LETTER

doi:10.1038/nature14274

Early reprogramming regulators identified by prospective isolation and mass cytometry

Ernesto Lujan[1,2,3]*, Eli R. Zunder[4]*, Yi Han Ng[1,3,5], Isabel N. Goronzy[1,3], Garry P. Nolan[4] & Marius Wernig[1,3]

In the context of most induced pluripotent stem (iPS) cell reprogramming methods, heterogeneous populations of non-productive and staggered productive intermediates arise at different reprogramming time points[1–11]. Despite recent reports claiming substantially increased reprogramming efficiencies using genetically modified donor cells[12,13], prospectively isolating distinct reprogramming intermediates remains an important goal to decipher reprogramming mechanisms. Previous attempts to identify surface markers of intermediate cell populations were based on the assumption that, during reprogramming, cells progressively lose donor cell identity and gradually acquire iPS cell properties[1,2,7,8,10]. Here we report that iPS cell and epithelial markers, such as SSEA1 and EpCAM, respectively, are not predictive of reprogramming during early phases. Instead, in a systematic functional surface marker screen, we find that early reprogramming-prone cells express a unique set of surface markers, including CD73, CD49d and CD200, that are absent in both fibroblasts and iPS cells. Single-cell mass cytometry and prospective isolation show that these distinct intermediates are transient and bridge the gap between donor cell silencing and pluripotency marker acquisition during the early, presumably stochastic, reprogramming phase[2]. Expression profiling reveals early upregulation of the transcriptional regulators *Nr0b1* and *Etv5* in this reprogramming state, preceding activation of key pluripotency regulators such as *Rex1* (also known as *Zfp42*), *Dppa2*, *Nanog* and *Sox2*. Both factors are required for the generation of the early intermediate state and fully reprogrammed iPS cells, and thus represent some of the earliest known regulators of iPS cell induction. Our study deconvolutes the first steps in a hierarchical series of events that lead to pluripotency acquisition.

Reprogramming somatic cells to a pluripotent state by forced transcription factor expression is typically an inefficient process involving heterogeneous populations that impede molecular analysis of productive reprogramming[1–11,14]. Previous studies have shown that reprogramming is a multi-stage process involving presumably early stochastic and late deterministic phases[2,3,5]. Progress has been made characterizing intermediates of the late phase given the appearance of well-known pluripotency markers at that time[1,2,7,8,10,15]. In contrast, not much is known about the early stochastic phase except the consistent observation that downregulation of donor cell markers is an early feature of successful reprogramming[1,2,7,8,10,15–20].

To identify surface markers of early reprogramming stages, we screened 176 antibodies on cells representing three stages of the reprogramming process: (1) mouse embryonic fibroblasts (MEFs), (2) a previously characterized partially reprogrammed cell (PRC) line[18,20,21] and (3) embryonic stem cells (ESCs). We identified 21 markers enriched or shared between these cell types and characterized their co-expression by single-cell mass cytometry using spanning-tree progression analysis of density-normalized events (SPADE), which groups similar cells into a defined number of clusters[22,23] (Fig. 1a, b and Extended Data Fig. 1). Next, we characterized their expression by mass cytometry during *Oct4*-, *Sox2*-, *Klf4*- and *c-Myc*-driven MEF reprogramming. By day 3, downregulation of the fibroblast expression program was evident (Fig. 1c and Extended Data Figs 2 and 3). At day 6, major branches were delineated by the PRC marker CD73 and ESC markers CD54, CD326 and SSEA1. Little co-expression was observed between these markers, suggesting several intermediates arise during early reprogramming or early expression of some of these markers may not be indicative of productive reprogramming. By day 9, CD326 and SSEA1 expression converged in a subpopulation and persisted on days 12 and 16 (Extended Data Fig. 2b). These clusters were heterogeneous for CD73, suggesting they may be derivatives of separate populations or a $CD73^{high}$ subpopulation becomes $CD326^{high}$, $SSEA1^{high}$. Over our time course, the ESC marker CD54 largely localized to fibroblast branches and did not cluster with $CD326^{high}$ and $SSEA1^{high}$ clusters, suggesting CD54 expression in pluripotent cells is a late event[7].

Hypothesizing that cells destined to successfully reprogram acquire surface markers in a stepwise, non-stochastic manner during early reprogramming, we assessed reprogramming efficiencies for cells with high or low expression of the above-characterized surface markers at early time points (Fig. 2a–c and Extended Data Fig. 4a). As mass culture experiments can be misleading because a single proliferative population can seed multiple secondary colonies, we conducted 96-well assays to assess unique reprogramming events. On the basis of the current literature it would be expected that reprogramming cells would be enriched by (1) low levels of fibroblast markers and (2) high levels of ESC markers independent of the time and state of reprogramming[2,7,8,10]. Indeed, by day 3, populations expressing low levels of all fibroblast markers except CD47 enriched for reprogramming populations compared with highly expressing populations (Fig. 2a). Surprisingly though, at these early time points, cells with high levels of the ESC markers SSEA1, CD54, CD326 or CD71 did not show significantly increased reprogramming (Fig. 2a–c). Day 9 fractions expressing high levels of CD326 or SSEA1 began to show greater but insignificant enrichment for reprogramming populations. We confirmed previous reports that SSEA1-sorted cells produce more iPS cell colonies in mass culture[1,8,10], emphasizing the critical importance of the 96-well assay (Extended Data Fig. 4b, c). Unlike what was previously assumed, our findings demonstrate that acquisition of markers that define the pluripotent state is a late event and early expression of ESC markers has little predictive value for successful reprogramming. Additionally these data support the idea that the mesenchymal-to-epithelial transition, as judged by the epithelial marker CD326 (EpCAM), is a late event[2,15–18,20].

Our surface marker screen identified several specific markers for stable, partially reprogrammed cells[4,9,18,21] (Fig. 1a). Although thought to be 'stuck' during reprogramming[4,18,21], we hypothesized a productive intermediate might arise between fibroblasts and iPS cells that share a subset of these markers. Indeed, day 6 fractions expressing high levels of the PRC markers CD73, CD49d and CD200 significantly enriched for reprogramming populations (Fig. 2b). Also, on day 9 the $CD73^{high}$

[1]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California 94305, USA. [2]Department of Genetics, Stanford University, Stanford, California 94305, USA. [3]Department of Pathology, Stanford University, Stanford, California 94305, USA. [4]Baxter Laboratory in Stem Cell Biology, Department of Microbiology and Immunology, Stanford University, Stanford, California 94305, USA. [5]Department of Microbiology and Immunology, Stanford University, Stanford, California 94305, USA.
*These authors contributed equally to this work.

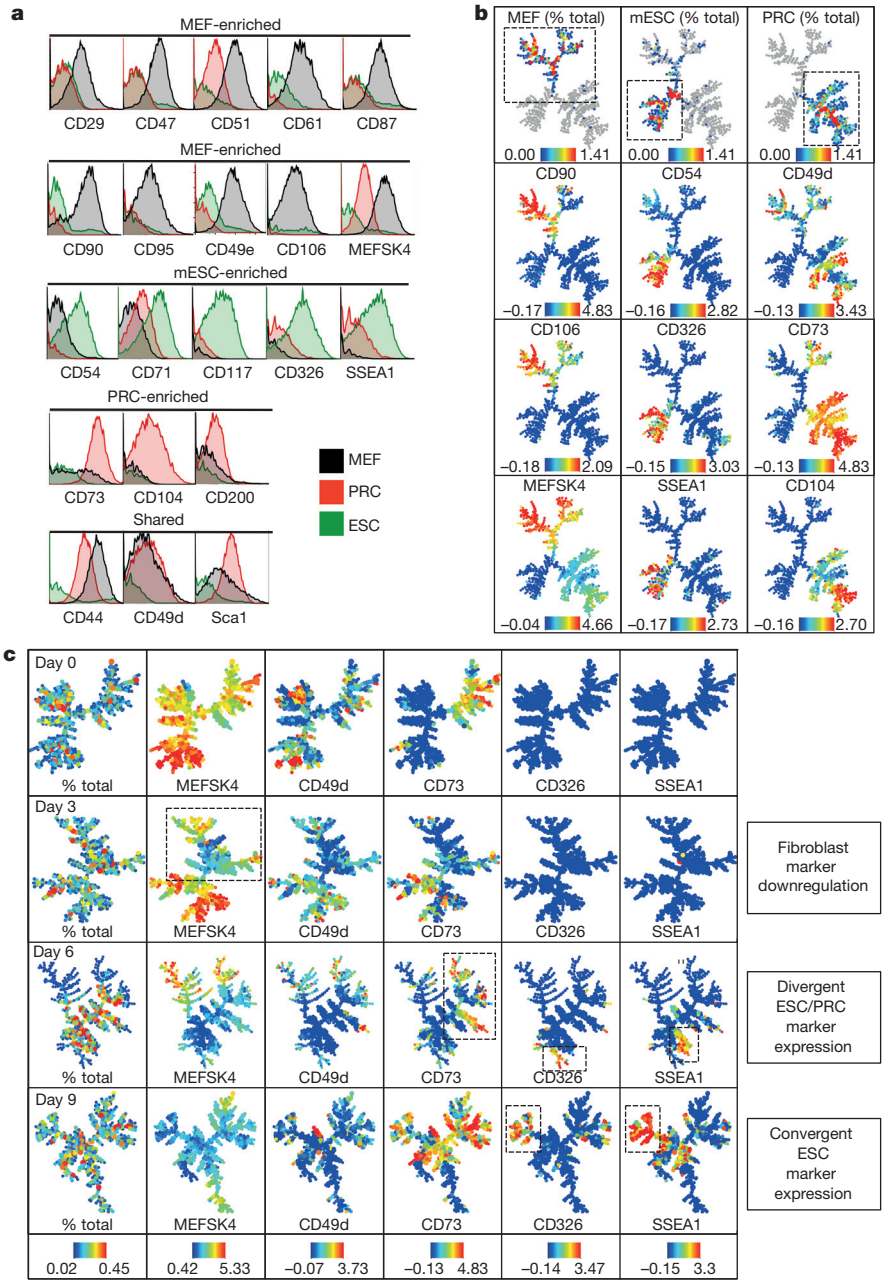


Figure 1 | Reprogramming surface marker profiling by mass cytometry. **a**, Histogram overlays show mass cytometry signal intensity for the three cell populations analysed: mouse embryonic fibroblasts (MEF, black), partially reprogrammed cells (PRC, red) and embryonic stem cells (ESC, green). **b**, SPADE analysis of combined MEF, PRC and ESC data sets. Colour bars represent absolute percentages (top row) and ArcSinh-transformed counts for each marker. **c**, SPADE analysis of infected reprogramming MEF populations at days 0, 3, 6 and 9. For each marker, the same colour scale was applied to every sample, allowing direct comparison between time points. Colour bars represent absolute percentages (left) and ArcSinh-transformed counts.

and CD49d^{high} populations contained higher reprogramming activity (Fig. 2c). In agreement with these results, the day 6 SPADE analysis showed CD73^{high}, CD49d^{high} and CD200^{high} branches largely clustering independently from branches enriched for ESC and MEF markers (Fig. 1c). These results demonstrate that distinct intermediate populations arise after fibroblast program repression but before ESC marker acquisition.

We next focused on the markers CD73 and CD49d. When corrected for plating efficiency, both CD73^{high} and CD49d^{high} populations showed remarkably high reprogramming efficiencies of 9.5% ± 3.5 and 12.5% ± 5.7, respectively (Fig. 2d–f). Similar enrichment of a reprogramming-prone population was observed in reprogramming tail tip fibroblasts and glial-restricted neural precursor cells, suggesting CD73 and CD49d may be universal markers of intermediate reprogramming stages (Fig. 3a–d). Finally, we explored their potential functional implications during reprogramming, and observed that adenosine, the enzymatic product of CD73, has a negative effect throughout and that CD49d activity is necessary during late reprogramming (Extended Data Fig. 4e).

