder, S. Gupta, S. Li, M. J. Abedin, A. Sajini, Y. Segal, J. M. Slack, J. R. Dutton, Brief report: Analysis of endogenous Oct4 activation during induced pluripotent stem cell reprogramming using an inducible Oct4 lineage label. *Stem Cells* **30**, 2596–2601 (2012).
44. K.-P. Kim, Y. Wu, J. Yoon, K. Adachi, G. Wu, S. Velychko, C. M. MacCarthy, B. Shin, A. Röpke, M. J. Arauzo- Bravo, M. Stehling, D. W. Han, Y. Gao, J. Kim, S. Gao, H. R. Schöler, Reprogramming competence of OCT factors is determined by transactivation domains. *Sci. Adv.* **6**, eaaz7364 (2020).
45. C. Vogel, E. M. Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
46. D. Del Vecchio, R. M. Murray, *Biomolecular Feedback Systems* (Princeton Univ. Press, 2015).
47. D. Del Vecchio, H. Abdallah, Y. Qian, J. J. Collins, A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Systems* **4**, 109–120.e11 (2017).
48. X. Bofill-De Ros, S. Gu, Guidelines for the optimal design of miRNA-based shRNAs. *Methods* **103**, 157–166 (2016).
49. R. T. Rodriguez, J. M. Velkey, C. Lutzko, R. Seerke, D. B. Kohn, K. S. O'Shea, M. T. Firpo, Manipulation of OCT4 levels in human embryonic stem cells results in induction of differential cell types. *Exp. Biol. Med.* **232**, 1368–1380 (2007).
50. A. L. Szymczak, C. J. Workman, Y. Wang, K. M. Vignali, S. Dilioglou, E. F. Vanin, D. A. Vignali, Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22**, 589–594 (2004).
51. M. Q. Liu, M. Zhao, W. H. Kong, J. S. Peng, F. Wang, H. Y. Qiu, Z. R. Zhu, L. Tang, M. Sang, J. G. Wu, W. Z. Ho, W. Zhou, Antiretroviral therapy fails to restore levels of HIV-1 restriction miRNAs in PBMCs of HIV-1-infected MSM. *Medicine* **94**, e2116 (2015).
52. N. Plachta, T. Bollenbach, S. Pease, S. E. Fraser, P. Pantazis, Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat. Cell Biol.* **13**, 117–123 (2011).

