

Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors

Itay Maza^{1,2}, Inbal Caspi¹, Asaf Zviran¹, Elad Chomsky¹, Yoach Rais¹, Sergey Viukov¹, Shay Geula¹, Jason D Buenrostro³, Leehee Weinberger¹, Vladislav Krupalnik¹, Suhair Hanna^{1,4}, Mirie Zerbib¹, James R Dutton⁵, William J Greenleaf³, Rada Massarwa¹, Noa Novershtern¹ & Jacob H Hanna¹

Somatic cells can be transdifferentiated to other cell types without passing through a pluripotent state by ectopic expression of appropriate transcription factors^{1,2}. Recent reports have proposed an alternative transdifferentiation method in which fibroblasts are directly converted to various mature somatic cell types by brief expression of the induced pluripotent stem cell (iPSC) reprogramming factors *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) followed by cell expansion in media that promote lineage differentiation^{3–6}. Here we test this method using genetic lineage tracing for expression of endogenous *Nanog* and *Oct4* and for X chromosome reactivation, as these events mark acquisition of pluripotency. We show that the vast majority of reprogrammed cardiomyocytes or neural stem cells obtained from mouse fibroblasts by OSKM-induced ‘transdifferentiation’ pass through a transient pluripotent state, and that their derivation is molecularly coupled to iPSC formation mechanisms. Our findings underscore the importance of defining trajectories during cell reprogramming by various methods.

Somatic cell transdifferentiation involves ectopic expression of lineage master regulators that induce conversion into a different somatic cell type without passing through a pluripotent configuration. For example, expression of *C/EBP α* converts Pro-B cells into macrophage-like cells^{1,7}. Recently, a new approach to somatic transdifferentiation, called OSKM-mediated transdifferentiation (OSKM-TD), has been described in which Yamanaka’s four original pluripotency reprogramming factors⁸ are briefly expressed for periods as short as 3–10 d to induce an intermediate, partially reprogrammed and presumably nonpluripotent ‘plastic’ state^{3–6}. Next, lineage-specifying media that lack conventional pluripotency-promoting cytokines, such as leukemia inhibitory factor (LIF), are provided to shift these intermediate cells toward a desired somatic cell fate without their ever becoming pluripotent^{3–6}. The authors of these studies supported their conclusion that OSKM-TD circumvents pluripotency in several ways^{3–5}. Brief OSKM induction of <10 d was deemed insufficient to yield iPSCs. Culture conditions,

particularly the absence of LIF and the presence of JAK1 small-molecule inhibitors to block Stat3 signaling, were designed to prevent acquisition of pluripotency. However, lineage-tracing tools that could unequivocally determine whether the cells attained pluripotency were not used. Thus, it remains unclear whether somatic cells produced by OSKM-TD transdifferentiate or, alternatively, go through a transient state of induced pluripotency and then differentiate to a somatic lineage according to the media conditions applied. Addressing this question is fundamental to understanding mechanisms of cellular reprogramming and to evaluating the safety of cells reprogrammed by this technique in potential clinical applications.

Our interest in this issue arose from our observation that *Nanog*-GFP⁺ iPSCs appear at low efficiency during reprogramming with different doxycycline (Dox)-inducible OSKM transgenic systems^{2,9,10} after as few as 3 d of Dox induction in conditions of 15% FBS, 5% KSR and LIF (leukemia inhibitory factor) (Fig. 1a). Furthermore, when we induced OSKM transgene expression^{2,9,10} with Dox (doxycycline) in *Oct4*-GFP secondary reporter fibroblast cells using cardiogenic or neural stem cell (NSC) growth conditions instead of conventional LIF-containing pluripotency conditions, we obtained GFP⁺ embryonic stem cell (ESC)-like colonies and ‘hybrid’ colonies with *Oct4*-GFP⁺ cells in the center and, on the edges, no *Oct4*-GFP⁺ cells and clear signs of neuronal differentiation (Supplementary Fig. 1a–c). These results emphasized the need to exclude the possibility that iPSCs may form rapidly under suboptimal reprogramming conditions and may be a source of ‘transdifferentiated’ cells generated by OSKM-TD approaches^{3,4}.

We next engineered a system to track transient acquisition of pluripotency during *in vitro* reprogramming. Authors of previous OSKM-TD studies argued that the lack of *Nanog* reactivation in the bulk of reprogramming cultures^{3,4} proved that the cells did not become pluripotent. Thus, we designed a genetic tracing system for reactivation of endogenous *Nanog*, which occurs at the late stages of iPSC formation^{3,4,11,12}. In our system, transient activation of *Nanog* is monitored by permanent tdTomato fluorescence, indicating that pluripotency was achieved, even if only temporarily. For this purpose

¹The Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. ²The Department of Gastroenterology, Rambam Health Care Campus & Bruce Rappaport School of Medicine, Technion Institute of Technology, Haifa, Israel. ³Department of Genetics, Stanford University School of Medicine, Stanford, California, USA. ⁴The Department of Pediatrics and the Pediatric Immunology Unit, Rambam Health Care Campus & Bruce Rappaport School of Medicine, Technion Institute of Technology, Haifa, Israel. ⁵Stem Cell Institute, University of Minnesota, Minneapolis, Minnesota, USA. Correspondence should be addressed to J.H.H. (jacob.hanna@weizmann.ac.il), R.M. (rada.massarwa@weizmann.ac.il) or N.N. (noa.novershtern@weizmann.ac.il).

Received 27 June 2014; accepted 1 June 2015; published online 22 June 2015; doi:10.1038/nbt.3270