We then used our day 6 SPADE analysis to identify heterogeneously expressed markers that could subdivide the CD73^{high} reprogramming-prone population and conducted single-cell efficiency assays (Fig. 3e, f). Within the CD73-positive population, the CD44^{high}, CD71^{low} and CD326^{high} fractions failed to reprogram, while CD49d^{high} and CD326^{low} fractions enriched for a reprogramming population. Thus, a CD73^{high} CD49d^{high} CD326^{low} CD44^{low} signature best describes the population undergoing productive reprogramming on day 6. Overlaying this signature onto the day 6 SPADE tree allowed determination of the exact cellular clusters most similar to this reprogramming-prone signature

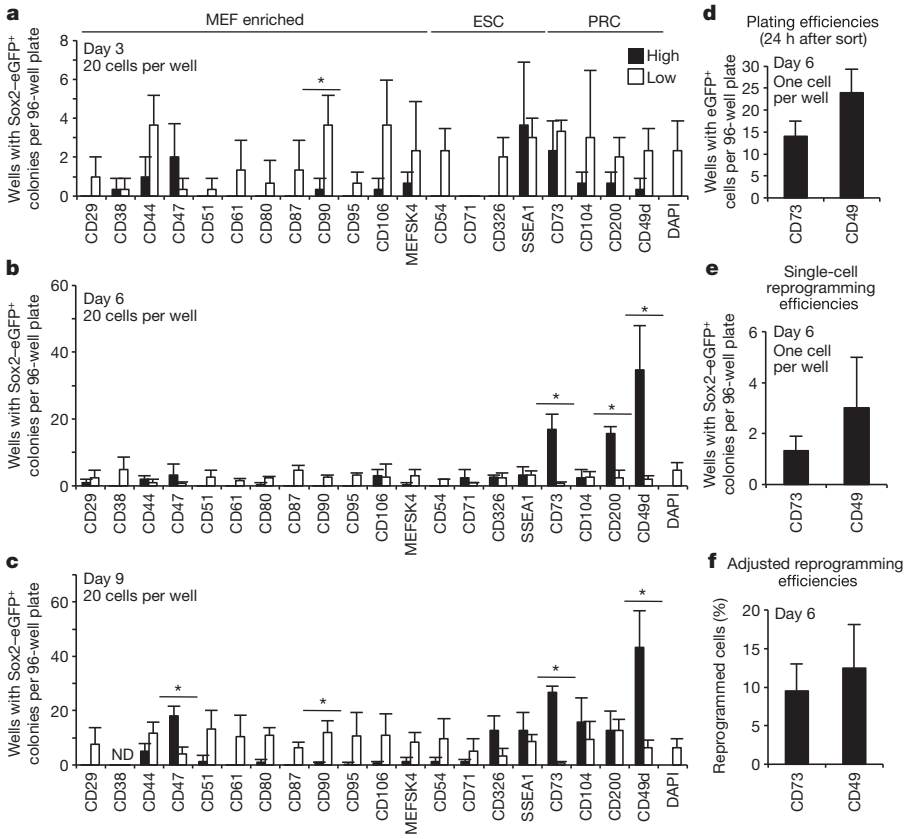


Figure 2 | A surface marker screen identifies an early CD73^{high} CD49d^{high} reprogramming intermediate. **a**–**c**, Ninety-six-well reprogramming assays on days 3, 6 and 9. Twenty cells per well sorted at days 3, 6 and 9. Sox2–eGFP^{+} colonies were assayed on day 24. Asterisks indicate two-sided *t*-test $P < 0.05$. **d**, Plating efficiencies for enhanced green fluorescent protein (eGFP)-expressing MEFs sorted on day 6 for CD73^{high} or CD49d^{high} (assayed 24 h after sort). **e**, Single-cell reprogramming efficiencies for day 6 CD73^{high} or CD49d^{high} fractions. **f**, Reprogramming efficiencies (**e**) adjusted for plating efficiencies (**d**). Error bars, s.d.; $n = 3$ independent experiments for all assays.

(Fig. 3g). As expected, MEF markers were low in these poised populations. This suggests that this intermediate arises after loss of mesenchymal markers, but before completion of mesenchymal-to-epithelial transition, as indicated by reprogramming enrichment in the CD326^{low} fraction.

To identify subsequent reprogramming stages, we conducted continuation analysis where cells were sorted on day 6 and characterized by mass cytometry on day 16 (Fig. 3h and Extended Data Fig. 4–6). By day 10, reprogramming-prone populations formed distinct colonies with ESC-like morphology, while CD73^{low} cells were highly proliferative but failed to develop into mature colonies (Extended Data Fig. 5b). Continuation analysis on day 16 revealed that while reprogramming-prone and non-prone populations contained CD326-expressing cells, broad overlap between CD326^{high} and SSEA1^{high} clusters was only in mature reprogramming-prone populations (Fig. 3h and Extended Data Fig. 6). These clusters did not overlap with the ESC marker CD54, and were heterogeneous for CD73 and CD49d. We conclude a distinct CD326^{high}, SSEA1^{high}, CD54^{low} intermediate arises after the CD73^{high}/CD49d^{high} intermediate and before pluripotency acquisition.

We then used the intermediates stages to gain molecular insights into transcriptional regulation of early reprogramming. Gene expression analysis intriguingly showed these intermediates precede activation of the majority of transcription factors thought to be predictive markers for pluripotency induction (Fig. 4a and Extended Data Fig. 7)[2,7]. The observation that these intermediates arise before key pluripotency regulators suggests a separate combination of early transcription factors must be induced to generate early intermediates and poise them for pluripotency acquisition. We found the transcription factors *Nr0b1* and *Etv5* preferentially expressed in reprogramming-prone populations and highly expressed in ESCs, suggesting a functional role in poising early reprogramming (Fig. 4a and Extended Data Fig. 7b–d).

To assess whether these genes were necessary to induce the early intermediate populations, we generated three short hairpins against each gene (Extended Data Fig. 7e, f). We then assessed the ability of reprogramming MEFs infected with these short hairpins to induce CD73^{high}/CD49d^{high} intermediates 9 days after reprogramming induction (Fig. 4b and Extended Data Fig. 7g). Surprisingly, while MEFs infected with a control short hairpin were able to induce the CD73^{high}/CD49^{high} intermediate (6.70 ± 2.27%), reprogramming MEFs infected with short hairpins targeting *Etv5* or *Nr0b1* were significantly impaired (Fig. 4b). This phenotype could be rescued by complementary DNA (cDNA) overexpression in combination with a hairpin targeting the untranslated region for the gene of interest (Fig. 4c). Further, when Nanog^{+} colonies or Sox2–eGFP^{+} colonies were assessed 24 days after reprogramming induction, a dramatic decrease in reprogramming efficiencies was observed (Fig. 4d, e). In contrast, knockdown of either gene in ESCs did not affect survival or proliferation (Extended Data Fig. 7h). While this paper was under review, an independent report confirmed *Nr0b1* as necessary for reprogramming[24]. These data indicate that *Etv5* and *Nr0b1* are required to generate the CD73^{high}/CD49d^{high} poised intermediate necessary to induce the canonical pluripotency program and definitive iPS cell formation.

We then wondered whether a similar intermediate population arises in high-efficiency reprogramming systems[12,13]. Published expression analysis of two high-efficiency systems showed transient CD73 upregulation, suggesting the presence of a similar intermediate (Extended Data Fig. 8a, b). We then characterized one of these systems, the *Mbd3*$^{fl/-}$ secondary MEF system, in greater detail[12]. After confirming reported

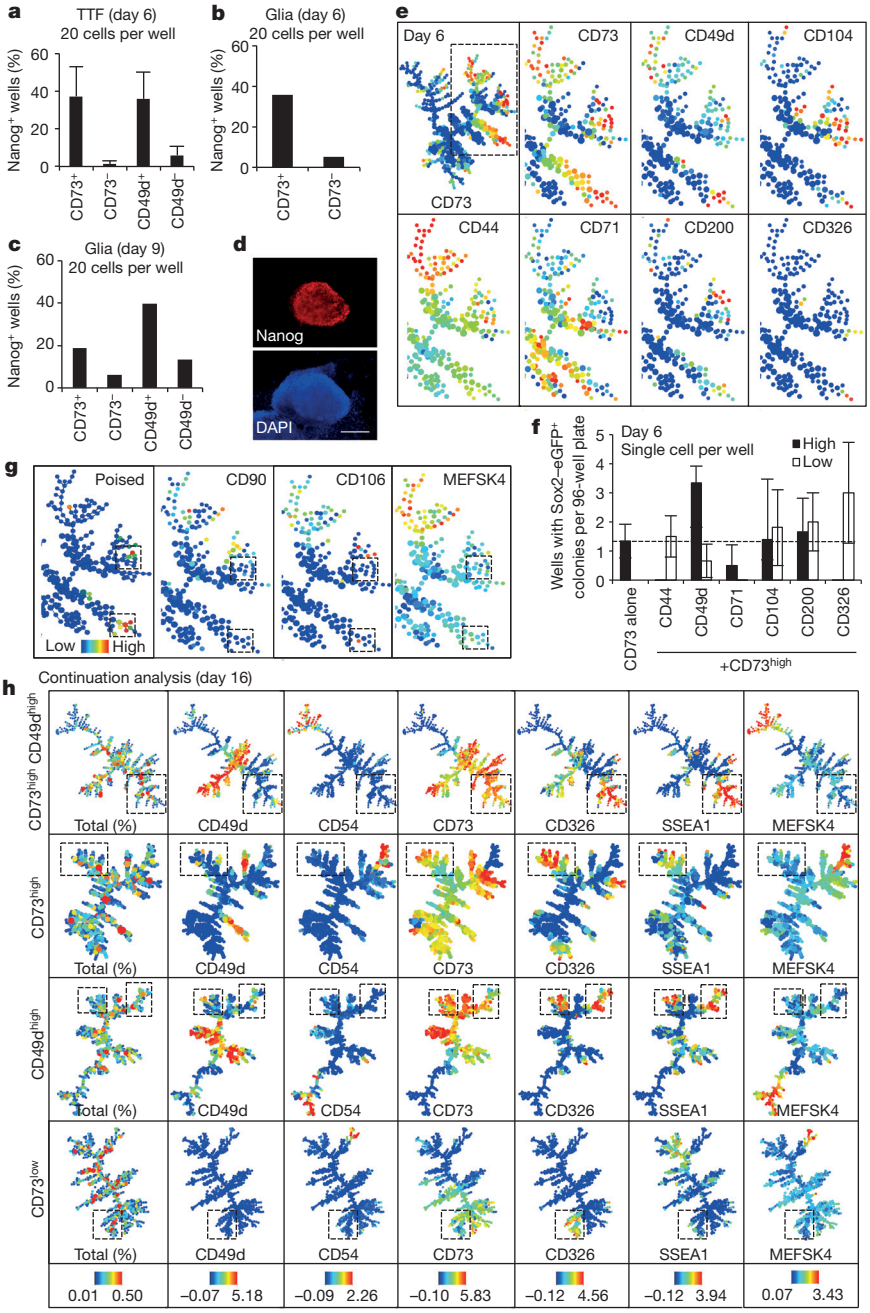


Figure 3 | Characterization of CD73^{high} and CD49d^{high} intermediates. **a**–**c**, Percentage of wells with *Nanog*-expressing cells on day 24 for reprogramming tail tip fibroblasts (TTF) (**a**) ($n = 3$) and glia (**b**, **c**) ($n = 1$, independent primary cells and infections). **d**, Representative Nanog immunostaining. Scale bar, 200 μm. **e**, Heterogeneously expressed markers in day 6 CD73^{high} population. **f**, Single-cell 96-well assays for day 6 CD73^{high} fraction with additional surface markers ($n = 3$). **g**, Refined poised signature. Clusters are low for mesenchymal markers. **h**, Continuation analysis shows SSEA1^{high} CD326^{high} branch unique to poised populations (boxed). All experiments represent independent biological replicates. Error bars, s.d.

reprogramming efficiencies, we analysed this system by mass cytometry (Extended Data Figs 8–10). By day 3, fibroblast marker repression was evident, and CD73 was upregulated within this population (Extended Data Figs 8e and 9b). Within the CD73^{high}/MEFSK4^{low} population, CD49d (*Itga4*) upregulation was not apparent, but we noticed the emergence of a separate integrin, CD104 (*Itgb4*). By day 4, the major CD73^{high} branch clearly overlapped with the CD104^{high} branch and persisted into day 5. SSEA1^{high} and CD326^{high} expression was present on day 4, but clear co-expression was not seen until day 5. By day 9, CD73 and CD104 expression was dramatically reduced while CD326 and SSEA1 expression remained high. These data demonstrate a transient CD73^{high}/CD104^{high} population arises after donor cell program repression and before ESC marker acquisition, even in a highly efficient reprogramming system. Similar to CD49d and CD73, CD104 is not highly expressed in ESCs (Fig. 1a). And similar to viral reprogramming, adenosine treatment abolished reprogramming in the Mbd3 reprogramming system, albeit only at late stages, whereas compounds affecting CD49d function had little effect (Extended Data Fig. 4f).