53. C. J. Lengner, F. D. Camargo, K. Hochedlinger, G. G. Welstead, S. Zaidi, S. Gokhale, H. R. Scholer, A. Tomilin, R. Jaenisch, Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell* **1**, 403–415 (2007).
54. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. *PLOS Biol.* **4**, 1707–1719 (2006).
55. A. Ochab-Marcinek, M. Tabaka, Bimodal gene expression in noncooperative regulatory systems. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22096–22101 (2010).
56. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
57. R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, L. S. Weinberger, Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17454–17459 (2012).
58. J. J. Gam, B. DiAndreth, R. D. Jones, J. Huh, R. Weiss, A 'poly-transfection' method for rapid, one-pot characterization and optimization of genetic systems. *Nucleic Acids Res.* **47**, e106 (2019).
59. S. Guo, X. Zi, V. P. Schulz, J. Cheng, M. Zhong, S. H. Koochaki, C. M. Megyola, X. Pan, K. Heydari, S. M. Weissman, P. G. Gallagher, D. S. Krause, R. Fan, J. Lu, Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156**, 649–662 (2014).
60. X. Hu, Q. Wu, J. Zhang, J. Kim, X. Chen, A. A. Hartman, A. E. Eastman, I. H. Park, S. Guo, Reprogramming progressive cells display low CAG promoter activity. *Stem Cells* **39**, 43–54 (2021).
61. Y. S. Michaels, M. B. Barnkob, H. Barbosa, T. A. Baeumler, M. K. Thompson, V. Andre, H. Colin-York, M. Fritzsche, U. Gileadi, H. M. Sheppard, D. J. Knapp, T. A. Milne, V. Cerundolo, T. A. Fulga, Precise tuning of gene expression levels in mammalian cells. *Nat. Commun.* **10**, 818 (2019).
62. A. M. Chiarella, K. V. Butler, B. E. Gryder, D. Lu, T. A. Wang, X. Yu, S. Pomella, J. Khan, J. Jin, N. A. Hathaway, Dose-dependent activation of gene expression is achieved using CRISPR and small molecules that recruit endogenous chromatin machinery. *Nat. Res.* **38**, 50–55 (2020).
63. M. Jost, D. A. Santos, R. A. Saunders, M. A. Horlbeck, J. S. Hawkins, S. M. Scaria, T. M. Norman, J. A. Hussmann, C. R. Liem, C. A. Gross, J. S. Weissman, Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* **38**, 355–364 (2020).
64. R. V. Desai, X. Chen, B. Martin, S. Chaturvedi, D. W. Hwang, W. Li, C. Yu, S. Ding, M. Thomson, R. H. Singer, R. A. Coleman, M. M. Hansen, L. S. Weinberger, A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science* **373**, eabc6506 (2021).
65. A. Pfeifer, M. Ikawa, Y. Dayn, I. M. Verma, Transgenesis by lentiviral vectors: Lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2140–2145 (2002).
66. L. Warren, C. Lin, mRNA-based genetic reprogramming. *Mol. Ther.* **27**, 729–734 (2019).
67. J. M. Polo, E. Anderssen, R. M. Walsh, B. A. Schwarz, C. M. Nefzger, S. M. Lim, M. Borkent, E. Apostolou, S. Alaei, J. Cloutier, O. Bar-Nur, S. Cheloufi, M. Stadtfeld, M. E. Figueroa, D. Robinton, S. Natesan, A. Melnick, J. Zhu, S. Ramaswamy, K. Hochedlinger, A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
68. N. Maherali, T. Ahfeldt, A. Rigamonti, J. Utikal, C. Cowan, K. Hochedlinger, A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3**, 340–345 (2008).
69. A. Somers, J. C. Jean, C. A. Sommer, A. Omari, C. C. Ford, J. A. Mills, L. Ying, A. G. Sommer, J. M. Jean, B. W. Smith, R. Lafyatis, M. F. Demierre, D. J. Weiss, D. L. French, P. Gadue, G. J. Murphy, G. Mostoslavsky, D. N. Kotton, Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells* **28**, 1728–1740 (2010).
70. D. Hockemeyer, F. Soldner, E. G. Cook, Q. Gao, M. Mitalipova, R. Jaenisch, A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**, 346–353 (2008).
71. R. Jones, Y. Qian, V. Siciliano, B. DiAndreth, J. Huh, R. Weiss, D. D. Vecchio, An endoribonuclease- based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat. Commun.* **11**, 5690 (2020).
72. M. Li, J. C. I. Belmonte, Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.* **20**, 382–392 (2018).
73. K. Nishimura, T. Kato, C. Chen, L. Oinam, E. Shiomitsu, D. Ayakawa, M. Ohtaka, A. Fukuda, M. Nakanishi, K. Hisatake, Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming. *Stem Cell Rep.* **3**, 915–929 (2014).
74. S. I. Kim, F. Oceguera-Yanez, R. Hirohata, S. Linker, K. Okita, Y. Yamada, T. Yamamoto, S. Yamanaka, K. Woltjen, KLF4 N-terminal variance modulates induced reprogramming to pluripotency. *Stem Cell Rep.* **4**, 727–743 (2015).
75. D. Strebinger, C. Deluz, E. T. Friman, S. Govindan, A. B. Alber, D. M. Suter, Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions. *Mol. Syst. Biol.* **15**, e9002 (2019).
76. Y. Gao, J. Chen, K. Li, T. Wu, B. Huang, W. Liu, X. Kou, Y. Zhang, H. Huang, Y. Jiang, C. Yao, X. Liu, Z. Lu, Z. Xu, L. Kang, J. Chen, H. Wang, T. Cai, S. Gao, Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
77. J. Chen, L. Guo, L. Zhang, H. Wu, J. Yang, H. Liu, X. Wang, X. Hu, T. Gu, Z. Zhou, J. Liu, J. Liu, H. Wu, S. Q. Mao, K. Mo, Y. Li, K. Lai, J. Qi, H. Yao, G. Pan, G. L. Xu, D. Pei, Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* **45**, 1504–1509 (2013).
78. J. Shu, C. Wu, Y. Wu, Z. Li, S. Shao, W. Zhao, X. Tang, H. Yang, L. Shen, X. Zuo, W. Yang, Y. Shi, X. Chi, H. Zhang, G. Gao, Y. Shu, K. Yuan, W. He, C. Tang, Y. Zhao, H. Deng, Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963–975 (2013).
79. N. Montserrat, E. Nivet, I. Sancho-Martinez, T. Hishida, S. Kumar, L. Miquel, C. Cortina, Y. Hishida, Y. Xia, C. R. Esteban, J. C. I. Belmonte, Reprogramming of human fibroblasts to pluripotency with lineage specifiers. *Cell Stem Cell* **13**, 341–350 (2013).
80. Y. Rais, A. Zviran, S. Geula, O. Gafni, E. Chomsky, S. Viukov, A. A. Mansour, I. Caspi, V. Krupalnik, M. Zerbib, I. Maza, N. Mor, D. Baran, L. Weinberger, D. A. Jaitin, D. Lara-Astiaso, R. Blecher-Gonen, Z. Shipony, Z. Mukamel, T. Hagai, S. Gilad, D. Amann-Zalcenstein, A. Tanay, I. Amit, N. Novershtern, J. H. Hanna, Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
81. P. W. Hill, R. Amouroux, P. Hajkova, DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: An emerging complex story. *Genomics* **104**, 324–333 (2014).
82. T. A. Hore, F. V. Meyenn, M. Ravichandran, M. Bachman, G. Ficz, D. Oxley, F. Santos, S. Balasubramanian, T. P. Jurkowski, W. Reik, Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12202–12207 (2016).
83. S. Jaffer, P. Goh, M. Abbasian, A. C. Nathwani, Mbd3 promotes reprogramming of primary human fibroblasts. *Int. J. Stem Cells* **11**, 235–241 (2018).
84. K. Adachi, W. Kopp, G. Wu, S. Heising, B. Greber, M. Stehling, M. J. Araúzo-Bravo, S. T. Boerno, B. Timmermann, M. Vingron, H. R. Schöler, Esrrb unlocks silenced enhancers for reprogramming to naive pluripotency. *Cell Stem Cell* **23**, 266–275.e6 (2018).
85. N. Mor, Y. Rais, D. Sheban, S. Peles, A. Aguilera-Castrejon, A. Zviran, D. Elinger, S. Viukov, S. Geula, V. Krupalnik, M. Zerbib, E. Chomsky, L. Lasman, T. Shani, J. Bayerl, O. Gafni, S. Hanna, J. D. Buenrostro, T. Hagai, H. Masika, G. Vainorius, Y. Bergman, W. J. Greenleaf, M. A. Esteban, U. Elling, Y. Levin, R. Massarwa, Y. Merbl, N. Novershtern, J. H. Hanna, Neutralizing Gatad2a-Chd4-Mbd3/NuRD complex facilitates deterministic induction of naive pluripotency. *Cell Stem Cell* **23**, 412–425.e10 (2018).
86. G. J. Markov, T. Mai, S. Nair, A. Shcherbina, Y. X. Wang, D. M. Burns, A. Kundaje, H. M. Blau, Ap-1 is a temporally regulated dual gatekeeper of reprogramming to pluripotency. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104841118 (2021).
87. H. Park, B. Cho, J. Kim, Rad50 mediates DNA demethylation to establish pluripotent reprogramming. *Exp. Mol. Med.* **52**, 1116–1127 (2020).
88. R. Navarro, L. C. Chen, R. Rakhit, T. J. Wandless, A novel destabilizing domain based on a small-molecule dependent fluorophore. *ACS Chem. Biol.* **11**, 2101–2104 (2016).
89. Y. Miyazaki, H. Imoto, L. C. Chen, T. J. Wandless, Destabilizing domains derived from the human estrogen receptor. *J. Am. Chem. Soc.* **134**, 3942–3945 (2012).
90. J. A. Casas-Mollano, M. H. Zinselmeier, S. E. Erickson, M. J. Smanski, CRISPR-Cas activators for engineering gene expression in higher eukaryotes. *CRISPR J.* **3**, 350–364 (2020).
91. R. Morita, M. Suzuki, H. Kasahara, N. Shimizu, T. Shichita, T. Sekiya, A. Kimura, K. I. Sasaki, H. Yasukawa, A. Yoshimura, ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 160–165 (2015).
92. H. Xie, M. Ye, R. Feng, T. Graf, Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
93. L. Warren, P. D. Manos, T. Ahfeldt, Y. H. Loh, H. Li, F. Lau, W. Ebina, P. K. Mandal, Z. D. Smith, A. Meissner, G. Q. Daley, A. S. Brack, J. J. Collins, C. Cowan, T. M. Schlaeger, D. J. Rossi, Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).
94. S. Liedtke, J. Enczmann, S. Waclawczyk, P. Wernet, G. Kögler, Oct4 and its pseudogenes confuse stem cell research. *Cell Stem Cell* **1**, 364–366 (2007).
95. Z. Redshaw, A. J. Strain, Human haematopoietic stem cells express Oct4 pseudogenes and lack the ability to initiate Oct4 promoter-driven gene expression. *J. Negat. Results Biomed.* **9**, 2 (2010).
96. X. Duportet, L. Wroblewska, P. Guye, Y. Li, J. Eyquem, J. Rieders, T. Rimchala, G. Batt, R. Weiss, A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42**, 13440–13451 (2014).