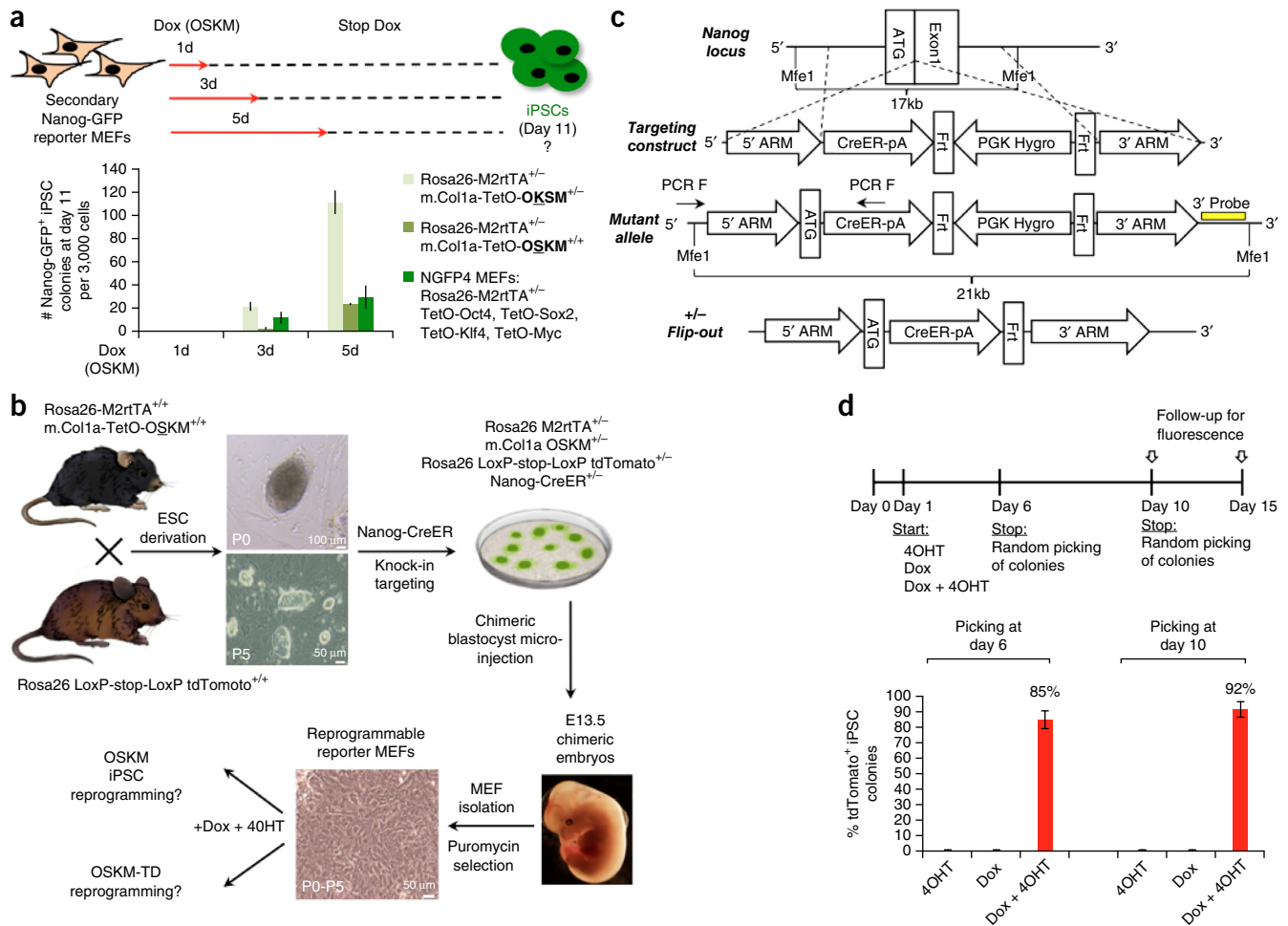


Figure 1 Lineage tracing for endogenous *Nanog* reactivation during reprogramming. **(a)** MEFs from three different secondary reprogramming systems, all carrying a *Nanog*-GFP knock-in reporter for pluripotency, were subjected to Dox-induced reprogramming. Dox was applied for the indicated times and then withdrawn. iPSC formation was evaluated at day 11 without passaging. Error bars, s.e.m. \pm mean of biological triplicates (one out of two representative experiments is shown). **(b)** Scheme illustrating generation of quadruple knock-in-allele reporter in reprogrammable MEFs, used for either OSKM-iPSC or OSKM-TD reprogramming. **(c)** *Nanog*-CreER knock-in targeting strategy. **(d)** Reprogrammable *Nanog*-CreER MEFs were subjected to iPSC reprogramming in the presence of Dox, 4OHT or both, which were withdrawn at day 6 or 10 as indicated. 48 colonies were randomly subcloned and validated as iPSCs (*Nanog*, *SSEA1* and *AP* staining) from each condition and time point, and then scored for the presence of *tdTomato*⁺ signal. Bar plot showing percentage of *tdTomato*⁺ iPSCs clones obtained indicates the sensitivity of the system. Error bars, s.e.m. \pm mean of biological duplicates (one out of three representative experiments is shown).

we first mated the *Rosa26 M2rtTA/m.Col1a OSKM*^{+/+} reprogrammable mouse strain⁹ with *Rosa26 LoxP-stop-LoxP tdTomato*^{+/+} reporter mice and derived and validated mouse ESC (mESC) lines from embryonic day E3.5 blastocysts that are heterozygous for the triple knock-in alleles: *Rosa26 M2rtTA*^{+/-}, *Rosa26 LoxP-Stop-LoxP tdTomato*^{+/-} and *m.Col1a-TetO-OSKM*^{+/-} (**Fig. 1b**). Next, we generated a bacterial artificial chromosome (BAC)-recombineered *Nanog*-CreER targeting construct, which introduces a knock-in tamoxifen-inducible CreER cassette under the control of the endogenous *Nanog* promoter (**Fig. 1c** and **Supplementary Fig. 2a**). The quadruple knock-in mESCs were microinjected into host blastocysts to generate chimeric animals from which transgenic fibroblasts were extracted for experimental analysis (**Fig. 1b**).

We validated the sensitivity and specificity of this system. mESCs reactivated *tdTomato* rapidly after 48 h of 4-hydroxytamoxifen (4OHT) induction (95–100% of ESC colonies), and no fluorescence was detected in mESCs passaged over extended periods of time in the absence of 4OHT (**Supplementary Fig. 2b**). NCAM⁺-sorted cells

from early embryoid bodies did not activate the *tdTomato* reporter after 4OHT treatment indicating that *Nanog*-CreER was not active in nascent neuroectoderm or mesoderm (**Supplementary Fig. 3**). Mouse embryonic fibroblasts (MEFs) expanded with or without Dox in the absence of 4OHT did not show any reactivation of *tdTomato*, further indicating the specificity of the system and excluding leakiness during expansion or reprogramming. The fidelity of the reporter was evaluated by reprogramming the transgenic MEFs to pluripotency, by adding Dox and 4OHT to the medium and withdrawing both molecules at days 6 and 10. Adding 4OHT alone did not produce any colonies whereas adding Dox alone produced mESC-like colonies, all of which were *tdTomato* negative (**Fig. 1d** and **Supplementary Fig. 2c**). mESC-like colonies formed in each of the conditions at days 6 and 10 (48 colonies for each group) were randomly subcloned and grown for another 5–6 d, and the presence of *tdTomato*⁺ clones was assessed as a percentage of the total number of developed and validated Dox-independent iPSC clones. We found the fidelity of our system to be ~85% and ~92% for iPSCs obtained after 6 and 10 d of Dox + 4OHT

induction, respectively (Fig. 1d). The <100% sensitivity might be explained by limitations of the Nanog knock-in allele activation, biallelic expression of Nanog, or a reflection of a true biological outcome, given that Nanog activity can be dispensable during iPSC generation¹². Nevertheless, these results validate the ability of the system to document acquisition of pluripotency during iPSC reprogramming with high specificity.

Next, we applied the Nanog tracing system and reporter-engineered MEFs to previously reported OSKM-TD experiments on converting mouse fibroblasts into NSCs³ (TD-NSC) and cardiomyocytes

(TD-CM)⁴. The published protocols were meticulously followed (Fig. 2a), including the use of conditions without LIF or feeder cells.

The TD-NSC protocol³ generated Sox1⁺ NSC colonies that could be subcloned and expanded as stable NSC lines (Fig. 2b and Supplementary Fig. 4), as previously shown. RNA-seq analysis confirmed the transcriptional similarity of TD-NSCs to NSCs derived from primary embryos, and their difference from MEFs, ESCs and iPSCs (Supplementary Fig. 5a). Notably, ~82% of Sox1⁺ TD-NSC colonies were tdTomato⁺, indicating acquisition of pluripotency during their conversion from MEFs (Fig. 2c and Supplementary Figs. 5b, 6 and 7a,b).