The stage-specific framework provided in this study bridges the previously unexplored gap between donor program silencing and pluripotent marker acquisition (Fig. 4f). We demonstrate a transient, 'poised' intermediate present across multiple reprogramming systems, suggesting a general property of iPSC reprogramming. We note that, similar to SSEA1, TRA-1-60 enriches for reprogramming-prone intermediates at

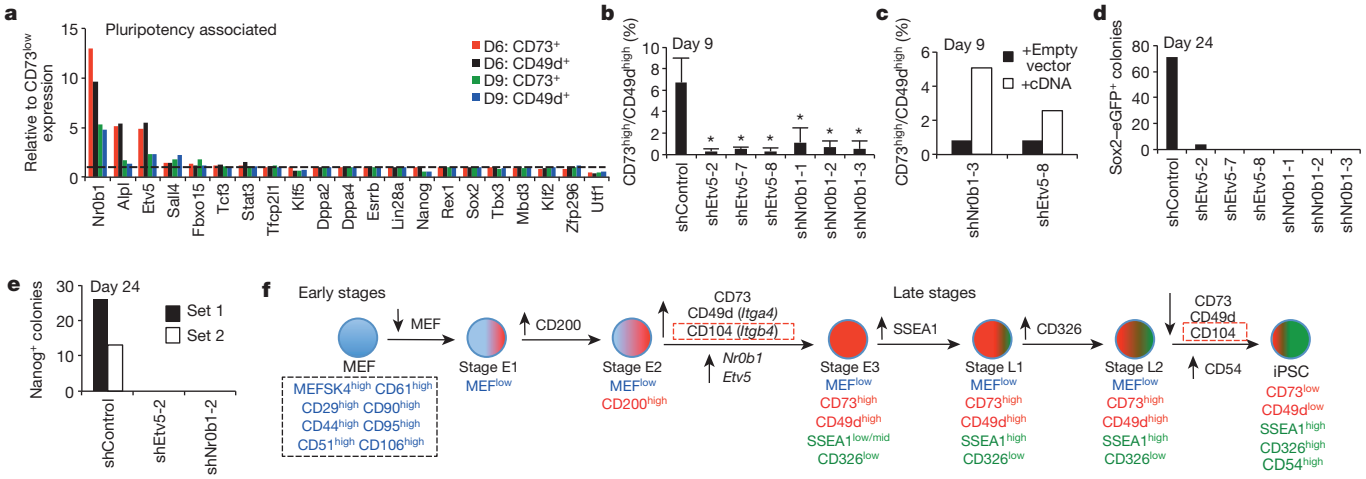


Figure 4 | Reprogramming regulators identified with CD73^{high}/CD49d^{high} intermediates. **a**, Day 6 and 9 reprogramming-prone and non-prone pluripotency-associated gene differential expression. Dotted line represents value of 1 (no difference). **b**, **c**, Day 9 CD73^{high}/CD49d^{high} quantification for knockdown (**b**, $n = 3$ independent experiments) and rescue (**c**) experiments. Gating shown in Extended Data Fig. 7g. Asterisks, two-sided *t*-test $P < 0.05$. **d**, **e**, Day 24 Sox2–eGFP^{+} (**d**, $n = 1$) and Nanog^{+} colonies (**e**, $n = 2$ independent experiments) from *Rosa-rtTA*$^{\pm}$, *Sox2-eGFP*$^{\pm}$ and *Rosa-rtTA*$^{\pm}$ MEFs, respectively. **f**, Early and late reprogramming model. Dotted red boxes distinguish CD104 observed in the *Mbd3*$^{fl/-}$ system. Error bars, s.d.

later time points during human reprogramming[11,25], and we speculate a similar transient intermediate arises in the human system.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March 2014; accepted 2 February 2015.
Published online 1 April; corrected online 20 May 2015 (see full-text HTML version for details).

1. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2,** 151–159 (2008).
2. 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).
3. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11,** 769–782 (2012).
4. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genet.* **45,** 34–42 (2013).
5. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).
6. Ho, R., Papp, B., Hoffman, J. A., Merrill, B. J. & Plath, K. Stage-specific regulation of reprogramming to induced pluripotent stem cells by Wnt signaling and T cell factor proteins. *Cell Rep.* **3,** 2113–2126 (2013).
7. O'Malley, J. *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* **499,** 88–91 (2013).
8. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–1632 (2012).
9. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).
10. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).
11. Tanabe, K., Nakamura, M., Narita, M., Takahashi, K. & Yamanaka, S. Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. *Proc. Natl Acad. Sci. USA* **110,** 12172–12179 (2013).
12. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).
13. Di Stefano, B. *et al.* C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506,** 235–239 (2014).
14. Hou, P. *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* **341,** 651–654 (2013).
15. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single cell mass cytometry. *Cell Stem Cell* **16,** 323–337 (2015).
16. Ichida, J. K. *et al.* A small-molecule inhibitor of Tgf-β signaling replaces *Sox2* in reprogramming by inducing *Nanog*. *Cell Stem Cell* **5,** 491–503 (2009).
17. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
18. Maherali, N. & Hochedlinger, K. Tgfβ signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.* **19,** 1718–1723 (2009).
19. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
20. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
21. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25,** 1177–1181 (2007).
22. Bandura, D. R. *et al.* Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81,** 6813–6822 (2009).
23. Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nature Biotechnol.* **29,** 886–891 (2011).
24. Zhang, J. *et al.* Dax1 and Nanog act in parallel to stabilize mouse embryonic stem cells and induced pluripotency. *Nature Commun.* **5,** 5042 (2014).
25. Takahashi, K. *et al.* Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nature Commun.* **5,** 3678 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Lovelace, R. Finck, K. M. Loh, K. Tanabe and S. Marro for advice. We thank J. Hanna for his the secondary *Mbd3*$^{fl/-}$ MEFs. We also thank S. Knöbel for the mEF-SK4 antibody. This work was supported by the California Institute of Regenerative Medicine grant RB2-01592 (to G.P.N.), the Institute for Stem Cell Biology and Regenerative Medicine at Stanford, and a New York Stem Cell Foundation-Robertson Investigator Award. E.L. was supported by the California Institute for Regenerative Medicine Predoctoral Fellowship TG2-01159 and National Science Foundation Graduate Research Fellowship DGE-114747. E.R.Z. was supported by National Institutes of Health National Research Service Award F32 GM093508-01. M.W. is a New York Stem Cell Foundation-Robertson Investigator and a Tashia and John Morgridge Faculty Scholar at the Child Health Research Institute at Stanford.

Author Contributions E.L., E.R.Z., G.P.N. and M.W. designed research. E.L. conducted reprogramming and sorting experiments. E.R.Z. conducted mass cytometry analysis and data processing. Y.H.N. and I.N.G. assisted with sample processing. E.L., E.R.Z., G.P.N. and M.W. analysed data. E.L., E.R.Z., G.P.N. and M.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W. (wernig@stanford.edu).

METHODS

No statistical methods were used to predetermine sample size.

Cell culture. Embryonic fibroblasts were isolated from embryonic day (E)13.5 embryos derived from B6;129S-*Sox2*$^{eGFP/+}$ mated with B6.Cg-*Gt(ROSA)26Sor*$^{tm1(rtTA*M2)Jae}$ (allele referred to as *Rosa-rtTA*) as previously described[26,27]. Tail tip fibroblasts were derived from 1-week-old B6 mice. This was done by chilling the animals on ice for 5 min, cutting the tail, and mincing in a 6 cm dish in 1 ml 0.25% trypsin. Another 1 ml of 0.25% trypsin was added after mincing and this was incubated at 37 °C for 10 min. The mixture was then resuspended in 15 ml MEF medium (described below) and plated on a 0.2% gelatinized 15 cm plate. Glial cultures were prepared from CD1 mice as previously described[26]. MEFs, tail tip fibroblasts and glial cells were cultured in MEF medium, which consisted of 10% cosmic calf serum (Thermo Scientific) in DMEM (Invitrogen) supplemented with non-essential amino acids (Invitrogen), penicillin–streptomycin (Invitrogen), sodium pyruvate (Invitrogen) and 2-mercaptoethanol (Invitrogen). Primary *Mbd3*$^{fl/-}$, *Oct4-GFP* secondary MEFs[12] were a gift from J. Hanna and grown in MEF medium. We verified modification of the *Mbd3* locus by western blot analysis (described below under 'Western blotting'; Extended Data Fig. 8f). Our partially reprogrammed line was previously characterized[15,21]. These were grown in MEF medium supplemented with 15% cosmic calf serum. *Oct4-neoR* knock-in mouse ESCs[19,21] were grown in mESC medium, which consisted of 12% knockout replacement serum (Invitrogen), 3% cosmic calf serum and supplemented with non-essential amino acids, penicillin–streptomycin, sodium pyruvate, 2-mercaptoethanol and leukaemia inhibitory factor (LIF). MEFs, PRCs and mESCs were routinely tested for mycoplasma contamination.

Surface marker screening. We screened mouse surface markers using the 'Mouse cell surface marker screening panel' Lyoplate (BD Biosciences, material number 562208) according to the manufacturer's instructions. Briefly, 150×10^6 cells were used for our partially reprogrammed cells (passage 15) and *Oct4-neoR* mES cells. Cells were plated on 0.2% gelatin-coated plates and treated with neomycin 3 days before screening; 175×10^6 cells were used for passage 4 *Sox2-eGFP* MEFs. Cells were washed with PBS-EDTA, dissociated in 10× TrypLE (Invitrogen) for 5 min, washed once in PBS before staining for 30 min in primary antibody in staining medium (PBS-EDTA supplemented with 0.5% BSA) according to manufacturer's instructions. After the primary antibody stain, cells were washed two times in PBS, stained for 30 min. on ice with biotinylated secondary antibodies in staining medium, washed two times in PBS, and then stained for 30 min. on ice with Alexafluor 647-conjugated streptavidin in staining medium. After the tertiary Alexafluor 647–streptavidin stain, the cells were washed twice in PBS and analysed on a BD LSR II Flow Cytometer with 96-well HTS module in staining medium. We used a high concentration of TrypLE to verify that all identified markers would not be cleaved by our dissociation reagent. We further validated a subset of these identified makers with our control populations dissociated in 1× TrypLE and stained with fluorophore-conjugated antibodies as described below.

Reprogramming. Reprogramming assays were conducted with passage 4 *Rosa-rtTA*$^{\pm}$; *Sox2*$^{eGFP/+}$ or *Rosa-rtTA*$^{\pm}$ (derived from the same litter: see mating scheme above) mouse embryonic fibroblasts, passage 2 B6 tail tip fibroblasts or passage 3 B6 glia as indicated. FUW-tetO lentiviral vectors and lentiviral packaging were used as previously described[26]. Passage 3 MEFs and passage 2 glial cells were split onto 10 cm 0.2% gelatin-coated plates at 250,000 cells per plate 1 day before infection. Passage 1 tail tip fibroblasts were split onto 10 cm 0.2% gelatin-coated plates at 100,000 cells per plate 1 day before infection. Cells were infected in MEF medium supplemented with polybrene (8 μg ml^{-1}; Sigma). One day after infection, MEF medium supplemented with doxycycline was added; this was considered to be day 0. Media were replaced every 2 days. On day 16, medium was replaced with mES cell medium without doxycycline. Owing to the complication of getting a large number of glial cells, we performed the reprogramming efficiency assays once per time point (days 6 and 9) but both were done with independently derived primary cells from different animals, independently derived virus and sorts.