97. N. W. Meza, A. Puyet, S. Pérez-Benavente, O. Quintana-Bustamante, A. Diez, J. A. Bueren, J. C. Segovia, J. M. Bautista, Functional analysis of gammaretroviral vector transduction by quantitative PCR. *J. Gene Med.* **8**, 1097–1104 (2006).
98. B. Fehse, O. S. Kustikova, M. Bubenheim, C. Baum, Pois(s)on—It's a question of dose. *Gene Ther.* **11**, 879–881 (2004).
99. S. Charrier, M. Ferrand, M. Zerbato, G. Précigout, A. Viornery, S. Bucher-Laurent, S. Benkhelifa-Ziyyat, O. W. Merten, J. Perea, A. Galy, Quantification of lentiviral vector copy numbers in individual hematopoietic colony-forming cells shows vector dose-dependent effects on the frequency and level of transduction. *Gene Ther.* **18**, 479–487 (2011).
100. J. R. Lin, M. Fallahi-Sichani, P. K. Sorger, Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat. Commun.* **6**, 8390 (2015).
101. A. Gerrits, B. Dykstra, O. J. Kalmykowa, K. Klauke, E. Verovskaya, M. J. Broekhuis, G. De Haan, L. V. Bystrykh, Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010).
102. D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
103. N. Zapata-Linares, S. Rodriguez, E. Salido, G. Abizanda, E. Iglesias, F. Prosper, G. Gonzalez-Aseguinolaza, J. R. Rodriguez-Madoz, Generation and characterization of human iPSC lines derived from a Primary Hyperoxaluria Type I patient with p.I244T mutation.. *Stem Cell Res..* **16**, 116–119 (2016).
103. N. T. Swaidan, S. Salloum-Asfar, F. Palangi, K. Errafii, N. H. Soliman, A. T. Aboughalia, A. H. S. Wali, S. A. Abdulla, M. M. Emara, Identification of potential transcription factors that enhance human iPSC generation. *Sci. Rep.* **10**, 21950 (2020).