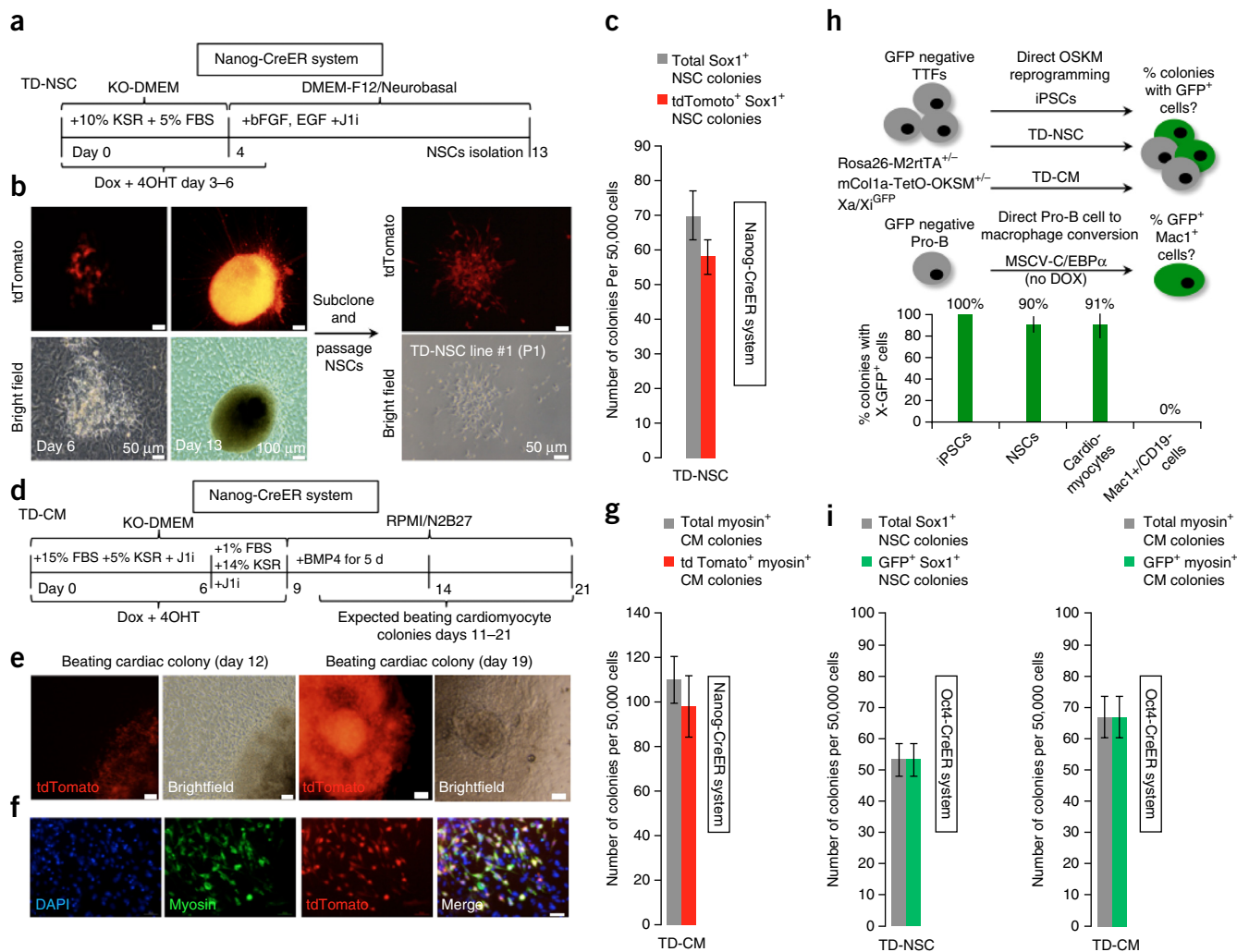


Figure 2 Frequent and transient acquisition of pluripotency during OSKM-TD. **(a)** Schematic illustration of NSC transdifferentiation with indicated time points for medium switch, small molecule and growth factor additions. **(b)** Left: representative images of NSC tdTomato⁺ colonies on the primary plate generated by protocol in **a**, at the start and end points of neural reprogramming. Right: passage of subcloned NSC colony grown in NSC medium, showing typical NSCs morphology and tdTomato⁺ fluorescence. Scale bars, 100 μ m for day 13, 50 μ m for all others. **(c)** Bar plot of immunostained Sox1⁺ TD-NSC colonies, showing the number of total and of tdTomato⁺ colonies obtained. Error bars, s.e.m. \pm mean of biological triplicate wells (one out of two representative experiments is shown). **(d)** Schematic illustration of TD-CM with indicated time points for medium switch, small molecules and growth factors additions. **(e)** Representative images of cardiomyocyte colonies reprogrammed by protocol in **d**, with tdTomato⁺ fluorescence, and repeated rhythmic beating motion after 12 or 19 d of reprogramming. Scale bars, 100 μ m. **(f)** Representative immunostaining for myosin in TD-CMs with overlay of tdTomato fluorescence. Scale bar, 50 μ m. **(g)** Bar plot quantifying validated immunostained myosin⁺ TD-CM colonies, showing total number of tdTomato⁺ colonies obtained. Error bars, s.e.m. \pm mean of biological triplicate wells (one out of two representative experiments is shown). **(h)** Tail tip fibroblasts (TTFs) were established from the indicated adult transgenic reprogrammable female mice. GFP negative populations were sorted, purified and subjected to OSKM iPSCs or TD protocols. Pro-B cells were purified from the same mice, and subjected to C/EBP α -mediated conversion to macrophage-like cells (without OSKM induction). Percentage of GFP⁺ cells or GFP⁺ cell containing colonies, detected at day 12–13 (for OSKM reprogramming) and day 8 (for Pro-B to macrophage conversion), is indicated. Error bars, s.e.m. \pm mean of biological duplicates (one out of three representative experiments is shown). **(i)** Bar plot of immunostained Sox1⁺ TD-NSC colonies and myosin⁺ TD-CM colonies derived from Oct4-CreER mT/mG reporter MEFs, showing the number of total colonies and of GFP⁺ colonies obtained. Error bars, s.e.m. \pm mean of biological triplicate wells (one out of two representative experiments is shown).

Similarly, the TD-CM protocol (Fig. 2d) yielded TD-CMs that expressed early and late cardiomyocytes markers (myosin, cardiac troponin T), many of which showed beating rhythmicity (Fig. 2e,f and Supplementary Fig. 5b–c and Supplementary Videos 1 and 2). Approximately 93% of myosin⁺ TD-CM colonies, and within each colony the majority of cells, were tdTomato⁺, indicating transient acquisition of pluripotency (Fig. 2e–g and Supplementary Figs. 5b,c and 7b,c).

It is well established that an additional late and specific hallmark of acquisition of naive pluripotency during reprogramming to iPSCs is reactivation of the silenced X chromosome. Subsequent differentiation of iPSCs involves random X inactivation¹³. To determine whether X chromosome reactivation occurs in OSKM-TD, we crossed female *Rosa26 M2rtTA^{+/+}/m.Col1a-TetO-OSKM^{+/+}* mice⁹ with male mice carrying

an X-linked GFP reporter transgene¹⁴, and purified GFP⁻ tail tip fibroblasts from female offspring; all GFP⁻ reprogrammable cells carried the GFP transgene on the inactive X chromosome (Fig. 2h). Notably, GFP fluorescence was identified in >90% of TD-NSC and TD-CM colonies (Fig. 2h), and in 100% of iPSC colonies obtained. Reprogramming of female-derived XGFP⁻ Pro-B lymphocytes into macrophage-like cells by *C/EBPα* overexpression¹⁵ did not yield any GFP⁺/Mac1⁺ colonies (Fig. 2h), thus excluding promiscuous, nonspecific X-chromosome reactivation in the transdifferentiation protocols applied.

Finally, we used MEFs derived from mice carrying an Oct4-CreER knock-in reporter and a *Rosa26 mTomato/mGFP (mT/mG)* reporter cassette; in these cells mTomato expression is switched to EGFP expression upon Oct4 locus reactivation in the presence of 4OHT (Supplementary Fig. 8a)¹⁶. Repeating OSKM-TD protocols

© 2015 Nature America, Inc. All rights reserved.

