Deconstructing transcriptional heterogeneity in pluripotent stem cells

Roshan M. Kumar1,2*, Patrick Cahan3*, Alex K. Shalek4, Rahul Satija3, Ajay DaleyKeyser1, Hu Li5, Jin Zhang5, Keith Pardee1,2, David Gennert5, John J. Trombetta5, Thomas C. Ferrante1, Aviv Regev5,7, George Q. Daley3 & James J. Collins1,2

Pluripotent stem cells (PSCs) are capable of dynamic interconversion between distinct substates; however, the regulatory circuits specifying these states and enabling transitions between them are not well understood. Here we set out to characterize transcriptional heterogeneity in mouse PSCs by single-cell expression profiling under different chemical and genetic perturbations. Signalling factors and developmental regulators show highly variable expression, with expression states for some variable genes heritable through multiple cell divisions. Expression variability and population heterogeneity can be influenced by perturbation of signalling pathways and chromatin regulators. Notably, either removal of mature microRNAs or pharmacological blockage of signalling pathways drives PSCs into a low—noise ground state characterized by a reconfigured pluripotency network, enhanced self-renewal and a distinct chromatin state, an effect mediated by opposing microRNA families acting on the Myc/Lin28/let-7 axis. These data provide insight into the nature of transcriptional heterogeneity in PSCs.

PSCs are defined by their unique capacity to differentiate into all the cell types of an organism, while self-renewing in culture. How PSCs reconcile pluripotency and self-renewal and decide among fate choices is a topic of intense interest, with relevance for regenerative medicine and developmental biology. Genomic maps of the regulatory circuits underlying pluripotency reveal a network of sequence-specific autoregulatory transcription factors targeting self-renewal genes that are active in PSCs, as well as repressed lineage-specific developmental regulators that exist in a poised state and are capable of driving cells towards differentiated fates1–3. These core transcription factors are thought to interact with chromatin modifiers, non-coding RNAs and external signalling pathways to maintain pluripotency. This self-sustaining transcriptional program becomes reactivated during reprogramming of somatic cells to pluripotency4.

The discoveries that levels of Nanog and other key PSC regulators fluctuate over time, that PSCs exist in multiple interconvertible states, and that distinct subpopulations of PSCs vary in their capacity to self-renew or differentiate, hint at the dynamism of the PSC transcriptional program5–13, which may be fundamental to pluripotency14–21. Here, we apply single-cell analytics to PSCs subjected to a range of perturbations to systematically dissect the factors underlying PSC heterogeneity. By doing so, we map the structure of gene expression variability in PSCs and identify regulatory circuits governing transitions between pluripotent cell states.

The landscape of gene expression variability in PSCs

To gain insight into the distinct substates of pluripotency, we used single-cell RNA-seq21–23 to characterize the transcriptome of 183 individual mouse embryonic stem cells (mESCs) grown under standard culture conditions, in the presence of serum and leukaemia inhibitory factor (LIF) (Extended Data Fig. 1, Supplementary Information and Supplementary Tables 1–3). Most cells (~92%) grouped together by principal component and cluster analysis, while 14 cells (8%) were characterized by reduced expression of fluctuating pluripotency regulators, which may indicate a distinct poised state (Extended Data Fig. 2 and Supplementary Information).

Some transcripts were detected in the vast majority of cells examined and showed a log-normal distribution of transcript abundance within the population, as for the core pluripotency regulator Oct4 (also called Pou5f1) (Fig. 1a). Other transcripts had bimodal expression, present in some cells and absent in others, as for the pluripotency regulator Esrb (Fig. 1a). Another set of genes displayed sporadic expression—that is, undetected in most cells but exhibiting relatively high expression in several cells—as for the lineage regulator and Polycomb target gene Neural1 (Fig. 1a, Extended Data Figs 3 and 4 and Supplementary Information). Expression distributions for 18 transcripts were validated by single-molecule fluorescence in situ hybridization (smFISH; Fig. 1b and Extended Data Fig. 3). Stable and variable regulators were also identified in neural precursor cells derived from mESCs (Supplementary Information, Supplementary Table 4 and Extended Data Fig. 4), suggesting that this may be a general feature of progenitor cell regulatory networks.

We examined functional gene categories for gene sets that showed more uniform or noisier expression in PSCs as compared to control gene sets. Genes involved in housekeeping and metabolic functions displayed relatively uniform expression, while previously identified targets26 of the Polycomb family of epigenetic regulators in PSCs exhibited greater variability (Fig. 1c and Extended Data Fig. 4). The Polycomb target genes include many developmental regulators and signalling factors governing lineage specification, and are thought to exist in a repressed yet poised state with a unique chromatin signature in PSCs26,27.
Figure 1 | Gene expression variability landscape of PSCs. a, Histograms of transcript distributions from single-cell RNA-seq of v6.5 mESCs cultured in serum+LIF. Arrow indicates high Neurod1-expressing cells. TPM, transcripts per million. b, Histograms and representative images of transcript distributions for Oct4, Esrrb and Neurod1 from single-molecule FISH. Scale bars, 40 μm. c, Gene categories showing high or low noise. d, Sporadic expression of the Polycomb target gene Bmp4 within an mESC colony as measured by smFISH. Scale bars, 40 μm.

Certain Polycomb target genes were expressed in some cells at levels comparable to pluripotency regulators, up to 60 transcripts per cell, despite the presence of the repressive H3K27me3 (trimethylation of lysine 27 on histone H3) chromatin mark associated with Polycomb activity (Fig. 1d, Extended Data Fig. 4 and Supplementary Table 5). As expected, Polycomb target genes showed higher average levels of H3K27me3, were detected in a smaller fraction of cells, and were expressed at lower levels than non-Polycomb target genes (Fig. 1e). However, within this set of genes, those with detectable expression showed lower average levels of H3K27me3 than those that were not detected (Student’s t-test $P = 4.82 \times 10^{-5}$), suggesting that dynamic fluctuations in chromatin state are associated with sporadic expression of certain Polycomb targets in PSCs. This subset of Polycomb target genes may represent regulators governing initial steps in lineage commitment, and may therefore be subject to particular dynamic regulation in PSCs.