Acknowledgments: We are grateful to J. Wrana for generously sharing his lineage tracking barcode library. We also thank the flow cytometry core at the Koch Institute's Robert A. Swanson (1969) Biotechnology Center for technical support. We thank R. V. Gayet for helpful comments on the manuscript. **Funding:** This work was funded by NIH NIBIB award number 5R01EB024591 (D.D.V., T.M.S., J.J.C., and R.W.) and MIT Amar G. Bose Research Grant (D.D.V. and R.W.). **Author contributions:** Conceptualization: K.I., N.S., T.B., R.D.J., J.J.C., R.W., D.D.V., and T.M.S. Methodology: K.I., N.S., T.B., R.D.J., M.M.K., E.A., S.B., J.J.C., R.W., D.D.V., and T.M.S. Software: K.I., N.S., T.B., T.D.J., and S.B.. Validation: K.I., N.S., T.B., R.D.J., and T.M.S. Formal analysis: K.I., N.S., T.B., R.D.J., and T.M.S. Investigation: K.I., N.S., T.B., R.D.J., M.M.K., S.P., and T.M.S. Resources: J.J.C., R.W., D.D.V., and T.M.S. Data curation: K.I., N.S., T.B., R.D.J., and T.M.S. Writing (original draft): K.I.. Writing (review and editing): K.I., N.S., T.B., R.D.J., S.B., J.J.C., R.W., D.D.V., and T.M.S. Visualization: K.I. and T.M.S. Supervision: J.J.C., D.D.V., and T.M.S. Project administration: J.J.C., D.D.V., and T.M.S. Funding acquisition: J.J.C., R.W., D.D.V., and T.M.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Plasmids will be made available to the scientific community through Addgene.

Submitted 26 January 2023
Accepted 27 October 2023
Published 29 November 2023
10.1126/sciadv.adg8495

ScienceAdvances

Synthetic genetic circuits to uncover the OCT4 trajectories of successful reprogramming of human fibroblasts

Katherine Ilia, Nika Shakiba, Trevor Bingham, Ross D. Jones, Michael M. Kaminski, Eliezer Aravera, Simone Bruno, Sebastian Palacios, Ron Weiss, James J. Collins, Domitilla Del Vecchio, and Thorsten M. Schlaeger

Sci. Adv. **9** (48), eadg8495. DOI: 10.1126/sciadv.adg8495