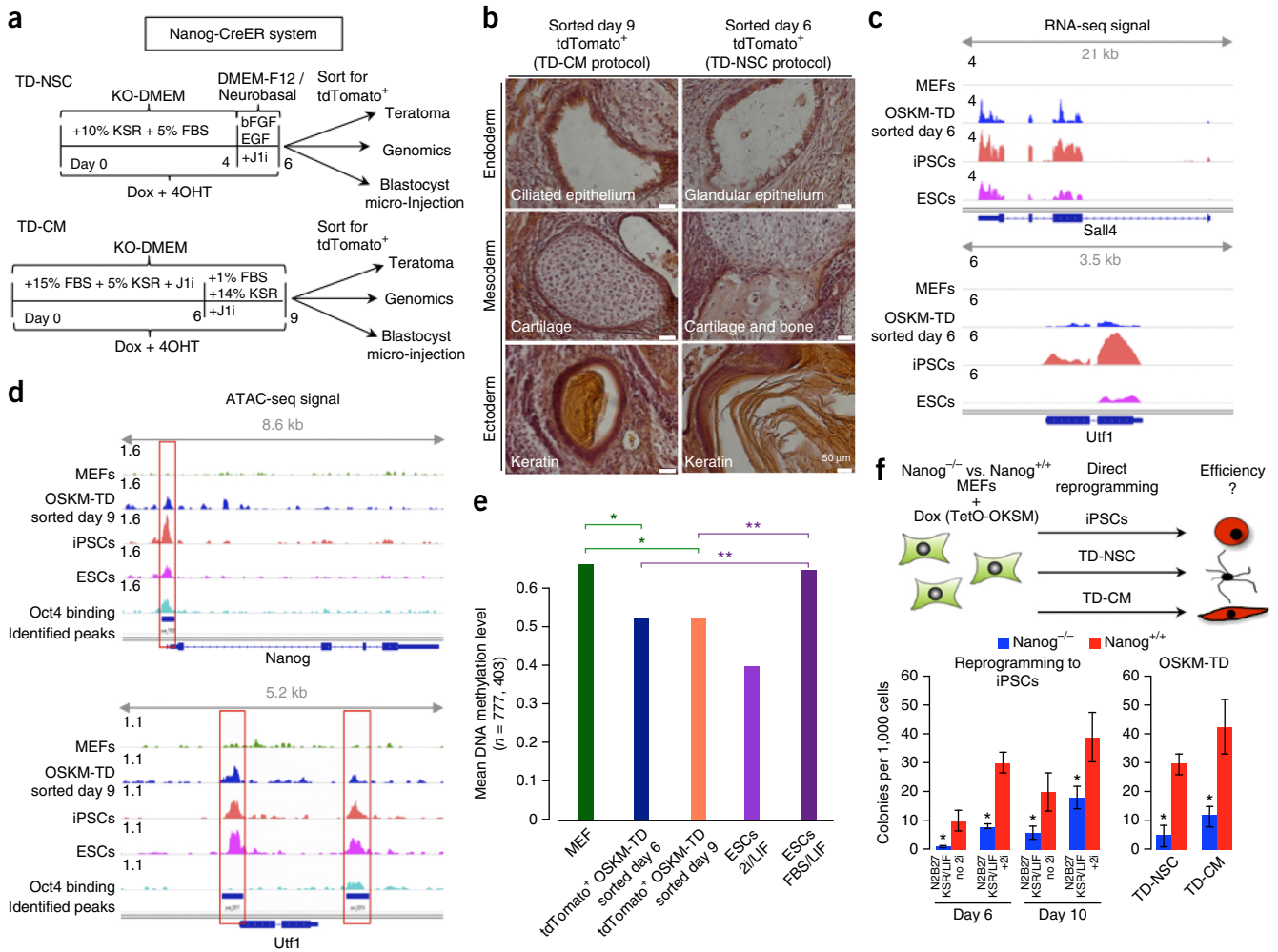


Figure 3 Molecular and functional evidence for transient acquisition of pluripotency during OSKM-TD. (a) Schematic representation of OSKM-TD “intermediate” Nanog-CreER reporter cells sorted for tdTomato⁺ at the end of OSKM induction step and used for teratoma assay, blastocyst microinjection or high-throughput genomics. OSKM-TD protocols were applied on Nanog-CreER reporter MEFs, and stopped at the end of OSKM induction step, at day 6 and 9, respectively. (b) Representative images of H&E stain of teratoma sections with representation of all three germ layers. Scale bars, 50 μm. (c) Transcriptional profile of *Sall4* and *Utf1* measured by poly-A RNA sequencing, in four samples. Data range is left side of each panel. (d) Chromatin accessibility signal measured by ATAC-seq in iPSCs, except for Oct4 binding measured by ChIP-Seq. All peaks identified by MACS (model based analysis for Chip-seq) are indicated by blue lines. Peaks which fall in promoters or in potential enhancers are enclosed by red rectangles. (e) Mean DNA methylation level measured in MEFs, OSKM-TD day 6 and day 9 tdTomato⁺-sorted cells, and ESCs (in ground state naive 2i/LIF conditions, or metastable naive FBS/LIF conditions). Mean methylation was measured using whole-genome bisulfite sequencing and calculated across all CpGs that were covered in all samples. *P < 10⁻¹⁵, **P < 2.2 × 10⁻¹⁶, both in Wilcoxon signed-rank test. (f) Reprogramming efficiencies of Nanog^{-/-} and Nanog^{+/+} MEFs to either iPSCs, TD-NSC or TD-CM, evaluated by AP, Sox1 or myosin colony staining, respectively. Reprogramming was conducted following transduction with FUW-M2rtTA and FUW-TetO-OKSM lentiviral vectors. Error bars, s.e.m. ± mean of biological triplicates (one out of two representative experiments is shown). *P < 0.05, Student *t*-test (in comparison to matched Nanog^{+/+} sample).

with either polycistronic lentiviral vectors or with individual Moloney leukemia retroviruses encoding OSKM (Fig. 2i) showed that 100% of Sox1⁺ TD-NSC and myosin⁺ TD-CM colonies lost mTomato expression and became GFP⁺, marking endogenous Oct4 reactivation during their reprogramming process (Fig. 2i and Supplementary Fig. 8b–d).

The higher efficiency of Oct4 reactivation (100%) compared with Nanog and X-chromosome reactivation that we detected may be explained by the fact that Oct4 is expressed in all states of pluripotency (primed, metastable and naive), whereas Nanog and X-chromosome reactivation are more stringent markers observed only in naive pluripotency^{11,17}. Alternatively, endogenous Oct4 reactivation in isolation may not unequivocally indicate the reacquisition of pluripotency¹⁸. Nevertheless, the Oct4, Nanog and X-chromosome lineage tracing results collectively corroborate the conclusion that the vast majority of mouse OSKM-TD cells pass through a pluripotent state before transitioning toward overt somatic differentiation. Acquisition of pluripotency in these secondary Nanog-CreER cells was not the result of higher levels of OSKM mRNA compared with those in ESCs or during reprogramming to iPSCs, and was demonstrated with several different OSKM delivery methods and reporters (Supplementary Fig. 9a,b).

The above results indicate that the majority of OSKM-TD cells pass through a transient Nanog⁺ pluripotent state before differentiating into various somatic lineages according to the culture conditions applied. To validate and evaluate features of this transient pluripotent state, we sorted Nanog-CreER-reprogrammed cells at the early end point of OSKM induction of the TD protocol (day 6 for TD-NSC and day 9 for TD-CM protocol) (Fig. 3a and Supplementary Fig. 10a,b). Sorted tdTomato⁺ cells were immediately subjected to various functional and molecular tests of pluripotency.

Both day 6- and day 9-sorted tdTomato⁺ cells obtained early during TD-NSC and TD-CM protocols, respectively, generated teratomas *in vivo*, with mature differentiation toward all three germ layers (Fig. 3b and Supplementary Fig. 10c). tdTomato⁺-sorted cells at both time points, but not pluripotent primed EpiSCs (epiblast stem cells) or somatic MEFs, contributed to mouse chimeric embryos after microinjection into host E3.5 blastocysts (Supplementary Fig. 10d,e).

RNA-seq analysis of sorted day 6 tdTomato⁺ TD-NSC protocol derived cells validated reactivation of bona fide pluripotency genes, including Nanog, Sall4, Nr5A2, Cdh1, Utf1 and Epcam (Fig. 3c and Supplementary Fig. 11). qPCR and western blot analysis for endogenous reactivation of several pluripotency factors (Supplementary Fig. 12a), further validated these results. The transcriptional profiles of these cells consistently clustered with those of iPSCs and ESCs and apart from those of established primary NSCs, TD-NSCs and MEFs (Supplementary Fig. 12b,c). As expected, they were not fully identical to those of established ESCs and iPSCs as these early-isolated tdTomato⁺ cells were reprogrammed in sub-optimal conditions and could not consolidate their pluripotency program under optimal naive pluripotency conditions and long-term passaging.

Global chromatin accessibility measured by an assay for transposase-accessible chromatin followed by sequencing (ATAC-seq)¹⁹ showed reactivation of enhancer and promoter regions of key bona fide pluripotency genes in day 9 tdTomato⁺-sorted cells from the TD-CM protocol, similar to what is observed in ESC and iPSC lines (Fig. 3d and Supplementary Figs. 13 and 14). Although the OSKM-TD protocol does not use 2i/LIF conditions, genome-wide DNA methylation assessment by whole-genome bisulfite sequencing showed significant global reduction in DNA methylation levels in tdTomato⁺-sorted cells toward what is typically measured in ground-state naive 2i/LIF

murine pluripotent cells (Fig. 3e). This result documents transient global DNA demethylation during reprogramming by OSKM without applying 2i and may be worth further investigation.