To determine the stability of expression states over rounds of cell division, we seeded individual mESCs onto culture plates, allowed them to form colonies over 3–4 days, and quantified the level of inter- and intra-colony variability for selected genes using smFISH (Fig. 2a). Individual colonies showed distinct expression states for Esrrb, with some locked into a high level, some locked into a low level, and others that displayed mixed expression suggesting a sudden switching between high and low states during the process of colony formation (Fig. 2b). Other pluripotency regulators, including Nanog, Nr0b1 (also called Dax1) and Myc, as well as lineage regulators and Polycomb targets Neurod1, Oct2, Olig2 and Pax6, also exhibited high inter-colony variability suggestive of slow fluctuations in expression states (Fig. 2c). As clusters of neighbouring cells or entire colonies tended to be in similar expression states (Extended Data Fig. 5), we estimate that transitions between transcriptional states for these variable regulators occur relatively infrequently with respect to the ES cell cycle, happening on the order of one to a few days, in line with measurements of Nanog transcriptional fluctuations. To confirm that ‘variable expression’ colonies were clonally derived, we performed time-lapse imaging to monitor colony formation over 4 days. Individual colonies formed from single cells showed substantial differences...
in growth rate and bimodal Nr0b1 expression (Fig. 2d), validating our approach and highlighting the pronounced variability and persistence of growth rate, morphology and expression state of mESCs grown in serum plus LIF (serum + LIF mESCs). This expression state persistence extended to the protein level (Fig. 2e and Extended Data Fig. 5), indicating that slow fluctuations in expression of certain pluripotency regulators might underlie distinct phenotypic responses of individual mESCs to external stimuli.

Clustering of pluripotency regulators revealed that they partitioned into several co-expressed modules, with some modules positively correlated with Polycomb target expression and others negatively correlated (Fig. 2f and Extended Data Fig. 5). Nanog, Nr0b1 and Zfp42 (also called Rest) were among those showing the strongest negative correlation with Polycomb target gene expression. To test these associations, we examined the dependence of selected Polycomb target genes on individual pluripotency regulators by two-colour FISH (Fig. 2g and Extended Data Fig. 5). Consistent with the RNA-seq data, Nanog was more likely to be expressed at high levels in cells expressing high levels of Sox2 (which also functions to specify the neural lineage) and low levels of Nr0b1, while Bmp4 expression was positively associated with Esr2b, and Otx2 expression was negatively associated with Nanog. Notably, the 8% of cells separated from the main population by principal component analysis tended to be in low expression states for Nr0b1, Nanog and Zfp42, and showed higher Polycomb target gene expression than did the majority of cells (Extended Data Figs. 2 and 5), suggesting that the RNA-seq-detected correlations reflect biologically meaningful states in which pluripotency factor expression influences the probability of lineage regulator expression.

**Regulatory pathways govern PSC heterogeneity and state**

For a better understanding of the factors governing PSC heterogeneity, we examined mouse PSCs cultured under different growth conditions13, treated with chemical inhibitors of epigenetic regulators, genetically modified to lack particular regulators30–34, or in different states of pluripotency13 (Extended Data Table 1, Supplementary Table 6 and Extended Data Fig. 6). While unimodally expressed genes such as Oct4 and Rest remained relatively invariant across a range of perturbations, the distributions of bimodally expressed genes such as Myc and Tcfcp2l1 shifted markedly in response to particular perturbations (Fig. 3a). Culturing mouse PSCs with inhibitors of Erk and GSK3 signalling (‘2i’ or ground state conditions) has previously been shown to promote self-renewal and block differentiation35. Notably, both culturing mESCs in 2i + LIF conditions13 and impairing microRNA (miRNA) production36,37 resulted in more uniform gene expression across 15 pluripotency regulators examined by single-cell quantitative polymerase chain reaction (qPCR), whereas knocking out PRC2 function through loss of the Polycomb-group (PcG) protein EED (ref. 31) resulted in greater population heterogeneity across all genes measured (Fig. 3b).

We next performed principal component analysis on the single-cell qPCR data and applied an automated classification algorithm to assign cells to discrete states (Fig. 3c–e and Extended Data Fig. 6). Surprisingly, a large fraction of both Dicer−/− and Dicer knockout mESCs, which lack mature miRNAs due to loss-of-function of separate miRNA processing factors38,39, were assigned to the ground state, suggesting that blocking external signalling pathways and removing miRNAs from PSCs results in common modes of self-renewal (Fig. 3e). PSCs in most other conditions were predominantly classified as a distinct state, which

---

**Figure 3 | Effect of perturbations on gene expression variability and cell state. a.** Population distributions for the unimodally expressed genes Oct4 and Rest, and the bimodally expressed genes Tcfcp2l1 and Myc. iPSCs, induced pluripotent stem cells; MEFs, mouse embryonic fibroblasts; TKO, triple knockout. See Extended Data Table 1 for a full list of conditions and abbreviations. **b.** Comparison of population heterogeneity as measured by difference in mean intra-condition distance between serum + LIF and listed condition. **c.** Principal component analysis (PCA) of single-cell qPCR data. **d.** PCA coloured by the most likely state classification. **e.** Cell state classification of PSCs exposed to different perturbations and conditions. **f.** Expression heatmap of genes contributing the most to the top three principal components, excluding housekeeping and fibroblast genes.
we term the ‘transition state’, as it is primarily associated with cells cultured under serum + LIF conditions and a higher probability of spontaneous differentiation. Epiblast stem cells (EpiSCs) were assigned to a separate state we term ‘primed’, in keeping with the view of these cells as being more developmentally advanced. Genes contributing the most to the principal components that distinguished Dgcr8<sup>−/−</sup> and Dicer knockout mESCs in serum + LIF and wild-type mESCs in 2i + LIF from mESCs in other conditions included Myc, Lin28a, Bmp4, Dnmt3b, and Dnmt3l, all of which showed sharply reduced expression in ground state cells (Fig. 3f). Our perturbation analysis therefore implicates miRNAs as key mediators of the transition and primed states, with their absence mimicking the inhibition of the Erk and GSK3 signalling pathways observed in 2i culture.

### 2i and miRNA deficiency promote the PSC ground state

To investigate further the apparent similarity between wild-type mESCs cultured in 2i + LIF conditions (2i + LIF mESCs) and Dgcr8<sup>−/−</sup> mESCs cultured in serum + LIF (Dgcr8<sup>−/−</sup> mESCs), we performed single-cell RNA-seq on mESCs from each condition (Supplementary Tables 2, 7, 8, 9). Gene expression changes between serum + LIF and 2i + LIF mESCs were highly correlated with changes between serum + LIF and Dgcr8<sup>−/−</sup> mESCs (Fig. 4a and Extended Data Fig. 7). As a population, 2i + LIF mESCs showed reduced heterogeneity as compared to serum + LIF mESCs, in keeping with the notion that inhibiting Erk and GSK3 signalling drives mouse PSCs into a noise-like ground state<sup>20,37,38</sup>. When compared across conditions, however, individual Dgcr8<sup>−/−</sup> mESCs were more similar to 2i + LIF than serum + LIF mESCs, supporting a model in which removal of miRNAs drives a portion of Dgcr8<sup>−/−</sup> mESCs towards the ground state, and results in a similar phenotype to that produced by Erk and GSK3 inhibition.