The following compounds were also added to reprogramming medium on the days indicated in the text: BIO 1211 (Tocris, 4 nM), adenosine (Sigma, 20 μM or 2 μM as indicated), fibronectin (Sigma, 4 μg ml^{-1}), AMP-CP (adenosine 5′-(α,β-methylene)diphosphate) (Sigma, 20 μM). These were chosen as CD73 (5′-nucleotidase, ecto) converts AMP to adenosine, CD49d (*Itga4*) is an integrin whose substrates are VCAM1 and fibronectin, α,β-methyleneadenosine 5′-diphosphate (AMP-CP) is a CD73 inhibitor and BIO1211 is a CD49d inhibitor.

Hairpins for *Nr0b1* and *Etv5* were designed with the pSico oligomaker (Supplementary Table 3) and cloned into the lentiviral pSico-puro vector[28] (http://web.mit.edu/jacks-lab/protocols/pSico.html). We assessed knockdown efficiencies in partially reprogrammed cells as they express these genes and grow well in MEF medium supplemented with 15% serum alone. To assess knockdown efficiencies, 50,000 PRCs were plated onto a six-well gelatin-coated well 1 day before transduction. Cells were transduced with the hairpin of interest in MEF medium supplemented with 15% serum and polybrene (8 μg ml^{-1}). The medium was exchanged the following day with MEF medium supplemented with 15% serum and puromycin (2 μg ml^{-1}). Cells were cultured for a further 3 days before RNA extraction (for a total of 4 days after infection). RNA was prepared with an RNeasy purification kit as described below.

To assess the effects of knockdown of these genes on the day 9 CD73^{high}/CD49d^{high} intermediate, 100,000 P4 *Rosa-rtTA* MEFs were plated on a 10 cm gelatin-coated plate. These were transduced the following day with the doxycycline-inducible FUW-tetO-*Klf4*, *Oct3/4*, *Sox2* and *c-Myc* vectors and indicated hairpins as described above. Medium was supplemented with doxycycline 1 day after infection (day 0). We note that because *Rosa-rtTA* MEFs contain a PGK-puromycin cassette, they were not selected with puromycin. On day 9, CD73-Alexa 488 (1:50) and CD49d-Alexa 750 (1:50) were analysed by fluorescence-activated cell sorting (FACS staining described below). To assess reprogramming efficiencies for these genes, 30,000 P4 *Rosa-rtTA* MEFs were plated on a 6 cm gelatin-coated plate or 15,000 P4 *Rosa-rtTA*, *Sox2-eGFP* MEFs were plated on a six-well gelatin-coated well and transduced the following day in the same manner. At day 16, the culture medium was switched to mES cell medium without doxycycline. Twenty-four days after reprogramming induction, Nanog^{+} colonies were assessed for *Rosa-rtTA* MEFs (immunofluorescence staining described below) or Sox2–eGFP^{+} colonies for *Rosa-rtTA*, *Sox2-eGFP* MEFs. In total, three independent reprogramming efficient experiments were conducted across the two different MEF lines. To validate further the specificity of the effects, we performed a 'rescue' experiment to demonstrate that upon re-expression of the cDNA (under knockdown conditions) the effect is eliminated. To this end, the cDNAs for *Nr0b1* and *Etv5* were cloned into the FUW-tetO vector. Fifteen thousand P4 *Rosa-rtTA* MEFs on six-well gelatin-coated plates were then transduced with the doxycycline-inducible FUW-tetO-*Klf4*, *Oct3/4*, *Sox2* and *c-Myc* vectors and indicated hairpins with the FUW-tetO-cDNA (*Etv5* or *Nr0b1*) or an empty vector. Medium was supplemented with doxycycline 1 day after infection (day 0). On day 9, CD73-Alexa 488 (1:50) and CD49d-Alexa 750 (1:50) were analysed by FACS (FACS staining described below). As the rescue experiment was supplementary to and consistent with the main knockdown experiments, we conducted only one rescue experiment for one hairpin.

To verify that mouse ESCs could survive and proliferate after knockdown of *Etv5* or *Nr0b1*, we infected 30,000 ESCs per well in gelatinized six-well plates in mESC medium and replaced the medium the following day with mESC medium supplemented with puromycin to select for successful transduction. These were then cultured for 3 days, dissociated with 0.25% trypsin and re-plated onto gelatinized six-well plates. These were then cultured for 3 days, fixed and stained for *Oct4* (described below).

To reprogram *Mbd3*$^{fl/-}$, *Oct4-GFP* secondary MEFs (the *Oct4-GFP* allele in these cells is not a targeted, but well-characterized, transgenic reporter[12]), we first assayed reprogramming conditions with 2000 *Mbd3*$^{fl/-}$ secondary MEFs on 2×10^6 mitomycin- (Sigma) treated B6 feeders, 10^5 *Mbd3*$^{fl/-}$ secondary MEFs without feeders and 2×10^5 *Mbd3*$^{fl/-}$ secondary MEFs without feeders on 10 cm dishes coated with 0.2% gelatin. We found optimal reprogramming efficiencies with 2,000 *Mbd3*$^{fl/-}$ secondary MEFs and 2×10^6 feeders as previously reported[12]. All reprogramming assays were done in non-hypoxic conditions. To reprogram cells, cells were cultured in media 1 (recombinant human LIF (10 ng ml^{-1}, Peprotech), doxycycline (1 μg ml^{-1}) and ascorbic acid (10 μg ml^{-1}, Sigma) in 15% cosmic calf serum (Thermo Scientific) in DMEM (Invitrogen) supplemented with non-essential amino acids (Invitrogen), penicillin–streptomycin (Invitrogen), sodium pyruvate (Invitrogen) and 2-mercaptoethanol (Invitrogen)) for 3 days and then media 2 (recombinant human LIF (10 ng ml^{-1}), doxycycline (1 μg ml^{-1}), PD0325901 (1 μM, Cell Signaling) and CHIR99021 (3 μM, Cayman) in 15% knockout replacement serum (Invitrogen) in DMEM (Invitrogen) supplemented with non-essential amino acids (Invitrogen), penicillin-streptomycin (Invitrogen), sodium pyruvate (Invitrogen) and 2-mercaptoethanol (Invitrogen)), until the completion of the experiment. Media were exchanged every 2 days.

Mass cytometry analysis. On designated days the reprogramming cultures were treated with 1× TrypLE (Invitrogen) for 5 min at 37 °C, dissociated into single-cell suspension by trituration, and then washed twice with PBS. The cell samples were then incubated with metal-conjugated antibodies (Supplementary Table 1) in PBS containing 5% FBS (Omega Scientific) for 30 min on ice, washed once with PBS containing 5% FBS, treated with 25 μM cisplatin for 1 min on ice for live–dead cell discrimination[29], washed once with PBS containing 5% FBS and then fixed with 1.6% paraformaldehyde at 20 °C (room temperature) for 10 min. Formaldehyde-fixed cell samples were then permeabilized with methanol on ice for 15 min, washed once with PBS containing 0.5% BSA, and then incubated at room temperature for 15 min with an iridium-containing DNA intercalator (DVS Sciences/Fluidigm) in PBS containing 1.6% paraformaldehyde. After intercalation/fixation, the cell samples were washed once with PBS containing 0.5% BSA and twice with water before measurement on a CyTOF mass cytometer (DVS Sciences/Fluidigm).

Normalization for detector sensitivity was performed as previously described[30], using polystyrene normalization beads containing lanthanum-139, praseodymium-141, terbium-159, thulium-169 and lutetium-175. Number of cells for each mass cytometry experiment are shown in Supplementary Table 2.

SPADE analysis. Density-dependent downsampling, hierarchical clustering, cluster upsampling and extraction of parameter medians was performed by the SPADE package (www.cytospade.org) as described in the main text and previously[23,31]. All assayed surface markers were used in the clustering step unless otherwise indicated, and the parameters for downsampling percentile and target number of clusters were set to 5% and 500, respectively.

The refined poised signature shown in Fig. 3g was determined by calculating the similarity of each SPADE cluster to the hand-gated CD73^{high} CD49d^{high} CD44^{low} CD326^{low} population. Similarity was calculated by the Manhattan distance metric using all measured surface markers, and is indicated by the coloured scale bar (low distance equals high similarity).

Immunofluorescence. Plates were fixed in 4% PFA for 10 min, washed three times with PBS, blocked and permeabilized in PBS supplemented with 5% CCS and 0.1% Triton-X 100 (Sigma) (blocking solution) for 10 min. Ninety-six-well plates were then incubated with mouse anti-Nanog (1:500, BD) or mouse anti-Oct4 (1:200, Santa Cruz Biotechnology) in blocking solution for 30 min, washed three times with PBS, incubated with donkey anti-mouse Alexa-555 (1:1,000, Invitrogen) or anti-mouse Alexa-488 (1:1,000, Invitrogen) in blocking solution for 30 min, washed three times with PBS and stained with 4′,6-diamidino-2-phenylindole (DAPI) for 3 min. Cells were then washed with PBS and visualized.

FACS and efficiency assays. Cells were washed with PBS-EDTA, dissociated in 1× TrypLE for 5 min, washed with PBS and incubated on ice with a fluorophore-conjugated antibody and DAPI for 30 min in PBS-EDTA supplemented with 0.5% BSA. The sources and detailed descriptions of all antibodies used are listed in Supplementary Table 1. For 96-well assays, single or 20 DAPI^{-} cells per well were double sorted on the indicated day into gelatinized 96-well plates supplemented with 400,000 feeders per plate in MEF medium supplemented with doxycycline. For the primary sort, cells were sorted into PBS supplemented with 0.5% BSA; for the secondary sort, cells were sorted directly into 96-well plates. Efficiency assays were conducted 24 days after transgene induction and determined by the number of wells with Sox2–eGFP^{+} colonies. For mass culture reprogramming, 10,000 SSEA1^{high} or SSEA1^{low} cells were double sorted 6 days after transgene induction onto 3 cm gelatinized plates supplemented with feeders, and Sox2–eGFP^{+} colonies were assayed 24 days after transgene induction. We used the same SSEA1 clone and vendor as previously used for determining reprogramming efficiencies[8,10]. Tail tip fibroblast and glial reprogramming efficiencies were determined by double sorting on the indicated days into 96-wells (as described above) and assaying for *Nanog* by immunofluorescence. Plating efficiencies were determined by infecting *Sox2-eGFP* MEFs with FUW-tetO-*hygroB-T2A-eGFP*, selecting for 5 days in hygromycin, and counting the number of wells with GFP^{+} cells 24 h after sorting. For all assays where CD73 was used for sorting, CD73-Alexa 647 was used (gating shown in Extended Data Fig. 4). For CD73 × CD49d analysis, CD73–Alexa 488 was used (Extended Data Fig. 7g).

RNA preparation and expression analysis. RNA of reprogramming populations for microarray analysis was prepared from *Rosa-rtTA*$^{\pm}$ day 6 and day 9 reprogramming cultures double sorted for CD73 or CD49d (as described above). RNA of control populations for microarray analysis was prepared from *Rosa-rtTA*$^{\pm}$ MEFs (passage 4), partially reprogrammed cells (passage 10) and V6.5 mouse ESCs (passage 11). RNA was prepared with RNeasy Mini Kit (Qiagen) and DNA was removed by on-column RNase-Free DNase treatment (Qiagen) according to the manufacturer's instructions. Mouse Gene 2.0 ST Arrays (Affymetrix) were prepared by the Stanford Protein and Nucleic Acid Facility. Data were normalized and gene names were assigned by Partek Genomic Suite. For all analysis, non-coding transcripts were removed. Preprocessing (floor = 100, ceiling = 20,000, min fold change = 2), *k*-means clustering ($k = 5$, seed value = 12345) and hierarchical clustering of *k*-means clusters (Pearson correlation, pairwise complete-linkage), and heat maps were generated by Gene Pattern (http://www.broadinstitute.org/cancer/software/genepattern/). Microarray data can be accessed with accession number GSE62957 from the National Center for Biotechnology Information database.