We next explored the functional role of endogenous Nanog reactivation in OSKM-TD. Endogenous Nanog reactivation has been shown to synergistically boost iPSC reprogramming by OSKM²⁰. We compared conversion of Nanog^{-/-} and Nanog^{+/+} MEFs into iPSCs and OSKM-TD cells²⁰. Nanog^{-/-} mESCs can be expanded in N2B27 2i/LIF/15% KSR conditions while maintaining their pluripotency^{21,22} (Supplementary Fig. 15a), and therefore we used these conditions for iPSC generation. Reprogramming of Nanog^{-/-} and Nanog^{+/+} MEFs to pluripotency was done by OSKM transduction, and both yielded validated pluripotent iPSCs (Supplementary Fig. 15b–d). However, the reprogramming efficiency of Nanog^{-/-} MEFs was only 10% that of Nanog^{+/+} MEFs at day 6 after OSKM induction, and only 45% at day 10 (Fig. 3f and Supplementary Figs. 15e and 16). Next, Nanog^{-/-} and Nanog^{+/+} MEFs were used in OSKM-TD protocols for generating TD-NSCs and TD-CMs. Transient reactivation of pluripotency markers such as Utf1 or SSEA1 was evident in Nanog^{-/-} MEFs in both of these protocols by the end of the OSKM induction steps (Supplementary Figs. 17 and 18). Upon completion of the reprogramming process, the Nanog^{-/-} cells yielded validated TD-NSCs and TD-CMs (Supplementary Fig. 19) at an efficiency of only ~25% compared with Nanog^{+/+} cells (Fig. 3f and Supplementary Figs. 15e and 16a). Collectively, these results show that endogenous Nanog reactivation is functionally coupled to OSKM-TD efficiency, and at an equivalent robustness to that observed during OSKM-iPSC formation.

Finally, OSKM-TD protocols make use of media devoid of LIF or with added JAK1 inhibitor, which blocks Stat3 pathway signaling, under the assumption that these conditions completely prevent the formation of pluripotent iPSCs^{3,4}. Given that OSKM transgenes have been found sufficient to support induction and expansion of pluripotent cells (ESCs and iPSCs) from nonpermissive mouse strains (i.e., nonobese diabetic mice) in suboptimal growth conditions lacking ERK1/2 small-molecule inhibitors²³, we wondered whether they could also stabilize mouse iPSCs in suboptimal conditions lacking exogenous LIF. To address this question, we reprogrammed Nanog-CreER reporter MEFs to pluripotency in LIF-free N2B27 medium with added JAK1 inhibitor (Supplementary Fig. 20a,b). Transgene-dependent iPSCs were derived, expanded and shown to be pluripotent by staining for alkaline phosphatase, SSEA1 and Nanog and by the potential to form mature teratomas (Supplementary Fig. 20c,d). These results corroborate previous findings that Stat3 signaling is dispensable for re-establishing pluripotency²⁴, and that simply blocking Stat3 signaling does not prevent acquisition of pluripotency when OSKM transgenes are provided, as the latter can substitute for the requirement for exogenous LIF to induce and maintain pluripotency²³.

Conversion of cell fate *in vitro* and *in vivo* holds great potential for regenerative medicine. There may be different routes of reprogramming to achieve a given cell type²⁵, and it is important to understand their technical and functional advantages and disadvantages and the molecular pathways and trajectories of each method^{26,27}. Although our study of OSKM-TD shows that the vast majority of the cells are reprogrammed to pluripotency and then rapidly differentiate according to the media used, we cannot exclude the possibility that some cells do transdifferentiate without passing through a pluripotent state. From the perspective of potential applications of OSKM-TD, both the advantages and disadvantages associated with iPSC generation will likely be relevant. It will also be useful to conduct similar studies on OSKM-TD of human cells^{5,6,28}. Finally, considering the continual discovery of factors, inhibitors and cytokines that induce and maintain

pluripotency^{18,29,30}, our findings underscore the importance of using molecular lineage tracing to track cell trajectories in different reprogramming protocols. An independent study published in this issue has reached similar conclusions³¹.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Addgene: #59721 and #60037; GEO: GSE67299.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

J.H.H. is supported by a generous gift from I. and P. Mantoux; the New York Stem Cell Foundation (NYSCF), FAMRI, the Kimmel Innovator Research Award, the ERC (StG-2011-281906), the Leona M. and Harry B. Helmsley Charitable Trust, Moross Cancer Institute, the Israel Science Foundation Regular research program, the ICRF Foundation, Helen and Martin Kimmel Institute for Stem Cell research (HMKISCR), the Benozio Endowment fund. J.H.H. and W.G. are supported by an HFSP research grant. J.H.H. is a New York Stem Cell Foundation - Robertson Investigator. We thank K. Hochedlinger for mutual exchange of results and discussions before publication. We thank Weizmann Institute management for providing critical financial and infrastructural support.

AUTHOR CONTRIBUTIONS

I.M. and J.H.H. conceived the idea for this project. I.M. designed and conducted experiments. I.M., N.N., R.M. and J.H.H. wrote the manuscript with contributions from most other authors. M.Z. conducted microinjections. S.G., I.C., V.K., S.H. assisted in tissue culture and reprogramming experiments. S.V., Y.R. and I.M. constructed and targeted Nanog-CreER construct. W.J.G. and J.D.B. optimized ATAC-seq protocol and assisted A.Z. in analysis. E.C. and L.W. conducted and analyzed whole genome bisulfite sequencing. A.Z. and N.N. conducted and analyzed RNA-seq analysis. J.R.D. generated and provided engineered Oct4-CreER reporter cells. R.M. assisted in and supervised all microscopy imaging and analysis by I.M. and A.Z. presented in this study. N.N. supervised the validity and analysis of all bioinformatics experiments and results in this manuscript. The authors have no competing interests as defined by Nature Publishing Group, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
- Vierbuchen, T. & Wernig, M. Direct lineage conversions: unnatural but useful? *Nat. Biotechnol.* **29**, 892–907 (2011).
- Kim, J. *et al.* Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc. Natl. Acad. Sci. USA* **108**, 7838–7843 (2011).
- Efe, J.A. *et al.* Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat. Cell Biol.* **13**, 215–222 (2011).
- Zhu, S. *et al.* Mouse liver repopulation with hepatocytes generated from human fibroblasts. *Nature* **508**, 93–97 (2014).
- Kurian, L. *et al.* Conversion of human fibroblasts to angioblast-like progenitor cells. *Nat. Methods* **10**, 77–83 (2013).
- Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Carey, B.W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat. Methods* **7**, 56–59 (2010).
- Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
- Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009).
- Carter, A.C., Davis-Dusenbery, B.N., Koszka, K., Ichida, J.K. & Eggan, K. Nanog-independent reprogramming to iPSCs with canonical factors. *Stem Cell Rep.* **2**, 119–126 (2014).
- Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
- Hadjantonakis, A.K., Cox, L.L., Tam, P.P. & Nagy, A. An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. *Genesis* **29**, 133–140 (2001).
- Rodriguez-Ubreva, J. *et al.* Pre-B cell to macrophage transdifferentiation without significant promoter DNA methylation changes. *Nucleic Acids Res.* **40**, 1954–1968 (2012).
- Greder, L.V. *et al.* 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).
- Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Festuccia, N. *et al.* Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* **11**, 477–490 (2012).
- Schwarz, B.A., Bar-Nur, O., Silva, J. & Hochedlinger, K. Nanog is dispensable for the generation of induced pluripotent stem cells. *Curr. Biol.* **24**, 347–350 (2014).
- Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513–524 (2009).
- Yang, J. *et al.* Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell* **7**, 319–328 (2010).
- Papp, B. & Plath, K. Epigenetics of reprogramming to induced pluripotency. *Cell* **152**, 1324–1343 (2013).
- Marro, S. *et al.* Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* **9**, 374–382 (2011).
- Ko, K., Araúzo-Bravo, M.J., Kim, J., Stehling, M. & Schöler, H.R. Conversion of adult mouse unipotent germline stem cells into pluripotent stem cells. *Nat. Protoc.* **5**, 921–928 (2010).
- Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
- Yuan, X. *et al.* Combined chemical treatment enables Oct4-induced reprogramming from mouse embryonic fibroblasts. *Stem Cells* **29**, 549–553 (2011).
- Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
- Bar-Nur, O. *et al.* Lineage conversion induced by pluripotency factors involves transient passage through an iPSC stage. *Nat. Biotechnol.* doi:10.1038/nbt.3247 (22 June 2015).