While some pluripotency regulators such as Esrrb and Tbx3 showed increased expression in 2i + LIF and Dgcr8<sup>−/−</sup> mESCs, other factors traditionally associated with pluripotency (for example, Myc and Lin28a) showed a sharp reduction in expression under both conditions (Fig. 4c). DNA methyltransferases Dnmt3a, Dnmt3b and Dnmt3l showed reduced expression in both 2i + LIF and Dgcr8<sup>−/−</sup> mESCs, suggesting that miRNA activity may be linked to the reduction in DNA methylation observed in the naive pluripotent state<sup>24,39</sup>. Regulators displaying bimodal expression patterns in serum + LIF mESCs, including the pluripotency factor Esrrb, showed altered distributions in both 2i + LIF and Dgcr8<sup>−/−</sup> mESCs. Expression was confirmed by smFISH and quantitative immunofluorescence, and was recapitulated in an independent mESC line (Fig. 4d, e and Extended Data Fig. 7). Thus, the pluripotency regulatory network adopts distinct configurations in the ground and transition states.

mESCs cultured in 2i + LIF exhibit a unique chromatin state characterized by lower levels of H3K27me3 at promoters and increased amounts of transcriptional pausing<sup>40</sup>. Both culture in 2i + LIF and removal of mature miRNAs resulted in a reduction of H3K27me3 at promoters (Fig. 4f, g, Extended Data Fig. 8 and Supplementary Table 5). Therefore, Dgcr8<sup>−/−</sup> mESCs manifest features of ground-state self-renewal including altered gene expression, a reconfigured pluripotency regulatory network, and a common chromatin state.

### miRNA balance and cell state

We profiled miRNA expression in two mESC lines cultured under both serum + LIF and 2i + LIF conditions, and mouse embryonic fibroblasts (MEFs) cultured under standard conditions (Supplementary Table 10). Notably, while levels of ES-cell-specific cell-cycle-regulating (ESCC) miRNAs<sup>41,42</sup>, which are known to be expressed at high levels in PSCs, remained elevated in both serum + LIF and 2i + LIF growth conditions, let-7 family members and miR-152, which act as tumour suppressors and are typically associated with differentiated cells<sup>43,44</sup>, were expressed at higher levels in 2i + LIF as compared to serum + LIF culture (Fig. 5a and Extended Data Fig. 9).

Prevailing models which posit that the ESCC and let-7 miRNA families are opposing classes of miRNAs that act in self-reinforcing loops to stabilize self-renewing and differentiated states<sup>46</sup> run contrary to our observed co-expression of ESCC and let-7 miRNAs in 2i + LIF mESCs. Analysing our own and published data<sup>46</sup>, we found that targets of let-7 showed significantly lower expression as a group in 2i + LIF as compared to serum + LIF mESCs ($P = 1.4 \times 10^{-5}$, Mann–Whitney U-test), as did miR-152 predicted targets ($P = 0.005$), suggesting that these miRNAs...
expression of ESCC miRNAs in 2i, their direct targets showed elevated expression as a class in 2i as compared to serum ($P = 6.412 \times 10^{-9}$; Mann–Whitney U-test, Extended Data Fig. 9), whereas let-7 and miR-152 direct targets were downregulated in 2i. These findings suggest that enforced expression of ESCC miRNAs elevates Dgcr8$^{-/-}$ mESCs out of ground-state self-renewal through indirect modulation of the Myc/Lin28/let-7 axis, and that in 2i let-7 and miR-152 oppose this effect of ESCC miRNAs through direct repression of a set of target genes including Myc, Lin28, Dmnt3b and Dmnt3l.

Both 2i+LIF and Dgcr8$^{-/-}$ mESCs give rise predominantly to compact colonies uniformly positive for alkaline phosphatase staining, as compared to serum+LIF mESCs which have a greater tendency to form larger colonies that are mixed for alkaline phosphatase staining (Fig. 5c). Compact, uniformly alkaline-phosphatase-positive colonies formed in serum+LIF culture displayed lower levels of Myc expression than did alkaline phosphatase mixed and negative colonies (Extended Data Fig. 9). We independently knocked down Lin28a expression in serum+LIF mESCs, and forced let-7 expression in doxycycline-inducible let-7 mESCs grown in serum+LIF$^{13}$. Lin28a blocks processing of let-7 precursor miRNAs into their mature form, and knockdown of Lin28a in mESCs grown in serum+LIF results in increased let-7 levels, increased colony-forming efficiency, and a higher proportion of smaller, uniformly alkaline-phosphatase-positive colonies (Fig. 5d and Extended Data Fig. 9). Sustained expression of let-7 in serum+LIF culture using the inducible system resulted in a higher proportion of smaller, uniformly alkaline-phosphatase-positive colonies that showed lower levels of Myc expression (Fig. 5e, f and Extended Data Fig. 9), suggesting that let-7 is capable of activating a positive feedback circuit that stabilizes ground state self-renewal. Acute inhibition of let-7 family members and miR-152 in wild-type mESCs cultured in 2i+LIF resulted in upregulation of Myc and Lin28a (Fig. 5g), further supporting the notion that these two miRNAs act in concert to maintain the ground state.

We tested the effect of reintroducing miRNAs into Dgcr8$^{-/-}$ mESCs on their self-renewal to see if ESCC-mediated elevation into the transition state could be counterbalanced by let-7. Transfection of stable mimics of let-7 resulted in sharply reduced self-renewal efficiency compared to controls, and transfection of mimics of the ESCC miRNA miR-294 also resulted in reduced self-renewal (Extended Data Fig. 9). Co-transfection of both miRNA family members together, however, resulted in higher self-renewal efficiency than did introduction of either family member alone. These results support a model in which expression of ESCC miRNAs alone drives PSCs into a transition state with a relatively high probability of spontaneous differentiation through indirect activation of Myc and Lin28a, while expression of let-7 miRNAs alone can drive differentiation. However, expression of a balance of ESCC miRNAs along with the opposing families let-7 and miR-148/152 as in 2i conditions, or neither as in the case of Dgcr8$^{-/-}$ mESCs, results in ground-state self-renewal through either direct repression of Myc and Lin28a by let-7 and miR-148/152, or removal of the activating force of ESCC miRNAs (Fig. 5h).

Conclusion

The diverse range of conditions under which pluripotency can be induced or maintained has been accompanied by reports of molecular and functional variation. Here we analysed the dynamic transcriptional landscape of pluripotent stem cells subject to a number of chemical and genetic perturbations. Applying single-cell analytics, we gleaned a number of essential insights. We found that different classes of genes manifest high or low expression variability in PSCs, with housekeeping and metabolic gene sets showing consistent expression across individual cells, while genes involved in signalling pathways and development were considerably more variable. Moreover, expression states of variable regulatory factors were coupled together, implying the presence of a regulated biological network. Analysis of chemical and genetic perturbations led to the discovery that depletion of miRNAs mimicked the transcriptional ground state of pluripotency routinely induced by culture in 2i+LIF,
conditions that block the dominant ERK and GSK3 signalling pathways that converge on the Myc/Lin28/let-7 axis. Our data shed light on the transcriptional dynamics of the pluripotent state at the single-cell level, and demonstrate how regulation of gene expression variation relates directly to the transition between pluripotency and differentiation.