For Fig. 4a, genes were selected on the basis of pluripotency-associated genes characterized in previous studies[2,7] or from differential expression of sorted populations and ESCs. *Oct4* is not shown as the probe failed to detect expression in mESCs.

For quantitative PCR analysis, cDNA was generated with a SuperScript First-Strand Synthesis System (Invitrogen). Data were generated with a 7900HT Real-Time PCR System (Applied Biosystems). Six-microlitre reactions were prepared with SYBR Green Real-Time PCR Master Mix (Life Technologies) under the following conditions: 50 °C for 2 min, 95 °C for 10 min, and 40 cycles of 15 s at 95 °C and 1 min at 60 °C. All expression was normalized to GAPDH before comparing with control hairpin expression levels. Supplementary Table 3 gives primer sequences.

Western blotting. Passage 3 *Rosa-rtTA*$^{\pm}$ MEFs or secondary *Mbd3*$^{fl/-}$, *Rosa26-CreER* MEFs were grown in 10-cm tissue culture plates. Secondary *Mbd3*$^{fl/-}$ MEFs were treated with 1 μM 4OH-tamoxifen for 24 h, then samples were cultured for a further 48 h and dissociated with 0.25% trypsin. The cell pellet was then lysed with cell lysis buffer (200 mM NaCL, 50 mM Tris pH8.0, 1% Trion X-100, 5% glycerol). Twenty micrograms of soluble protein was run on a 4–12% gradient Bis-Tris gel (Life Technologies) and blotted onto PVDF membrane. After blocking, membrane was incubated with primary antibody against Mbd3 (1:1,000, Bethyl Laboratories, A302-528A) for 1 h at room temperature, washed three times with PBS with 0.1% Tween-20 and incubated with secondary antibody anti-rabbit HRP (1:5,000, Jackson Immuno, A5441) for 1 h at room temperature.

26. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463,** 1035–1041 (2010).
27. Ellis, P. *et al.* SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev. Neurosci.* **26,** 148–165 (2004).
28. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33,** 401–406 (2003).
29. Fienberg, H. G., Simonds, E. F., Fantl, W. J., Nolan, G. P. & Bodenmiller, B. A platinum-based covalent viability reagent for single-cell mass cytometry. *Cytometry* **81,** 467–475 (2012).
30. Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytometry* **83,** 483–494 (2013).
31. Linderman, M. D. *et al.* CytoSPADE: high-performance analysis and visualization of high-dimensional cytometry data. *Bioinformatics* **28,** 2400–2401 (2012).

a

MEF Enriched

Antibody	MEF	PR	ESC
CD106	93.91	0.27	1.67
CD61	80.94	0.19	3.08
CD51	94.2	12.4	5.05
CD119	68.57	1.41	9.17
CD172a	66.72	0.59	15.46
CD140a	52.58	0.17	16.35
CD124	19.61	0.18	1.69
CD95	19.36	0.25	1.53
CD120A	16.82	0.27	1.52
IFNgRBchain	15.92	0.96	3.21

Partially Reprogrammed Enriched

Antibody	MEF	PR	ESC
CD73	24.1	99.19	41.06
CD80	12.39	98.24	3.72
CD200	3.58	75.98	7.24
Ly-6G/Lys-6C	1.37	12.51	2.84

ESC Enriched

Antibody	MEF	PR	ESC
CD326	0.45	19.49	91.38
CD26	12.98	0.42	90.58
CD54	26.85	0.25	87.87
NK-T/NK-Ant	3.57	0.18	78.2
CD117/c-Kit	0.68	0.87	73.93
SSEA1	0.47	5.59	53.49
CD13	19.49	0.25	49.3
T/B Act.Ant.	0.81	0.55	20.72
MAD-CAM-1	0.79	0.08	10.55

MEF-Partially Reprogrammed Enriched

Antibody	MEF	PR	ESC
CD44	98.27	98.17	3.75
CD49d	76.78	79.94	10.68
Sca-1/Ly5-A/E	82.49	99.4	21.89

MEF-ESC Enriched

Antibody	MEF	PR	ESC
CD38	59.24	0.57	78.85
CD90.2/Thy1	91.92	1.14	52.3
MAC-3	8.8	2.55	22.07

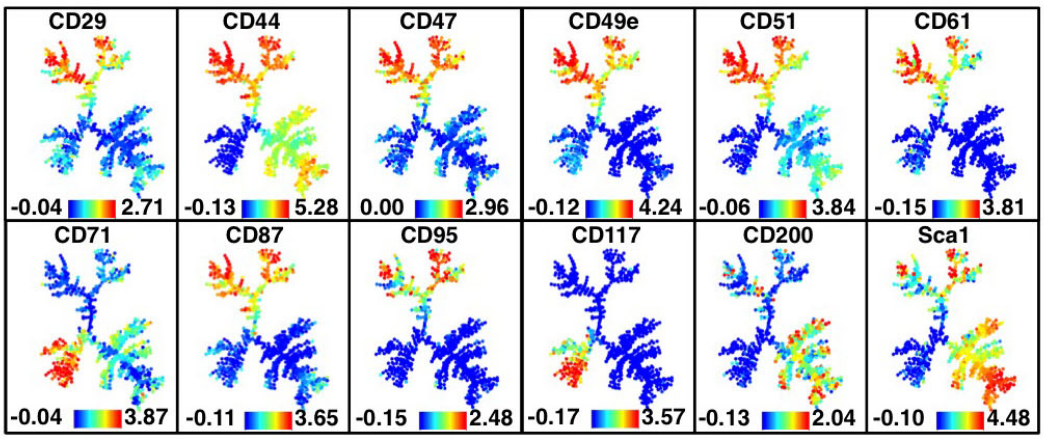
PR-ESC Enriched

Antibody	MEF	PR	ESC
CD71	4.31	44.37	67.96
CD104	0.27	49.12	40.67

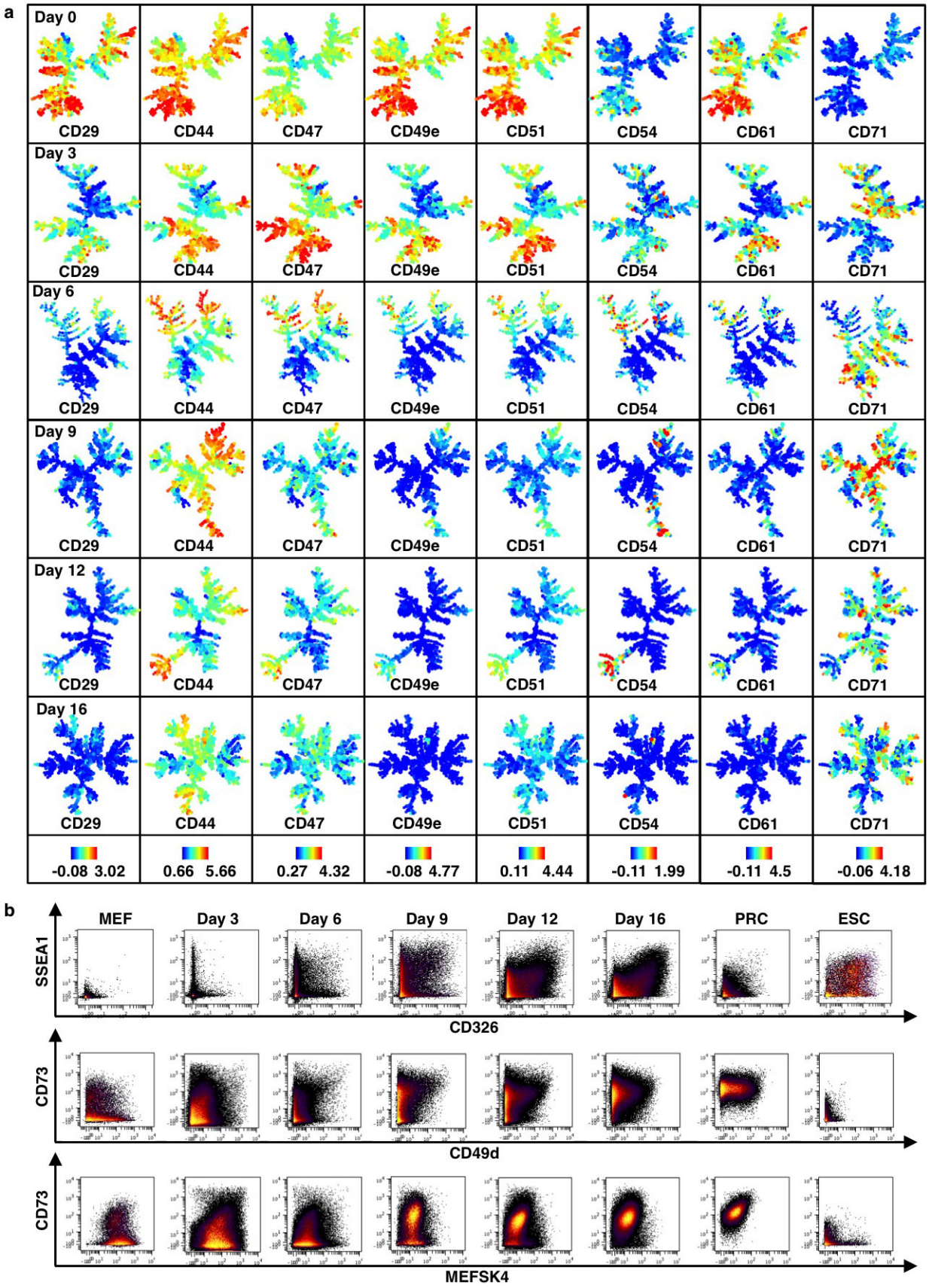
MEF-PR-ESC Enriched

Antibody	MEF	PR	ESC
CD9	92.67	98.4	98.37
CD24	78.82	99.19	71.00
CD29	93.12	95.38	99.21
CD47	79.66	98.21	99.12
CD49e	90.6	94.98	98.84
CD81	94.83	98.98	97.44
CD98	92.19	99.25	99.36
CD147	94.01	99.1	99.19
Crry/p65	92.13	99.22	98.9