ONLINE METHODS

Generation of reprogrammable Nanog-CreER reporter mice. Homozygote *Rosa26 LoxP-stop-LoxP tdTomato^{+/+}* mice were crossbred with the reprogrammable double homozygote *Rosa26-M2rtTA* and *m.Coll1a OSKM^{+/+}* mice⁹. At day E3.5 blastocysts were extracted and individually seeded in a flat bottom 96-well plate with plated gamma-irradiated feeder cells, in mESCs growth medium: DMEM (Gibco 41965-039) supplemented with human LIF (20 ng/ml prepared in house), 1% L-glutamine, 1% nonessential amino acids, 1% penicillin/streptomycin, 0.1 mM β -mercaptoethanol and 15% FBS (Biological Industries). Developed ESCs colonies were expanded as single clonal lines. To make a targeting construct for inserting *CreERT2* (also known as *Tg(KRT5-cre/ERT2)2Ipc*) in frame with mouse *Nanog*, we used Red/ET based recombineering technique (GeneBridges). Briefly, *CreERT-pgk-gb2-kanamycin* cassette was inserted into a BAC covering the murine *Nanog* gene region in a way that the starting codon of *Nanog* now became the starting codon of *CreERT2*. Next, the BAC fragment containing *CreERT* insert and appropriate homology arms were cloned into a pDTA plasmid encoding the diphtheria toxin gene cassette used for negative selection in mESCs. The kanamycin selection cassette was replaced with a PGK-hygromycin cassette by ligation. The resulting recombineered Nanog-CreER pgk-Hygro construct (Addgene: #59721) was linearized with *SalI* and electroporated into the triple allele knock-in ESCs. Hygromycin-resistant clones (10 d selection) were subcloned and screened by Southern blot analysis using a 3' external probe (Addgene: #60037) and a PCR creating a 2-kb amplification of the 5' arm, with upstream external forward primer and reverse primer from the Nanog CreER knock-in. Correctly targeted clones were transfected with a Flippase-encoding construct to remove the inverted hygromycin-resistance cassette, whereas no difference in tracing results was observed after this optional step. Properly targeted ESCs were microinjected into blastocysts to create chimeric animals. At day E12.5–E13.5, embryos were extracted and used to derive MEFs containing the reporter system with all four knock-in alleles. V6.5 ESCs were also similarly targeted with Nanog-CreER reporter construct and used for some of the analysis. Karyotype analysis was conducted to validated chromosomal stability of all lines used, and genotyping to validate their alleles. Mycoplasma testing is conducted biweekly in our laboratory (MycAlert Kit, Lonza), and all reported results were conducted on noncontaminated cells.

Southern blot analysis. Genomic DNA was extracted from each hygromycin-resistant targeted subclone. 10–15 μ g of genomic DNA was digested with *MfeI* restriction enzyme for 5 h and separated by gel electrophoresis. The DNA was transferred to a nitrocellulose membrane that was next hybridized with a radioactively labeled probe and developed using enhanced chemiluminescent substrate (Thermo Scientific).

iPSC reprogramming by OSKM. For mouse iPSC reprogramming experiments, secondary OSKM transgenic reporter MEFs were plated at day 0 in MEF medium: DMEM supplemented with 1% L-glutamine, 1% nonessential amino acids, 1% penicillin/streptomycin and 10% FBS, and the following day reprogramming was initiated by changing the medium to N2B27 2i/LIF: DMEM-F12 (BI 01-170-1a) and neurobasal medium (Gibco 21103-049) mixed by 1:1 ratio and supplemented with N2 (Insulin Sigma I-1882, Apo-Transferrin Sigma T-1147, Progesterone Sigma P8783, Putrescine sigma P5780, sodium selenite Sigma S5261), B27 (Gibco 17504-044) and 2i: CHIR (Axon 99021, 3 μ M) PD0325901 (Tocris 4192, 1 μ M), LIF and adding doxycycline hyclate (Sigma D9891, 4 μ g/ml), with 4-hydroxytamoxifen (Sigma H7904, 1 μ M) for activation of the CreER reporter system. The medium was replaced every 48 h, and colonies were analyzed at the indicated time points. Additional FUW-M2RtTA and/or TetO-STEMCCA lentiviral induction were optionally used to enhance reprogramming efficiency. *Nanog^{+/+}* and *Nanog^{-/-}* ESCs²⁰ were rendered transgenic for a constitutively expressed puromycin-resistance cassette³² and were subsequently microinjected into host blastocysts. Chimeric embryos were used to derive MEFs, which were selected by puromycin (2 μ g/ml) resistance for at least 3 d then used for transduction and reprogramming in 2i/LIF/15% KSR (Invitrogen) enriched conditions. In **Figure 1a** and **Supplementary Figure 1**, MEFs were obtained from *Rosa26-m2rtTA(+/+* or *+/-*); *m.Coll1a-OSKM(+/+* mice (Jackson Laboratories #011004),

Rosa26-m2rtTA(+/+/-); *m.Coll1a STEMCCA-OKSM(+/+/-)* mice (Jackson Laboratories #011001) that were bred with Oct4-GFP or Nanog-GFP knock-in reporter mice¹⁰. NGFP4 iPSC line was generated as previously described¹⁰.

Transdifferentiation reprogramming with OSKM. Reprogramming of MEFs to cardiomyocytes was done in media containing: knockout DMEM (Invitrogen) with supplemented 15% FBS and 5% KSR with added JAK1 inhibitor (Calbiochem 420099, 0.5 μ M) for 6 d, followed by switching to 1% FBS and 14% KSR and JAK1 inhibitor for 3 d. The medium was supplemented with 1% L-glutamine, 1% nonessential amino acids, 1% penicillin/streptomycin and 0.1 mM β -mercaptoethanol. Doxycycline hyclate (Dox) and 4-hydroxytamoxifen (4OHT) were added to the medium at day 1 and continuously kept in the medium until day 6–9. From day 9 onwards, cells were cultured in chemically defined RPMI (Gibco 21875-034)/N2B27 medium (0.5XN2, 1XB27) with the cardio-inductive growth factor BMP4 (Peprotech 120-05ET) for the first 5 d, and supplemented with BSA fraction V 0.05% (Gibco 15260-037) 1% L-glutamine, 1% nonessential amino acids, 1% penicillin/streptomycin and 0.1 mM β -mercaptoethanol. Colonies were allowed to develop for 21 d or until beating cardiomyocytes were noticed among the plated cells.

Reprogramming of MEFs to NSCs was carried out in knockout DMEM medium (Gibco 10829-018) with 10% KSR, 5% FBS, 1% L-glutamine, 1% nonessential amino acids, 1% penicillin/streptomycin, 0.05 mM β -mercaptoethanol and with Dox + 4OHT for the first 3–6 d. Thereafter, Dox and 4OHT were withdrawn from the medium, and cells were grown in neural reprogramming medium: DMEM-F12 and neurobasal medium mixed by 1:1 ratio and supplemented with BSA fraction V 0.05%, 1 \times N2, 1 \times B27, 1% L-glutamine, 1% nonessential amino acids, 1% penicillin/streptomycin and 0.1 mM β -mercaptoethanol with bFGF (Peprotech 450-33, 20 ng/ml) and EGF (Peprotech 315-09, 20 ng/ml), until NSCs colonies were formed by day 13. From day 4 onwards JAK1 inhibitor was added to the culture medium. At day 13, formed colonies were subcloned and grown in NSC medium on poly-D-lysine and laminin-coated plates. We noted that OSKM activation by Dox for 3 d was sufficient for OSKM-TD, though substantially fewer colonies developed at these minimally required activation periods, as reported in the original papers^{3,4}. When applying OSKM-TD protocols without 4OHT or with ethanol (EtOH) as a control replacement, all cardiomyocytes and NSCs colonies were negative for tdTomato signal, excluding leakiness of this system under OSKM-TD conditions, as similarly observed with iPSCs (**Fig. 1d**).

Genetic tracing for X-chromosome reactivation in reporter murine female cells. Female *Rosa26 M2rtTA^{+/+}/m.Coll1a-TetO-OSKM^{+/+}* mice⁹ were bred with male mice carrying a CMV-GFP knock-in to the X chromosome. Adult Female F₁ mice were used as a source for adult tail tip fibroblasts or Pro-B somatic cells. Additional FUW-M2RtTA and TetO-STEMCCA lentiviral induction were used to enhance reprogramming efficiency. GFP⁻ population were sorted on a FACSAria III five-laser equipped cell sorter (BD Biosciences) and further expanded for OSKM-iPSCs and OSKM-TD reprogramming experiments. Colonies were scored for GFP signal on Zeiss Axioscope D1 microscope, and as expected, iPSCs were homogeneously GFP⁺ (because they maintain both X chromosomes active), whereas OSKM-TD colonies had patchy GFP expression consistent with random inactivation of the X alleles. Yet, the presence of GFP⁺ cells in TD colonies indicates reactivation of the X chromosome in the clonal population, followed by random inactivation during differentiation.