Transcriptional heterogeneity is increasingly being recognized as a key component of many biological processes. It will be of interest to map stable and flexible regulatory nodes in networks governing other progenitor and differentiated cell types to discern common principles underlying network architecture and gene expression variability.

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

Received 13 February; accepted 7 October 2014.

1. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956 (2005).

2. Loh, Y.-H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature 457, 1127–1132 (2009).

3. Loh, Y.-H., Zilionis, R. & Daley, G. Q. Deconstructing pluripotency. Annu. Rev. Genet. 42, 165–185 (2011).

4. MacArthur, B. D., Ma’ayan, A. & Lemischka, I. R. Systems biology of stem cell fate specification and cellular reprogramming. Cell Stem Cell 4, 480–483 (2009).

5. Young, R. A. Control of the embryonic stem cell state. Nature Struct. Mol. Biol. 20, 311–316 (2013).

6. Chambers, I. et al. Functional specification of embryonic and adult stem cells. Nat. Rev. Mol. Cell Biol. 10, 757–763 (2009).

7. Reynolds, N. et al. Neurogenesis from mouse and human embryonic stem cells. Proc. Natl Acad. Sci. USA 104, 18097–18102 (2007).

8. Cherry, A. & Daley, G. Q. Another horse in the meta-stable state of pluripotency. Stem Cells Dev. 22, 1496–1505 (2008).

9. Kalmar, T. et al. The transcriptional and epigenomic foundations of ground state pluripotency. Cell 149, 515–524 (2012).

10. Reinhart, B. J. et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403, 901–906 (2000).

11. Murphy, H. A. et al. Pluripotency: links between transcriptional and epigenetic states. Cell Stem Cell 3, 391–401 (2008).

12. Graf, T. & Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. Cell Stem Cell 2, 196–207 (2008).

13. Marks, H. et al. Transcriptional heterogeneity enforces lineage plasticity in human embryonic stem cells. Stem Cells 30, 144, 940–954 (2011).

14. Nakamura, A. & Sakaguchi, K. Heterogeneity of mouse embryonic stem cells with distinct functional and epigenetic states. Cell Stem Cell 3, 391–401 (2008).

15. Wang, Y. et al. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nature Genet. 40, 1478–1483 (2008).

16. Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. Cell 149, 590–604 (2012).

17. Shalek, A. K. et al. Single-cell RNA-seq reveals bimodality in expression and biological noise: from microbes to mammals. Cell 144, 910–925 (2011).

18. Reinhart, B. J. et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403, 901–906 (2000).

19. Meton, C., Judson, R. L. & Blöb, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature 463, 621–626 (2010).

20. Zhu, H. et al. The Lin28/let-7 axis regulates glucose metabolism. Cell 147, 81–94 (2011).

21. Baldià, G. et al. The Lin28/let-7 axis regulates glucose metabolism. Cell 147, 81–94 (2011).

22. Haiey, J. et al. Self-organizing circuitry and emergent computation in mouse embryonic stem cells. Stem Cell Res. 8, 324–333 (2012).

23. Boyer, L. A. & Clevers, H. M. Pluripotency factors in embryonic stem cell regulate differentiation into germ layers. Cell 144, 841–851 (2011).

24. Halley, J. D. et al. Organizing circuitry and emergent computation in mouse embryonic stem cells. Stem Cell Res. 8, 324–333 (2012).

25. Koh, K. & Lim, B. A. Precise balance: pluripotency factors as lineage specifiers. Cell Stem Cell 10, 363–369 (2012).

26. MacArthur, B. D. & Lemieshka, I. R. Statistical mechanics of pluripotency. Cell 154, 484–489 (2013).

27. Silva, J. & Smith, A. Capturing pluripotency. Cell 132, 532–536 (2008).

28. Thomson, M. et al. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 144, 875–889 (2011).

29. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. Nature Rev. Mol. Cell Biol. 14, 357–368 (2013).

30. Calabrese, J. M., Seila, C. A., Yeo, G. W. & Sharp, P. A. RNA sequence analysis defines Dicer’s role in mouse embryonic stem cells. Proc. Natl Acad. Sci. USA 104, 18097–18102 (2007).

31. Martin-Belmonte, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells Dev. 22, 1496–1505 (2012).

32. Kaji, K. et al. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat. Cell Biol. 4, 891–896 (2002).

33. Tsumura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cell 11, 805–814 (2006).

34. Wang, X. et al. Dynamic heterogeneity of human embryonic stem cells. Nature 493, 380–385 (2007).

35. Tesar, P. J. et al. New cell lines from mouse epiblast blastuloid define features with human embryonic stem cells. Nature 448, 196–199 (2007).

36. Wang, Y. et al. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nature Genet. 40, 1478–1483 (2008).

37. Grün, D., Keister, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. Nature Methods 11, 637–640 (2014).

38. Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. Cell 149, 515–524 (2012).

39. Leitch, H. G. et al. Naive pluripotency is associated with global DNA hypomethylation. Nature Struct. Mol. Biol. 20, 311–316 (2013).

40. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. Cell 149, 590–604 (2012).

41. Boyerinas, B., Park, S. M., Hau, A., Murrman, A. E. & Peter, M. E. The role of let-7 in cell differentiation and cancer. Endocr. Relat. Cancer 17, F19–F36 (2010).

42. Tsun, T. et al. mir-152-5p microRNA that is maternally expressed by DNA hypermethylation in endometrial cancer. Cancer Res. 71, 6450–6462 (2011).

43. Blueme, J. W. et al. Phenotypic consequences of promoter-mediated transcriptional noise. Mol. Cell 24, 853–865 (2006).

44. Blau, W. J. et al. MicroRNA-dependent regulation of the G1-S transition and proliferation. J Cell Sci 123, 4633–4647 (2013).

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

Acknowledgements We thank members of the Collins and Daley laboratories for discussions. J.J.C. is supported by NIH grant R24DK092760 and the HHMI. Q.D. is supported by grants from the NIH (R01GM107536, R24DK092760, P50GM005590) and is an affiliate member of the Broad Institute and an investigator of the Harvard Center for Orphan Disease Research and the Howard Hughes Medical Institute. A.R. is supported by the Broad Institute, the Klarman Cell Observatory at the Broad Institute, and the NIH (R00HG009191). A.K.S. was awarded an NIH Pioneer Award (DP1OD003958-01) and the HHMI, R.M.K. is supported by the Wyss Institute. P.C. is supported by a Wellcome Trust New Investigator Research Project. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nature Rev. Mol. Cell Biol. 6, 633–637 (2005).