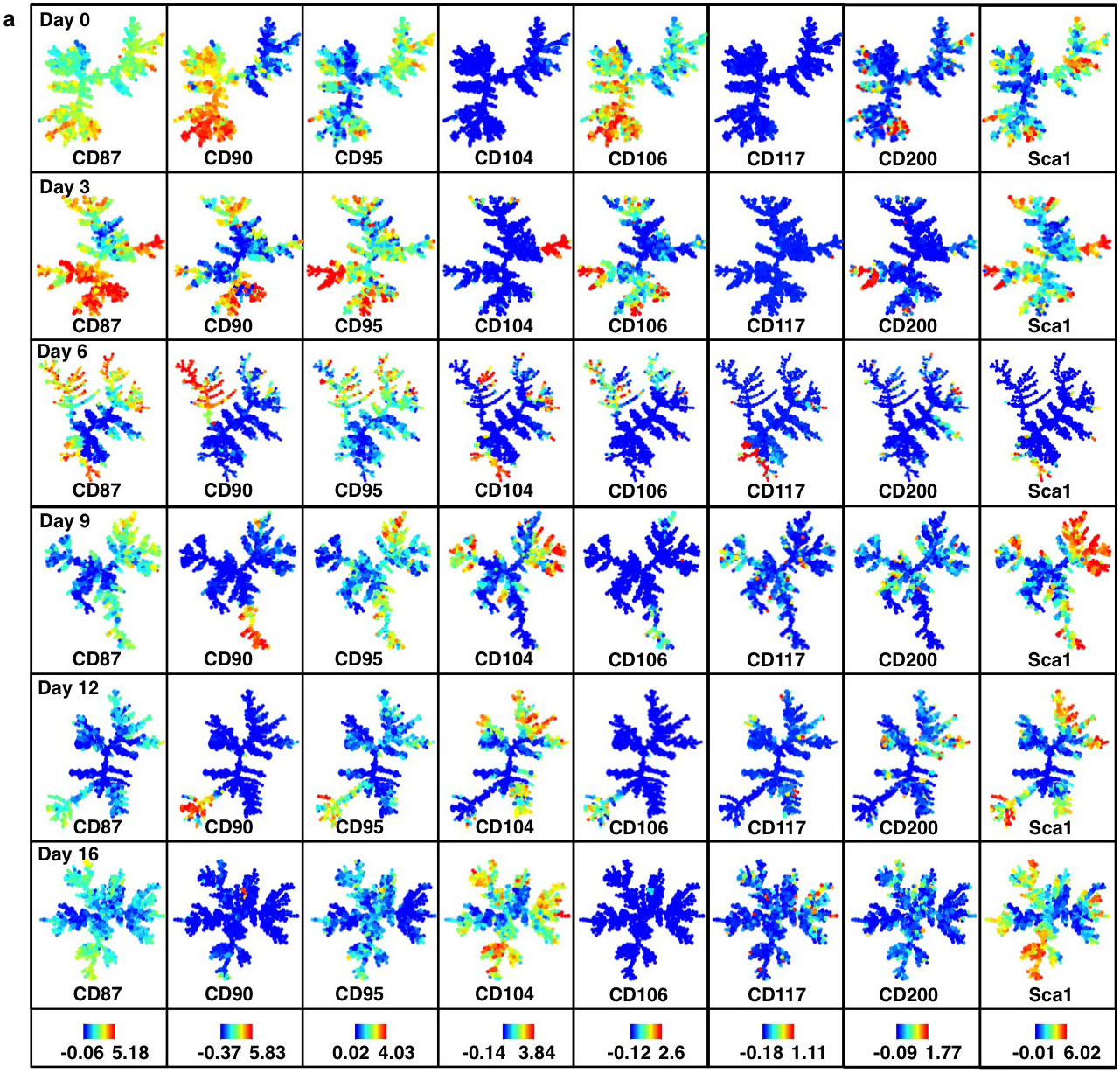
b



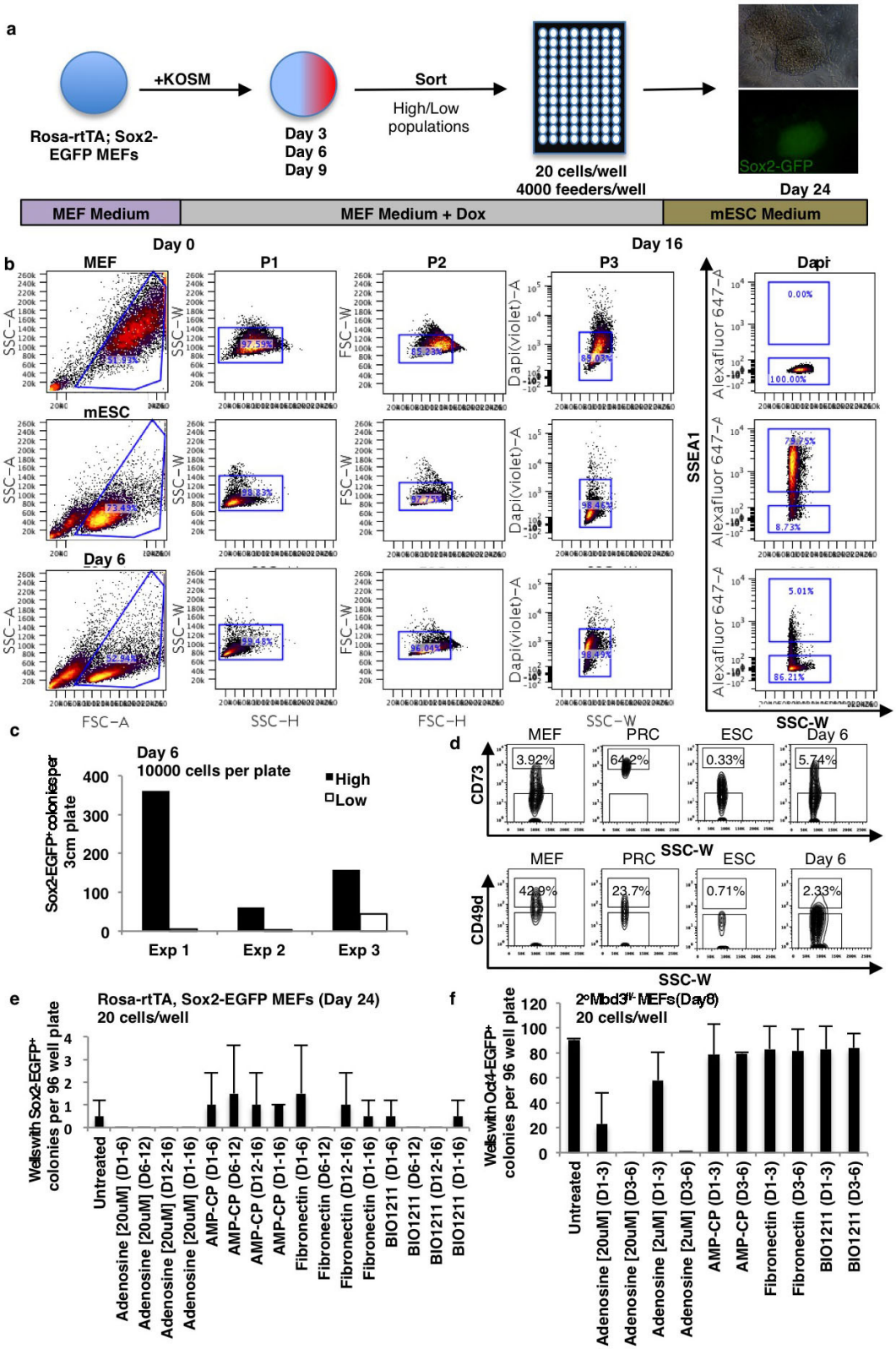
Extended Data Figure 1 | Results from surface marker screen. **a**, Shown are surface markers detected in MEFs, partially reprogrammed cells (PR) or ESCs analysed by flow cytometry. Numbers indicate the percentage of each population positive for the marker of interest, relative to isotype control samples. Markers are grouped for enrichment in single populations or shared between multiple populations. **b**, SPADE analysis for MEFs, mESCs and PRCs for surface markers analysed by mass cytometry (continued from Fig. 1b). Colour bars (bottom) represent ArcSinh-transformed counts for each marker.



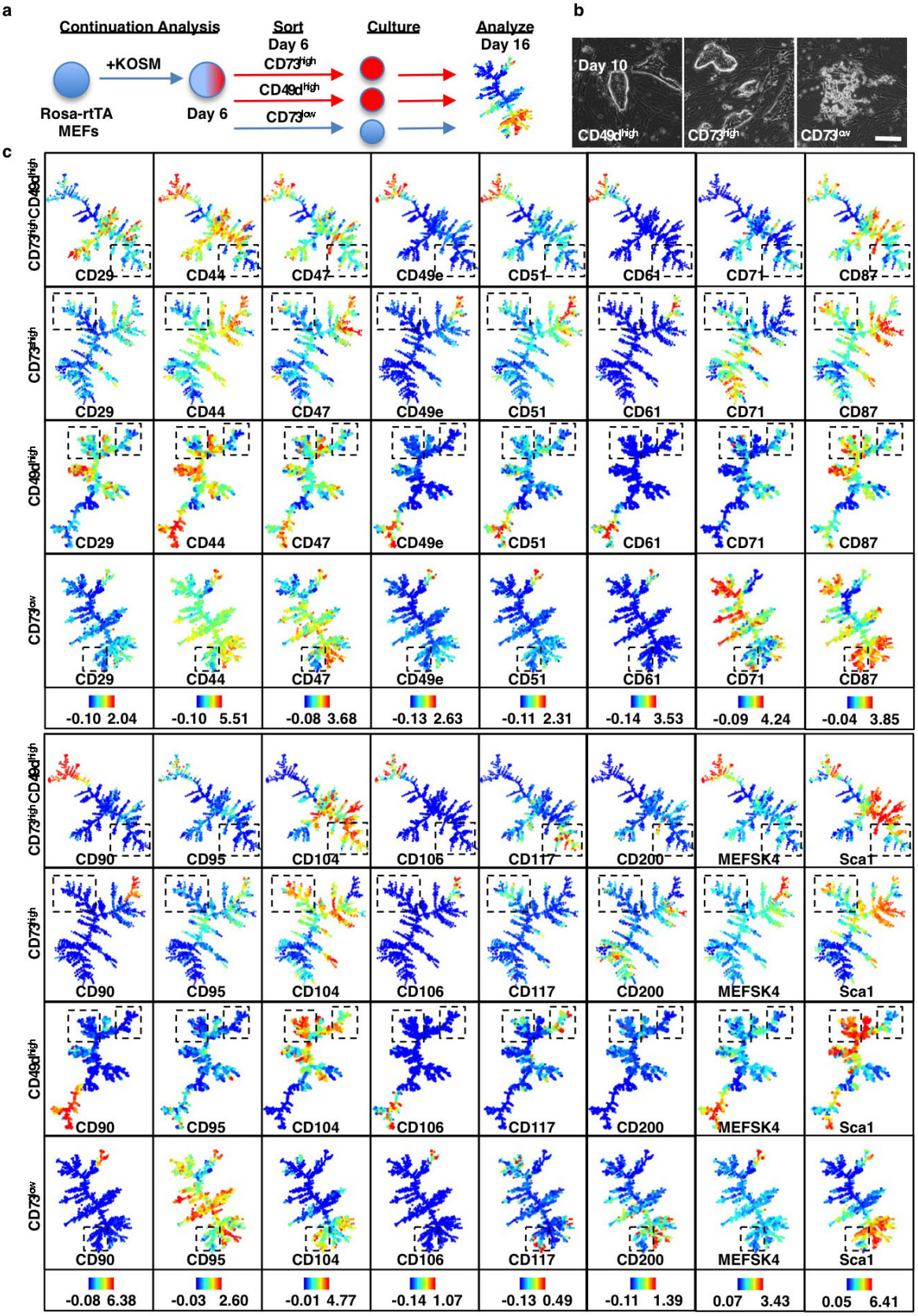
Extended Data Figure 2 | SPADE and biaxial analysis for MEF reprogramming. **a**, SPADE analysis of lentiviral-infected MEF reprogramming populations analysed by mass cytometry (continued from Fig. 1c). Colours bars (bottom) represent ArcSinh-transformed counts for each marker. **b**, Biaxial plots for selected markers in control populations and during MEF reprogramming.