Genetic tracing with Oct4 CreER mT/mG reporter MEFs. The Oct4-CreER reporter system has been previously reported and validated elsewhere¹⁶. Briefly, the tamoxifen-inducible Cre recombinase gene was inserted into the 3'-untranslated region (UTR) of *Oct4*. The knock-in allele was maintained on a C57BL/6 mice strain background. *Oct4-MerCreMer* mice are deposited with The Jackson Laboratory (Stock No. 016829). For lineage tracing, MEFs were derived from E12.5 embryos obtained following *Oct4-MerCreMer* mice mating with homozygous double-fluorescent *Rosa26 LoxP-mTomato-stop-LoxP-mEGFP* reporter mice (The Jackson Laboratory, Stock No. 007576). The latter system allows marking tamoxifen-dependent Cre recombination by replacement of membrane-targeted tdTomato with membrane-targeted enhanced green fluorescent protein (EGFP) expression.

Mouse embryo micromanipulation. Pluripotent ESCs, day-9 iPSCs, EpiSCs or Nanog-CreER (days 6 or 9) tdTomato⁺ cells were injected into BDF2 (F2 mating of siblings of female C57BL/6 and male DBA/2 F1 mice) diploid blastocysts (ten cells per embryo, unless specified otherwise), harvested from hormone-primed BDF1 6-week-old females. MEF cells constitutively labeled with tdTomato were used as negative controls (three MEF cells were injected per blastocyst). All cells were subjected to the same sorting conditions and were immediately used for microinjections. Microinjection into BDF2 E3.5 blastocysts placed in M2 medium under mineral oil was done by a flat-tip microinjection pipette. After microinjection, blastocysts were returned to KSOM media (Zenith) and kept at 37 °C until transferred to recipient females. Ten to 15 injected blastocysts were transferred to each uterine horn of 2.5-d post-coitum pseudo-pregnant females. 129 EpiSCs, constitutively labeled with tdTomato exogenous transgene, were treated with ROCK inhibitor 48 h before harvesting with Accutase (Sigma-Aldrich), to increase survival yield. The latter experiments were approved by the Weizmann Institute IACUC (00330111-2). Animals were not randomized and investigators were not blinded for this study. No animals were excluded from any analysis throughout this study. Accurate mouse embryos and/or animal group size is indicated in **Supplementary Figures 10c,e**.

Viral production. For primary cell reprogramming, $\sim 3 \times 10^6$ 293T cells in a 10-cm culture dish were transfected with a solution made of 770 μ l DMEM (Invitrogen) together with 50 μ l of TransIT-LT1, pPAX (3.5 μ g), pMDG (1.5 μ g) and 5 μ g of the lentiviral target plasmid (FUW-M2rtTA, FUW-TetO-STEMCCA-OKSM, FUW-TetO-OSKM). Viral supernatant was harvested 48 and 72 h after transfection, filtered through 0.45- μ m sterile filters (Nalgene) and added freshly to the reprogrammed cells. Murine stem cell virus C/EBP α retrovirus stocks were prepared by transient transfection of Phoenix-Eco cells using Eugene (Roche), and supernatants were harvested 48 h later and filtered. Pro-B cell expansion on OP9 cells and transduction with C/EBP α was conducted as previously described¹⁰.

Immunostaining. Nanog-CreER iPSCs in LIF-free N2B27 medium with supplemented JAK1 inhibitor and Dox, and Nanog^{-/-} iPSCs were cultured on glass cover slips (13 mm, 1.5 H; Marienfeld, 0117530), washed three times with PBS and fixed with 4% paraformaldehyde for 10 min at room temperature. Cells were then permeabilized and blocked in 0.1% Triton, 0.1% Tween and 5% FBS in PBS for 15 min at room temperature. Primary antibodies were incubated for 2 h at room temperature and then washed with 0.1% Tween and 1% FBS in PBS three times. Next, cells were incubated with secondary antibody for 1 h at room temperature, washed and counterstained with DAPI, mounted with Shandon Immu-Mount (Thermo Scientific) and imaged. All secondary antibodies were diluted 1:200.

For staining of transdifferentiated NSCs, cardiomyocytes and intermediate cells, all cells were reprogrammed in 6-well tissue culture plates, fixed and stained in the tissue culture wells. The following antibodies were used: polyclonal goat Gata4 antibody (Santa Cruz SC-1237, 1:200), mouse monoclonal myosin antibody (Hybridoma MF-20, 1:200), goat polyclonal Sox1 antibody (R&D AF-3369, 1:200), mouse monoclonal Tuj1 antibody (Covance MMS-435P, 1:500), mouse monoclonal Nestin antibody (Hybridoma Rat-401, 1:20), rabbit polyclonal Sox2 antibody (Millipore AB5603, 1:200), Chicken polyclonal MAP2 antibody (abcam ab5392, 1:5,000), rabbit polyclonal O2 antibody (Millipore AB9610, 1:1,000), mouse monoclonal IgM O4 antibody (R&D MAB1326, 1:500), mouse monoclonal Oct4 antibody clone C10 (Santa Cruz SC5279, 1:200), rabbit polyclonal Nanog antibody (Bethyl A300-397A, 1:200), mouse monoclonal IgM SSEA1 antibody (Hybridoma MC-480 clone, 1:20) and rabbit polyclonal Utf1 antibody (abcam ab24273 1:1,000). All antibodies in this study have been validated in the literature and by our internal tests on primary cell lines.

Microscopy and image analysis. Images were acquired with A1 AxioScope microscope (Carl Zeiss) equipped with DP73 camera (Olympus) or with Z1 AxioScope microscope (Carl Zeiss), using a 20X Plan-Apochromat objective (numerical aperture 0.4). All images were acquired in sequential mode. Images were processed with Zeiss Zenblue 2011 software (Carl Zeiss) and

Adobe Photoshop CS4. An in-house algorithm was developed and implemented in Matlab to process the images and analyze the percentage of tdTomato⁺/green-fluorescence⁺ overlay area out of total green-fluorescence⁺ area (**Supplementary Figs. 5c and 7**). The algorithm consists of the following steps: (i) noise reduction using sliding median filter; (ii) creating a binary mask image per channel using a detection threshold defined as the overall image median value plus bias (defined as 10% of the dynamic range); all pixels with values above the threshold are “detected”; (iii) overlaying tdTomato⁺ and green-fluorescence⁺ mask images and calculating the percentage of tdTomato⁺ & green-fluorescence⁺ pixels out of all green-fluorescence⁺ pixels.

Western blot analysis. Whole-cell protein extracts were isolated from wild-type Nanog^{+/+}, Nanog^{-/-} mESCs, Nanog-CreER reporter MEFs and sorted day 6 tdTomato⁺ intermediate cells. Blots were incubated with the following antibodies in 5% BSA/TBST: rabbit polyclonal Hsp90 β antibody (Calbiochem CA1016, 1:5,000), rabbit monoclonal Gapdh antibody (Epitomics 2251-1, 1:5000), rabbit polyclonal Oct4 antibody clone H-134 (Santa Cruz SC9081, 1:1,000), rabbit polyclonal Sall4 antibody (abcam ab29112 1:1,000) and rabbit polyclonal Nanog antibody (Bethyl A300-397A, 1:1,000). Secondary antibodies were horseradish peroxidase-linked goat anti-rabbit (1:10,000; Jackson). Blots were developed using ECL (Thermo Scientific).

Teratoma assay. In **Supplementary Figure 10c**, sorted Nanog-CreER tdTomato⁺ cells at day 6 and 9 of OSKM-TD protocols were injected subcutaneously into the flanks of immune-deficient nonobese diabetic severe combined immunodeficient gamma chain deficient (NSG) mice. Sorted ESCs were used as a positive control. For each time point, 2×10^5 -sorted tdTomato⁺ cells were collected in 25 \times diluted Matrigel (BD Bioscience), and subsequently injected subcutaneously in NSG mice. Three animals were injected (on one side) per each group. After 4 weeks, all injected mice were euthanized and the tumor mass extracted, measured and fixed in 4% paraformaldehyde overnight. In **Supplementary Figure 20**, Nanog-CreER iPSCs expanded in LIF-free N2B27 medium with supplemented JAK1 inhibitor and doxycycline and Nanog^{-/-} iPSCs grown in N2B27 2i/LIF cell lines were expanded for over eight passages and injected (2×10^6 cells) subcutaneously into the flanks of NSG mice and dissected 4–6 weeks afterwards. Slides were prepared from the paraffin-embedded fixed tissue, which were next stained with H&E and inspected for representation of all three germ layers. Differentiation results were confirmed by a board-certified histopathologist.