Author Contributions A.M. performed and designed the experiments, analysed the data and wrote the paper. P.C. analysed the data, developed analytical tools and wrote the paper. A.K.S., D.G. and J.J.T. performed ChIP-seq and single-cell RNA-seq experiments. R.S. helped to design the study and write the paper. Address correspondence and requests for materials should be addressed to G.Q.D. (george.daley@childrens.harvard.edu) or J.J.C. (jcollins@bu.edu).
Methods

Cell culture. Mouse embryonic stem cells (mESCs) and induced pluripotent stem cells (iPSCs) were cultured on γ-irradiated mouse embryonic fibroblasts (MEFs) (GlobalStem, CF-1 MEFs) or in feeder-free conditions as detailed in Extended Data Table 1. Feeder-free cells were grown on Matrigel (hESC-qualified, BD)-coated Nunc cell-culture treated plates unless otherwise specified. Cells cultured in Serum + LIF were grown in Dulbecco’s Modified Eagle Medium (DMEM) + Glutamax (Gibco) plus 15% fetal bovine serum (FBS), Statis stem-cell qualified, Gemini), 1× penicillin/streptomycin (Gibco), 1× non-essential amino acids (NEAA) (Gibco), 1× β-mercaptoethanol (β-Me) (Embodyx ES cell-qualified, Millipore), and 1000 U ml⁻¹ leukenia inhibitory factor (LIF) (ESGRO, Millipore). Cells cultured in 2× LIF were maintained in N2/B27 media, consisting of 50% DMEM/F12 (1:1) and 50% Neurobasal plus 0.5x N-2 supplement, 0.5x B-27 supplement, 1× penicillin/streptomycin, 1× Glutamax, and 1× NEAA (all from Gibco), and 1× β-Me (Millipore) supplemented with 1 μM MEK inhibitor PD0329501 (StemGent), 3 μM GSK3 inhibitor CHIR99021 (StemGent), and 1,000 U ml⁻¹ LIF (Millipore). For the 11 experiments shown in Extended Data Fig. 9h, either the MEK inhibitor PD0329501 or the GSK3 inhibitor CHIR99021 were omitted from the media, respectively. Except as described, all mouse PSCs were passaged with trypsin (Gibco). Epiblast stem cells (EpiSCs) were grown on MEF feeder cells in DMEM-F12 (Gibco) supplemented with 20% knockout serum replacement (Life Technologies), 5 ng ml⁻¹ FGF2 (RD System), 1 μM 2-mercaptoethanol (Sigma), 2 mM l-glutamine (Gibco), and 1× NEAA (Gibco), and passaged using collagenase type IV (Life Technologies).

MEFs were used in the miRNA profiling experiments described in Fig. 5a were from GlobalStem (C57/Bl6 MEFs), harvested at the second passage after thawing from frozen stocks, and cultured under standard conditions in DMEM + 10% FBS + 1× Pen/Strep/Glut.

For perturbation experiments described in Fig. 3, v6.5 mESCs were treated with 10 nM trichostatin A (Sigma), 1 μM valproic acid (Sigma), or 4 μM BIX01294 (Stemgent) for 6 days before analysis, passing twice in the interim.

Origin of cell lines used in this study. V6.5 mESCs were a gift from R. Jaenisch. E1Ht2ga mESCs were obtained from ATCC (CRL-1821). Efd⁻ mESCs were a gift from T. Magnuson. Dnmnt⁻/- Dnmnt3a⁻/- Dnmt3b⁻/- triple knockout mESCs were a gift from M. Okano. Dnmt1⁻/⁻ mESCs were obtained from Novus (NBA1-19349). M4d3⁻/⁻ mESCs were a gift from B. Hendrich. mEpiscs were a gift from P. Tesar. Dicer knockout mESCs and their matched Il1 wild-type control mESCs were a gift from P. Sharp. iPSCs and iShLin28a cells were derived as described below.

Generation of iPSCs. Lentivirus production and generation of the iPSCs profiled by single-cell qPCR were performed as previously described. Briefly, iPSCs were generated from C57BL/6 MEFs (GlobalStem) using the doxycycline-inducible Tet-STEMCCA cassette and a constitutive reverse tetracycline transactivator (rtTA; gifts of G. Mostoslavsky) packaged into lentiviral vectors. In the presence of doxycycline the Tet-STEMCCA cassette expresses OCT4, KLF4, SOX2 and Myc from a single multicistronic transcript. The MEFs (P2) were plated at 25,000 cells ml⁻¹ into a gelatin-coated, six-well plate and allowed to become adherent overnight. Cells were maintained in N2 media supplemented with bFGF/laminin to expand nestin-positive neural cells, and were then plated on to laminin and polyornithine-coated plates and shifted to LIF mESC media. At day 4, embryoid bodies were washed in PBS, spun down and resuspended at a concentration of 3 × 10⁶ cells ml⁻¹ of complete media, mixed 7:3 with C5 suspension reagent (Fluidigm), and loaded onto C5 Single-Cell Auto Prep chips (C5 chips; Fluidigm). After loading, each of the cell isolation chambers on the C5 chip was optically inspected for the presence of a cell. Subsequently, these cells were lysed and SMART-Seq² whole transcriptome amplified (WTA) products were prepared with the C5 System using the SMARTer Ultra Low RNA kit for Illumina Sequencing (Clontech) and protocols provided by Fluidigm.

Library preparation and RNA-seq. WTA products were harvested from the C5 chip, diluted to a concentration of 0.15 μg ml⁻¹, and cDNA libraries were prepared using Nextera XT DNA Sample Preparation reagents (Illumina) as per the manufacturer’s recommendations, with minor modifications. Specifically, reactions were run at one-quarter the recommended volume, the tagmentation step was extended to 10 min, and the extension time during the PCR step was increased from 30 to 60 s. After the PCR step, all 96 samples were pooled without library normalization, cleaned twice with 0.9X AMPure XP SPRI beads (Beckman Coulter), and eluted in buffer TE. The pooled libraries were quantified using a Quanti-IT DNA High-Sensitivity Assay kit (Invitrogen) and examined using a high-sensitivity DNA chip (Agilent). Finally, samples were sequenced deeply using either a HiSeq 2000 or a HiSeq 2500 sequencing kit (25-bp paired-end reads). Single-cell sequences were sequenced to an average depth of more than one million aligned reads per cell, with full values for number and fraction of aligned reads provided in the Supplementary Tables.

RNA-seq of population controls. Population controls were generated by extracting total RNA (RNAeasy plus Micro RNA kit; Qiagen), according to the manufacturer’s recommendations. Subsequently, 1 μl of RNA in water was added to 2 μl of C5 lysis reaction mix, thermocycled using cycling conditions 1 (as below). Next, 4 μl of the RT Reaction Mix was added and the mixture was thermocycled using cycling conditions II (as below). Finally, 1 μl of the total RT reaction was added to 9 μl of PCR mix and that mixture was thermocycled using cycling conditions III (as below). Products were quantified, diluted to 0.15 ng μl⁻¹ and libraries were prepared, cleaned and tested as described.