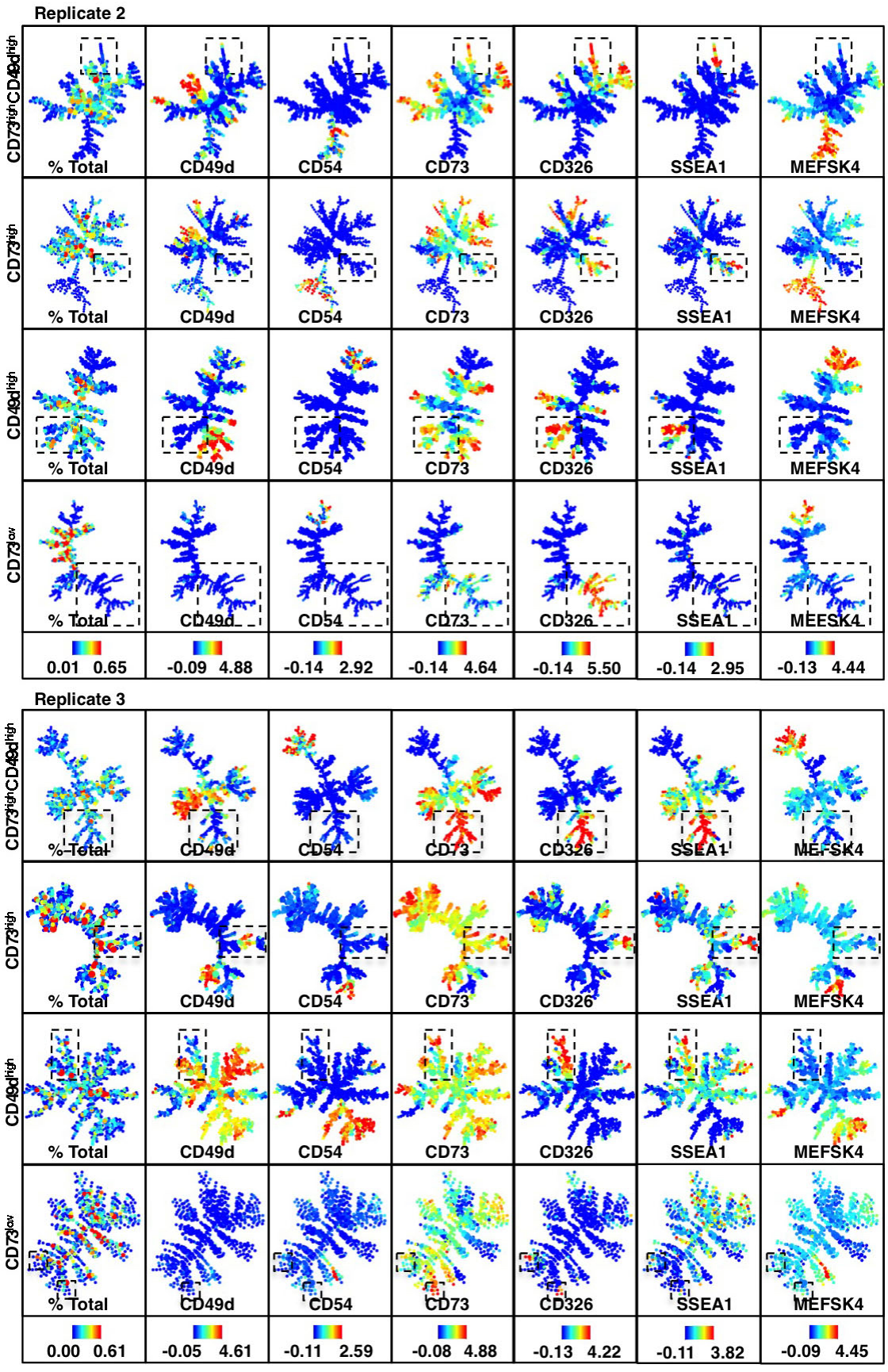
Extended Data Figure 3 | SPADE analysis for MEF reprogramming. **a**, SPADE analysis of lentiviral-infected MEF reprogramming populations analysed by mass cytometry (continued from Fig. 1c). Colours bars (bottom) represent ArcSinh-transformed counts for each marker. **b**, Day 12 and 16 time points for markers shown in Fig. 1c). Coloured bars for percentage total represent absolute percentages.



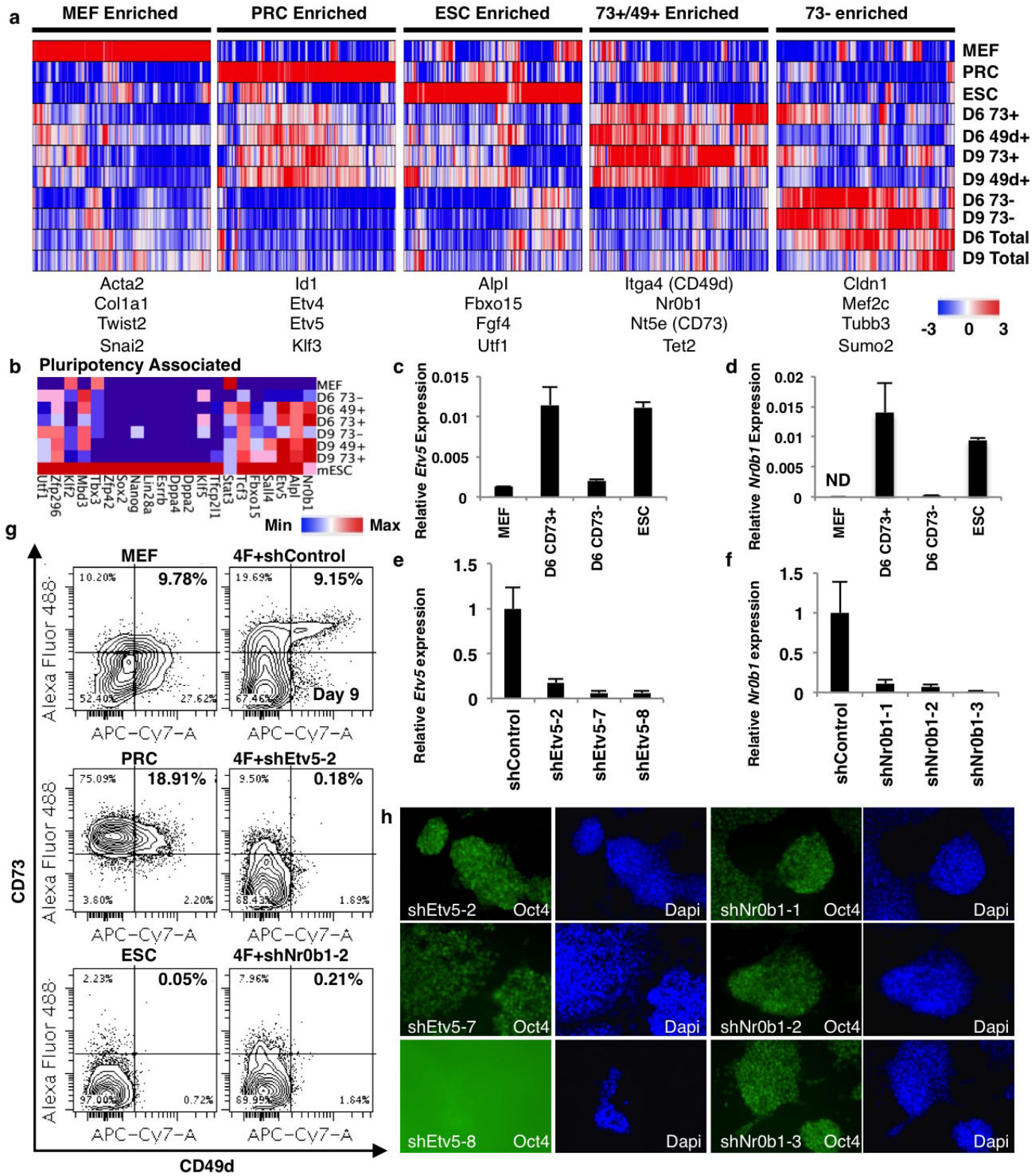
Extended Data Figure 4 | Details for sorting experiments and chemical treatment assay. **a**, Ninety-six-well reprogramming assay. Twenty cells per well sorted at days 3, 6 and 9. Sox2–eGFP^{+} colonies were assayed on day 24. **b**, Gating strategy for SSEA1 in controls and day 6 reprogramming population. High- and low-expressing populations were determined on the basis of MEF and ESC control levels. **c**, Ten thousand SSEA1^{high} (black bar) or SSEA1^{low} (white bar) were sorted onto 3 cm gelatinized plates with feeders. Sox2–eGFP^{+} colonies were counted on day 24 ($n = 3$ independent experiments). **d**, Gating strategy for CD73 and CD49d in controls and day 6 reprogramming population. High- and low-expressing populations were determined on the basis of MEF, PRC and ESC control levels. **e**, **f**, Treatment of reprogramming populations with compounds affecting CD73 and CD49d. Shown are 96-well reprogramming efficiencies for infected Rosa-rtTA Sox2–eGFP MEFs (**e**) or secondary *Mbd3*$^{fl/-}$ MEFs (**f**). The *y* axis displays wells with Sox2–eGFP^{+} colonies 24 days after infection (**e**) or wells with Oct4–eGFP^{+} colonies 8 days after transgene induction (**f**) and treated with the indicated compounds for the days (D) indicated ($n = 2$ independent experiments).



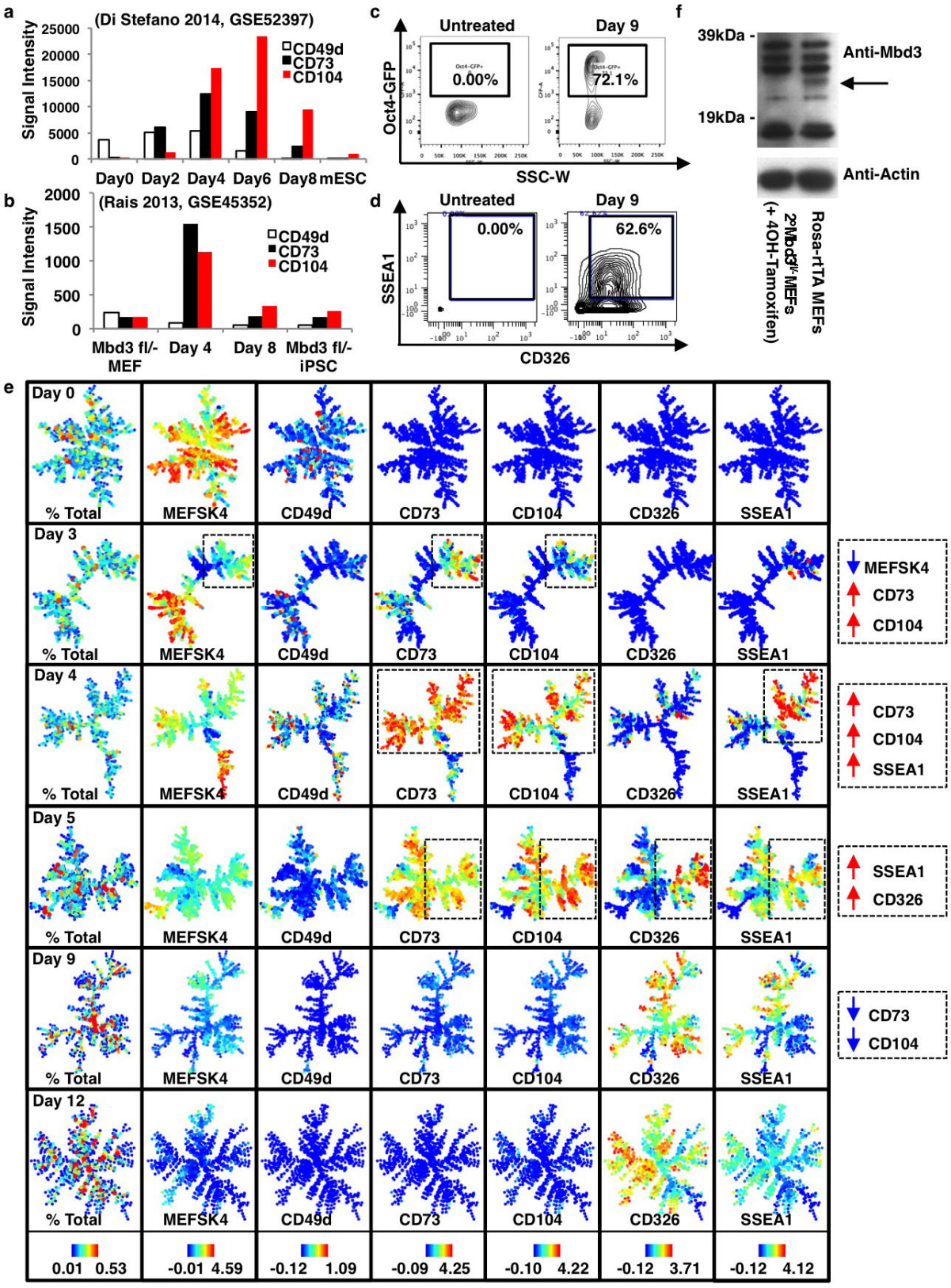
Extended Data Figure 5 | Day 6 continuation analysis on day 16. **a**, Schematic of continuation analysis. Reprogramming populations were sorted for poised ($CD73^{high}$ or $CD49^{high}$) and non-poised ($CD73^{low}$) populations on day 6, cultured for 10 days on a 3 cm plate and analysed by mass cytometry on day 16. **b**, Morphology of $CD49d^{high}$, $CD73^{high}$ or $CD73^{low}$ cells sorted on day 6 and inspected on day 10. Poised $CD49d^{high}$ or $CD73^{high}$ cells form compact colonies within several days of sorting while non-poised $CD73^{low}$ cells fail to do so. **c**, SPADE analysis (day 16) of cells sorted at day 6 for $CD73^{high}/CD49d^{high}$ $CD73^{high}$, $CD49^{high}$ and $CD73^{low}$ expression (continued from Fig. 3h). Boxes highlight a $SSEA1^{high}$ $CD326^{high}$ branch that is unique to the poised populations. Colours bars (bottom) ArcSinh-transformed counts for each marker.