Alkaline phosphatase staining. Cells were grown on 6-well tissue culture plates and were washed three times with PBS and fixed with 4% paraformaldehyde for 2 min at room temperature. Alkaline phosphatase staining was performed according to the manufacturer’s protocol (Millipore SCR004).

RNA sequencing and gene expression profiling analysis. RNA was extracted from Trizol pellets and used for RNA-seq by TruSeq RNA Sample Preparation Kit v2 (Illumina) according to the manufacturer’s directions. DNA sequencing was conducted on Illumina HiSeq1500. Tophat software version 2.0.10 was used to align reads to mouse mm10 reference genome (UCSC, December 2011). Read counts per exon were calculated over all 628,052 exons in mm10 ensemble GTF (UCSC, December 2011), using bedtools coverage command (version 2.16.2). Exons annotated as protein coding, pseudogene or long noncoding (linc)RNA ($n = 459,556$) were selected for further analysis. Exon counts were normalized by the exon length in kbp and by million number of aligned reads per sample, to give RPKM (read-per-kilobase-per-million-reads) values. Gene expression was defined by the average expression level (RPKM) of all exons associated with a certain gene. Sample hierarchical clustering was done over primary brain-derived NSCs, MEFs, mESCs and two subclones of TD-NSCs (**Supplementary Fig. 5a**); only exons with at least one RPKM call > 2 were selected, resulting in 371,807 exons corresponding to 19,755 genes. Additional hierarchical clustering was done over all samples (**Supplementary Fig. 12c**); only exons with at least one RPKM call > 2 were selected, resulting in 379,435 exons corresponding to 20,419 genes. Hierarchical clustering was carried out in both figures, using Matlab (version R2011b) clustergram command, with Spearman correlation as a distance metric, average linkage

and per-row standardization, such that the mean of each row is 0 and its s.d. is 1. Gene profiles (Fig. 3c and Supplementary Figs. 11 and 12b) were generated using IGV version 2.3.26.

Whole-genome bisulphite sequencing for DNA methylation. Nanog-CreER reporter MEFs were reprogrammed in TD-NSC or TD-CM protocols and sorted for tdTomato⁺ at the end of the Dox induction step, at day 6 and 9, respectively. Similarly, cells were collected from MEF cultures, as well as V6.5 mouse ESCs grown in FBS/LIF (taken from GSE41923) or in naive ground-state KSR/2i/LIF conditions. DNA was isolated from snap-frozen intermediate sorted cells using the Quick-gDNA mini prep kit (Zymo). DNA was then converted by bisulfite using the EZ DNA Methylation-Gold kit (Zymo). Sequencing libraries were created using the EpiGnome Methyl-Seq (Epicentre) and sequenced on an Illumina HiSeq 1500 system. The sequencing reads were aligned to the Mouse Genome Build 37 (mm9) using a proprietary script based on Bowtie 2. Due to the low sequencing depth of some of the samples, the methylation calls were downsampled so that each CpG was covered by a single call (being either 'methylated' or 'nonmethylated'). Mean methylation was then calculated over all CpGs that had calls in all samples (either methylated "1" or nonmethylated "0"). Whole genome bisulfite sequencing data are deposited under GEO submission: GSE67299.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq). Nanog-CreER reporter MEFs were reprogrammed using OSKM-TD protocol until the last OSKM induction step, at which point tdTomato⁺ intermediate cells were sorted and applied for ATAC sequencing. Briefly, 50,000 sorted cells were centrifuged at 500g for 3 min, followed by a wash using 50 µl of cold PBS and centrifugation at 500g for 3 min. Cells were lysed using cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂ and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500g for 10 min using a refrigerated centrifuge. Next, the pellet was resuspended in the transposase reaction mix (25 µl 2× TD buffer, 2.5 µl transposase (Illumina) and 22.5 µl nuclease-free water). The transposition reaction was carried out for 30 min at 37 °C and immediately put on ice. Directly afterwards, the sample was purified using a Qiagen MinElute kit. Following purification, the library fragments were amplified using custom Nextera PCR primers 1 and 2 for a total of 12 cycles. Following PCR amplification the libraries were purified using a Qiagen MinElute Kit.

ATAC-seq analysis. Reads were aligned to mm10 mouse genome using Bowtie2 with the parameter -X2000 (allowing fragments up to 2 kb to align). Duplicated aligned reads were removed using Picard MarkDuplicates tool with the command REMOVE_DUPLICATES = true. To identify chromatin accessibility signal we considered only short reads (≤100 bp) that corresponded to a nucleosome-free region¹⁹. To detect and separate accessible loci in each sample, we used MACS version 1.4.2-1 with -call-subpeaks flag (PeakSplitter version 1.0). Next, summits in previously annotated spurious regions were filtered out using a custom blacklist targeted at mitochondrial homologs.

To develop this blacklist, we generated 10,000,000 synthetic 34-mer reads derived from the mitochondrial genome. After mapping and peak calling of these synthetic reads we found 28 high-signal peaks for the mm10 genome. For all subsequent analysis we discarded peaks falling within these regions. Each peak in each sample was represented by a 300-bp region around the summit center. The peaks from all samples were unified and merged (using bedtools unionbedg and merge commands), to create a single list of 94,855 identified accessible loci. Accessibility signal and peaks, alongside previously published Oct4 binding signal (GSE49767) were visualized using IGV version 2.3.26 (Fig. 3d and Supplementary Figs. 13 and 14a).

To quantify the change in accessibility between samples, we estimated read coverage in all accessible loci ($n = 94,855$) using bedtools coverageBed command (version 2.16.2). Read coverage was normalized by peak length in kbp and by million number of aligned reads per sample, to give RPKM values. Further analysis was done using Matlab version R2011b. Differential peaks were defined by fourfold change difference between MEF and ESC samples, resulting in 20,540 peaks. The correlation matrix was calculated using Spearman correlation over all 20,540 differential peaks, followed by hierarchical clustering performed using Spearman correlation as a distance metric and average linkage (Supplementary Fig. 14b). ATAC-seq data are deposited under GEO no. GSE67299.

PCR analysis. Total RNA was isolated using the RNeasy kit (Qiagen). 3 µg of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). 1 µg of DNase-I-treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately resuspended in 100 µl of water. Quantitative PCR analysis was done in triplicate using 1/50 of the reverse transcription reaction on Vii7 platform (Applied Biosystems). Error bars, s.d. of technical triplicate for each measurement. For mouse pluripotency factors relative expression, the following validated mouse-specific primer sequences were used: *Oct4* forward primer: 5' ACA TCG CCA ATC AGC TTG G 3' reverse primer: 5' AGA ACC ATA CTC GAA CCA CAT CC 3'. *Sox2* forward primer: 5' ACA GAT GCA ACC GAT GCA CC 3' reverse primer: 5' TGG AGT TGT ACT GCA GGG CG 3'. *Klf4* forward primer: 5' GCA CAC CTG CGA ACT CAC AC 3' reverse primer: 5' CCG TCC CAG TCA CAG TGG TAA 3'. For genotyping the following primers were used: *M2rtTA* genotyping: Forward primer 5' AAA GTC GCT CTG AGT TGT TAT 3'; Reverse primer 5' GCG AAG AGT TTG TCC TCA ACC 3'. *Col1a* knock-in allele genotyping: Forward primer 5' GCA CAG CAT TGC GGA CAT G 3'; Reverse primer 5' TTG CTC AGC GGT GCT GTC CA 3'. *tdTomato* genotyping: Forward primer 5' CTG TTC CTG TAC GGC ATG G 3'; Reverse primer 5' GGC ATT AAA GCA GCG TAT CC 3'. *CreER* genotyping: Forward primer 5' TCC ATT TGT CAC GTC CTG CAC G 3'; Reverse primer 5' GAA GGA ACC TGG CTT TGC CCT G 3'.

32. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).