For the cell lysis mix we used the following composition and cycling conditions. C5 loading reagent (stock concentration 20X; volume 0.6 μl), SMARTer Kit RNase Inhibitor (40X; 0.30 μl), SMARTer Kit 3’ SMART CDS Primer II A (12 μM; 4.20 μl), SMARTer Kit dilution buffer (1X; 6.90 μl); cycling conditions I: (a) 72 ºC, 3 min, (b) 4 ºC, 10 min, (c) 25 ºC, 1 min.

For the reverse transcription (RT) reaction mix we used the following composition and cycling conditions. C5 loading reaction (stock concentration 20X.0; volume 0.45 μl), SMARTer kit 5X First-Strand Buffer (RNase-Free) (5/0; 4.20 μl), SMARTer kit dithiothreitol (100 mM; 0.53 μl), SMARTer kit dNTP Mix (dATP, dCTP, dGTP and dTTP, each at 10 mM) (10 μl; 2.10 μl), SMARTer kit SMARTer Kit II Aligoumamide (12 μl; 2.10 μl) SMARTer kit RNase inhibitor (40X; 0.53 μl), SMARTer kit SMARTCDS reverse transcriptase (100X; 2.10 μl); cycling conditions II: (a) 42 ºC, 90 min, (b) 70 ºC, 10 min.

For the PCR mix we used the following composition and cycling conditions. PCR water (35.52 μl), 10X Advantage 2 PCR buffer (stock concentration 20X.0; volume 5.65 μl), 50X dNTP mix (10 mM; 2.2 μl), IS PCR primer (12 μl; 2.2 μl), 50X Advantage 2 Polymerase mix (50X; 2.25 μl), C1 loading reagent (20X.0; 2.5 μl), C1 loading reagent (20X.0; 2.5 μl), (c) 95 ºC, 1 min; (d) 5 cycles of 95 ºC, 1 min; 58 ºC, 1 min; (e) 68 ºC, 1 min; (f) 68 ºC, 20 s, (g) 58 ºC, 4 min, (h) 68 ºC, 6 min; (i) 68 ºC, 6 min; (j) 95 ºC, 20 s, (k) 64 ºC, 30 s, (l) 68 ºC, 6 min; (m) 72 ºC, 10 min. Calculation of expression levels and statistics. Mapping of reads and calculation of expression values was performed as described. A Bowtie index was created based on the UCSC knownGene transcriptome, and paired-end reads were directly aligned to this index. RSEM was run on these alignments to generate expression level estimates, and these estimates (t) were multiplied by 1,000,000 to yield estimates of transcripts per million (TPM) for each gene. These TPM estimates were then transformed into log space by taking ln(TPM+1), which we hereafter refer to as lnTPM. Gene expression levels referred to in the manuscript represent ln-transformed TPM values.

For the fraction of cells a gene was detected in, we considered as detected any gene showing an ln[TPM] value >0 after filtering out outlier cells (see below). Average single-cell expression levels were calculated by first averaging TPM levels in linear space, and then log-transforming by taking ln(linear average TPM+1). Gene expression levels in expressing cells, as shown in Fig. 1e, were calculated by averaging TPM levels in log space but only over the number of cells in the population that the gene was detected in. Standard deviations (s) were calculated using gene expression values in log space, and coefficients of variation were calculated by dividing the population standard deviation by the average single-cell expression value as defined above.

We used Saffish to estimate the abundances of non-coding RNAs in the single-cell RNA-seq libraries. Saffish was executed using a transcript index file (generated as described below) and fastq files of the library reads to generate transcript level expression data, referred to in the manuscript as lnTPM.
abundances, which we summed on a per gene basis to generate gene level estimates. To generate a transcript index file, we used Ensembl BioMart (accessed on 4 June 2014), using the Ensembl Genes 75, GCRm38.2 protein_coding, miRNA, and lncRNA annotations, using a kmer size = 17 when compiling the Sailfish index. We used the same process to identify reliably detected genes using SailFish-processed data as described in Supplementary Information section 'Single-cell RNA-seq reproducibility and sensitivity'. We judged genes detected in at least 21% of cells, or showing a maximum Sailfish expression level of ln(TPM) ≥ 1.83 and detected in at least 10% of cells to be reliably detected.

Assessment of library quality and exclusion of cells from analysis. Library quality metrics, including genomic mapping rates, were calculated using PicardTools version 1.42 (http://broadinstitute.github.io/picard/). Barcoded RNA-seq libraries were inspected for alignment rates and number of genes detected, and criteria for removal of outlier cells were empirically determined from a combination of these factors. A large majority of cells grouped together with similar numbers of genes detected and alignment rates (Extended Data Fig. 1). Cells showing substantially fewer numbers of genes detected and/or lower alignment rates than these groups were filtered out, leaving 183 individual wild-type mESCs in serum + LIF, 94 individual wild-type mESCs in 2i + LIF, 84 Dgcr8−/− mESCs in serum + LIF, and 55 ESC-derived neural precursor cells that passed this quality control step and were used in subsequent analyses.

Definition of Polycomb target genes for expression comparisons. Polycomb target genes in mESCs were defined as those called bound by the Polycomb repressive complex 2 (PRC2) subunits SUZ12 and EZH2, the Polycomb repressive complex 1 (PRC1) subunit Ring1B, and carrying the H3K27me3 mark in ref. 26. Identification of gene sets enriched for variable genes. We used a permutation approach to determine whether gene sets (Gene Ontology Biological Process annotations plus a list of Polycomb target genes identified in ref. 26) tended to have coefficients of variation (COV)’s higher or lower than expected by chance. First, we computed the average COV of all genes in a gene set. Then, to determine whether it was higher or lower than expected by chance, we randomly sampled x genes where x is the number of genes in the gene set being assessed, and computed the average COV. We repeated this procedure 10,000 times, and computed the empirical P value as 1 + the number of random samples in which the mean COV ≥ observed COV/(1 + 10,000). Similarly, P values for genes having lower COV than expected by chance were computed as 1 + the number of random samples in which the mean COV ≤ observed COV/(1 + 10,000). Because COV depends on μ, and genes in some gene sets have high or low μ (for example, housekeeping genes and lineage regulators, respectively), we weighted the gene sampling such that the randomly sampled gene sets had equivalent μ distributions as the gene set being assessed. These P values were adjusted by applying Holm’s method to correct for multiple tests. Only significant gene sets are listed in Fig. 1c even though this panel shows the nominal P values.

Correlation and distance analyses. The Pearson correlation coefficient was used to quantify gene expression correlation in Fig. 2f. All dendrograms were generated by unsupervised hierarchical clustering using 1 − Pearson correlation coefficient as the distance metric and average linkage as the linkage method. All intra- and inter-condition distances (Figs 3b and 4b and Extended Data Figs 4g and 7c) were computed as 1 − Pearson correlation coefficients. For Fig. 3b, pluripotency genes examined were Pou5f1 (Oct4), Nanog, Etf, Rex1, Sall4, Dmnt1, Tcfcp21, Zfx, Lin28a, Dmnt3l, Dmnt3b, Ronin, Stella, Sall1 and Myc, and housekeeping genes were Actb, Gapdh, Rps6 and Ppia.