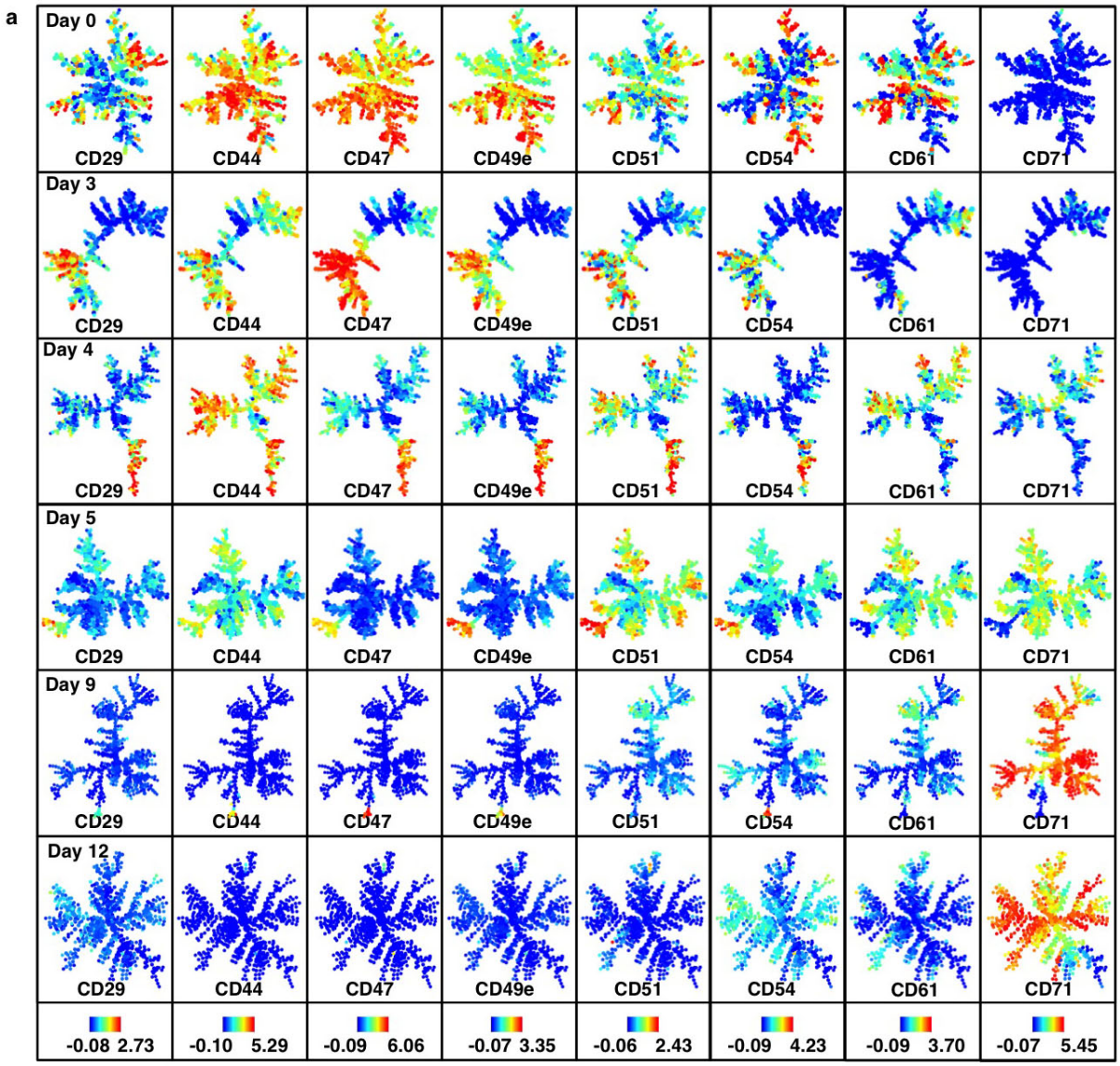
Extended Data Figure 6 | Continuation analysis replicates confirm a SSEA1^{high} CD326^{high} branch that is unique to poised populations. Continuation analysis replicates for reprogramming-prone (CD73^{high}/CD49d^{high}, CD73^{high}, CD49d^{high}) and non-prone (CD73^{low}) populations. Boxes highlight a SSEA1^{high} CD326^{high} branch that is unique to the poised populations. Colours bars (bottom) represent absolute percentages (left panel) and ArcSinh-transformed counts for each marker.



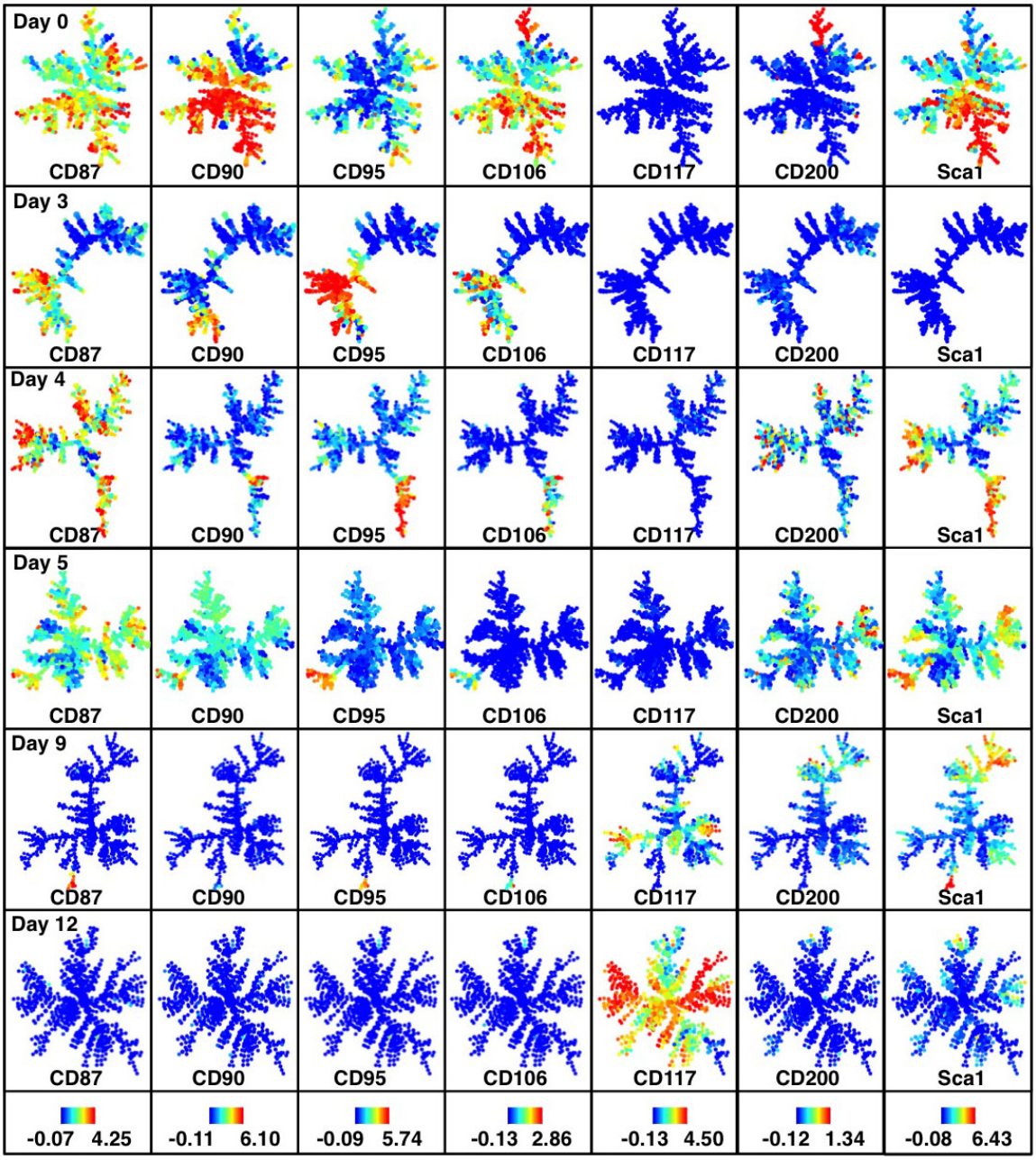
Extended Data Figure 7 | Molecular characterization of reprogramming-prone intermediates. **a**, Genes differentially expressed between reprogramming-prone (day 6 or day 9 CD73^{high} or CD49d^{high}) and non-prone (CD73^{low}) populations. Genes with more than twofold differential expression between reprogramming-prone and non-prone were selected and *k*-means clustered ($k = 5$) with control and total reprogramming population expression values. **b**, Heat map of pluripotency-associated genes shown in Fig. 4a ($\log_2$). **c**, **d**, Quantitative PCR verification of *Etv5* (**c**) and *Nr0b1* (**d**) expression levels ($n = 3$ technical replicates). **e**, **f**, *Etv5* (**e**) and *Nr0b1* (**f**) knockdown qPCRs ($n = 3$ technical replicates). **g**, Representative FACS plots for day 9 CD73^{high}/CD49d^{high} quantification shown in Fig. 4b. **h**, Demonstration of ESC self-renewal after infection with *Etv5* and *Nr0b1* hairpins All infected ESCs continue to express *Oct4* after passaging except ESCs infected with sh*Etv5–8* ($n = 1$).



Extended Data Figure 8 | Characterization of high-efficiency reprogramming systems. **a**, **b**, Expression analysis for CD49d, CD73 and CD104 for previously reported highly efficient reprogramming systems generated by transient expression of C/EBPα[13] (**a**) or *Mbd3* depletion[12] (**b**). **c**, *Oct4-GFP* transgene reporter signal and **d**, SSEA1 and CD326 levels for the $Mbd3^{fl/-}$ secondary reprogramming MEFs for untreated (left) and 9 days after induction (right). **e**, SPADE analysis for reprogramming $Mbd3^{fl/-}$ secondary MEFs at days 0, 3, 6, 9 and 12 using all surface markers by mass cytometry. Percentage totals of cells and representative markers are shown for each time point. Remaining markers are shown in Extended Data Figs 9 and 10. Colours bars represent absolute percentages (left) and ArcSinh-transformed counts for each marker. **f**, Verification of Mbd3 loss in passage 3 *Rosa26-CreER*, $Mbd3^{fl/-}$ secondary MEFs after treatment with 4OH-tamoxifen. Mbd3 levels were compared with passage 3 $Rosa\text{-}rtTA^{\pm}$ MEFs. While there are several unspecific bands, there is clearly one band around the expected size of Mbd3 absent in 4OH-tamoxifen-treated cells (arrow).



Extended Data Figure 9 | SPADE and biaxial analysis for secondary *Mbd3*$^{fl/-}$ MEFs. **a**, SPADE analysis for 2° *Mbd3*$^{fl/-}$ MEF reprogramming populations (continued from Extended Data Fig. 8e). Colour bars (bottom) represent ArcSinh-transformed counts for each marker. **b**, Biaxial plots for selected markers.



Extended Data Figure 10 | SPADE analysis for secondary *Mbd3*$^{fl/-}$ MEFs. SPADE analysis for 2° *Mbd3*$^{fl/-}$ reprogramming populations (continued from Extended Data Fig. 8e). Colours bars (bottom) represent ArcSinh-transformed counts for each marker.