Identification of differentially expressed genes. To identify differentially expressed genes between mESCs cultured in serum + LIF, in 2i + LIF, or Dgcr8−/− mESCs, we first quantile normalized ln(TPM) expression values across all three data sets. We performed pairwise comparisons of all three conditions, identifying differentially expressed genes using t-test and adjusting P values with Holm’s method to correct for multiple tests. We defined genes as differentially expressed as those with corrected P values <0.01 and |log( ratios )| > 1. To identify genes differentially expressed between mESCs cultured in serum + LIF and ESC-derived neural precursor cells, we quantile normalized raw ln(TPM) expression values across these two data sets, and then identified differentially expressed genes as above.

RNA fluorescence in situ hybridization (FISH). For RNA FISH experiments, mESCs were seeded in the absence of feeder cells at a density of 200–500 cells per well of a Matrigel (BD)-coated 24-well glass bottom plate (MatTek, No. 1.5 glass, 13 mm diameter). This seeding density was empirically determined to yield sparse individual cells attaching to the matrigel-coated glass. Individual cells were allowed to form colonies over 3–4 days of culture, and then fixed with 4% formaldehyde for 30 min at room temperature and stained with RNA FISH probes (Panomics, Quant-iGene ViewRNA ISH Cell Assays) according to manufacturer’s recommendations, except that protease treatment was omitted. After staining with FISH probes, cells were treated with 1× DAPI and 1:100 CellMask Green stain (Life Technologies) for 10 min at room temperature, washed, treated with ProLong Gold (Life Technologies), covered with a cover slip and allowed to cure overnight. Individual colonies were then identified manually and entire colonies or portions of them were scanned on a Leica SP5 X MP inverted confocal microscope outfitted with a tunable white light laser at 63× magnification (1.3 NA, glycerol immersion) in x, y and z using HyD detectors set to photon counting mode. z- intervals were spaced 0.3 μm apart, and colonies were scanned to their upper and lower limits resulting in on average 50–60 slices per colony. Five-channel imaging was performed to obtain brightfield, DAPI (ex 405 nm), CellMask Green (ex 486 nm), probe 1 (ex 564 nm), and probe 2 (ex 647 nm, if applicable) images.

To count numbers of transcripts and assign them to individual cells, we first generated two-dimensional projections of the three-dimensional stacks for the FISH probe channels, and then determined the number of transcripts per colony in Matlab using a modified version of a previously described analysis package58. Briefly, numbers of spots were counted at a range of threshold values, spot count versus threshold plots were manually inspected to determine a threshold plateau range over which transcript number was relatively insensitive to the threshold value, and this was set as the threshold to use. These thresholds were then applied in CellProfiler60, which was used to identify individual cells and assign transcripts to them. Nuclei were identified from two-dimensional projections of DAPI images in a semi-automated process, with manual editing as needed for nuclei that CellProfiler was not able to correctly identify and segment. Identified nuclei were then used as seeds to propagate outwards and identify cell boundaries on the basis of two-dimensional CellMask Green projections. Transcript spots were identified from images thresholded with the values obtained in Matlab, and then assigned to individual cells. The number of transcripts identified by this method agreed well with the number found by three-dimensional spot-finding in Matlab, indicating that the live-cell characterization, particularly when using the HyD detectors in photon counting mode) and relatively flat morphology of the ESC colonies grown in the absence of feeders enables two-dimensional transcript quantification.

For all RNA FISH distributions shown, kernel density plots, which are estimates of the probability densities of the underlying distributions, were generated using the density function in R with default settings. Where histograms are shown, counts were binned into 20–60 bins and histograms were plotted using the hist function in R and overlaid with kernel densities.

Live cell tracking experiments. For the live cell tracking experiments shown in Fig. 2d, cells were plated and cultured as described above, except that cells were stained with 1 μM CellTrace calcine AM-red-orange (Life Technologies), allowed to settle for 2 h, and imaged on a Zeiss Axio Observer fluorescence microscope before returning to the incubator. Cells were then stained with 1 μM calcein for 15 min daily, imaged, and returned to the incubator after imaging. Control experiments established that staining with 1 μM calcein did not affect cell growth rate or colony morphology. Cells were grown for 4 days, and then fixed and processed for RNA FISH experiments 16 h after the final calcein staining and imaging to allow for loss of calcein signal.

ChIP-seq. For ChIP–seq experiments, mESCs were grown in the absence of feeder cells and cross-linked when they were at ~50% confluence with 1% formaldehyde for 10 min at room temperature. Cross-linking was quenched by the addition of 125 mM glycine, and cells were then washed with PBS containing protease inhibitors, harvested by scraping and cell pellets were flash-frozen in liquid nitrogen. ChIP-seq was then performed as previously described54.

Antibodies used for chromatin immunoprecipitations were as follows: H3K4me3 (Millipore 07–473), H3K27me3 (Upstate 07–449), H3K36me3 (Abcam ab9050), H3K9me3 (Abcam ab8898), H3K27ac (Abcam ab4729) and RNAP2 (Covance MMS–128P).

To quantify ChIP–seq levels of H3K4me3, H3K27me3, H3K36me3, H3K9me3, H3K27ac and RNAP2 on a per gene level, we calculated the number of reads which mapped to the promoter region of each gene. Promoter regions were defined as the region from 1.5 kb upstream to 0.5 kb downstream of the transcription start site of each annotated gene in the NCBI37/mm9 RefSeq annotations.

Single-cell qRT–PCR. Single-cell qRT–PCR of mESCs was performed on a Fluidigm BioMark according to manufacturer’s recommendations (full details available at http://www.fluidigm.com), with modifications as noted below. Briefly, cells were harvested from culture, stained using an antibody to SSEA1 (Clone MC–480, BD Biosciences), and individual mESCs were sorted into 8-well PCR strips or 96-well PCR plates on the basis of SSEA1 expression and gated such that the top 30% of SSEA1-expressing cells were collected. The viability dye Sytox blue (Life Technologies) was used to exclude dead cells. Cells were sorted into 9 μl of reaction buffer containing a mixture of 96 different TaqMan Gene Expression Assays (Life Technologies) and reagents from the CellsDirect One-Step qRT–PCR kit (Life Technologies), and RT–PCR was performed with 22 cycles of specific target amplification (STAA). Amplification products were diluted 1:5 with TE(8.0), and samples were
checked for the presence of amplified cDNA by duplex microtitre-scale qPCR using Gapdh-FAM and ActB-VIC TaqMan probes (Applied Biosystems). This step was performed to exclude wells that did not contain an actual cell or in which amplification failed, as initial experiments had found that a substantial fraction of wells did not contain amplification products, presumably due to inefficiencies in single-cell sorting. For wells showing successful amplification of both Gapdh and ActB (typically ~50–75% of samples on a plate), 5 μl of diluted samples were submitted for expression profiling, along with a set of population and negative control samples, on 96 × 96 gene expression chips on the Fluidigm BioMark. Ct values were determined with Fluidigm Real-time PCR analysis software, applying the auto (detectors) method and using a linear (derivative) baseline correction with a quality threshold of 0.65. When expression of a gene was not detected in a cell, its Ct value was assigned a value of 40 for the purposes of analysis. Data for the 84 probes that were common to all Fluidigm expression arrays performed are shown in Extended Data Fig. 6b and listed in Supplementary Table 4, along with TaqMan probe ID numbers. Amplification efficiencies for TaqMan probes were determined using standard curve analysis, and only probes with an efficiency between 0.8–1.25 per cycle were included in the study.

For the comparison of technical versus biological variability shown in Extended Data Fig. 6a, three pools of 100 or 10 cells or individual cells were sorted into separate wells of PCR strips and processed as described above, and pre-amplified cDNA from each pool or single cell was run in triplicate on a Fluidigm gene expression chip. To compare technical variability, the average of the three standard deviations of the technical triplicates is indicated by the error bars in the x-direction in Extended Data Fig. 6a. To compare biological variability, the average of the Ct value for the three separate pools or individual cells is indicated by the y error bars.

For the samples labelled ‘Hoechst’ in Fig. 3 and Extended Data Fig. 6, v6.5 mESCs cultured in serum + lIF were incubated with 10 μg ml−1 Hoechst 33342 in serum + lIF growth media for 30 min at 37°C before sorting, and single cells in the G0/G1 phase of the cell cycle were isolated for analysis.

To determine expression distributions changes significantly in response to perturbations, we compared the expression distribution of each gene (Ct value) in each condition to that in the v6.5 serum + lIF sample using the two-sample Kolmogorov–Smirnov test. P values were adjusted using the Holm method to correct for multiple tests.

**Principal components analysis of qRT–PCR data and cell state classification.**

The principal components analysis of single-cell qRT–PCR data (Fig. 3c) was performed by applying the prcomp R function to the complete data matrix of 1,144 individual cells and 81 genes. Genes contributing the most (with absolute loadings > 0.18) to the first three principal components distinguishing mESC perturbations (excluding MEFs) are displayed in Fig. 3c. To identify distinct transcriptional states of the qPCR data, we modelled the first three principal components as multivariate normal mixtures, allowing the model selection step of mClust to test 1 to 9 components and all model geometries available. We used the model parameters that maximized the Bayesian Information Criteria, resulting in a model of five components (that is, clusters or groups of transcriptional states), for further analysis (Extended Data Fig. 6). The resulting model is a mixture density of the form:

$$f(x) = \sum_{i=1}^{C} P_i f_i(x)$$

where $i =$ component, $P_i =$ the probability that the observation $x$ comes from the component (cluster or group) $i$, and $f_i$ is the probability density of component $i$. This model can be used to compute the posterior probability that a given cell comes from one of the $C$ components using Bayes theorem:

$$P(x \in \text{cluster } A) = \frac{P_A f_A(x)}{\sum_{i} P_i f_i(x)}$$

Cells were classified based on the component that maximized this posterior probability. However, if the maximum posterior probability did not exceed 0.50, then the cell was not assigned a state (unclassified in Fig. 3e).

**Quantitative immunofluorescence.** For quantitative immunofluorescence experiments, mESCs were grown in the absence of feeder cells on Matrigel (BD)-coated 24-well glass bottom plates (MatTek, No. 1.5 glass, 13 mm diameter) and cross-linked with 4% paraformaldehyde (Alfa Aesar, catalogue no. 43368) for 15 min at room temperature. Immunostaining was performed according to standard protocols. Briefly, cells were washed in PBS, permeabilized with 0.2% Triton X-100 plus either 1% normal goat serum (NGS, Life Technologies) or bovine serum albumin (BSA, Sigma) if goat primary antibodies were used, washed in PBS + NGS or BSA, and incubated in primary antibody overnight at 4°C (see below for antibodies and concentrations used). The next day, samples were washed with PBS + NGS/BSA and incubated with 1:200 secondary antibody for 1 h at room temperature (see below for list of secondary antibodies used). Samples were then washed with PBS, treated with 1 μg ml−1 DAPI (Life Technologies) for 15 min at room temperature, washed and treated with ProLong Gold antifade reagent (Life Technologies), then covered with a cover slip and allowed to cure overnight.

Images were collected on a Leica SP5 X MP inverted confocal microscope outfitted with a tunable white light laser at 20× magnification (0.7 NA, dry) in x, y and z. For the data shown in Fig. 2e and Extended Data Fig. 5b, c, images were acquired using standard detectors, and voltage and gain settings were maintained at a constant level for each set of images acquired for each pair of transcription factors of interest, as shown in Fig. 4e. Extended Data Fig. 5d. Images were acquired using HyD detectors in photon counting mode. z-intervals were spaced 1 μm apart, and 10 slices were taken centred around the area of maximum intensity for each region imaged. Four-channel imaging was performed to obtain brightfield, DAPI (ex 405 nm), TF (1 ex 564 nm) and TF (2 ex 647 nm) images.

Relative protein levels within individual cells were quantified with CellProfiler. Two-dimensional maximum projections of confocal z-stacks were input into CellProfiler, and cells were identified on the basis of OCT4 staining to exclude differentiated cells from the analysis. Performing the analysis by identifying cell nuclei on the basis of DAPI staining gave qualitatively similar results.

Antibodies used for quantitative immunofluorescence were as follows. Primary antibodies: OCT4, sc-5279 mouse monoclonal (1:100, Santa Cruz) or ab19857 rabbit polyclonal (1:100, Abcam); NANOG, ab80892 rabbit polyclonal (1:100, Abcam); ESRRB, PP-H6705-00 mouse monoclonal (1:100, R&D Systems); NR0B1, sc-841 rabbit polyclonal (1:100, Santa Cruz). Secondary antibodies: goat anti-mouse Cy5 (1:200, Life Technologies); goat anti-rabbit Alexa 568 (1:200, Life Technologies).

**miRNA expression analyses.** Cells used in the miRNA expression profiling experiment described in Fig. 5a were cultured under the conditions described above, and RNA was extracted and purified according to manufacturer’s instructions using a Qiagen miRNeasy kit. miRNAs were then quantified according to manufacturer’s protocols on a NanoString nCounter using the NanoString mouse miRNA Assay Kit, starting with 100 ng of purified RNA per sample. Data were normalized by dividing by the geometric mean of four housekeeping mRNAs (Actb, B2m, Gapdh and Rpl19). The NanoString mouse miRNA probe set includes eight negative control probes. Analysis of miRNAs was limited to those 427 probes with normalized counts that exceed the maximum negative control value in at least one of the six conditions profiled. Differentially expressed miRNAs were defined as those with |log2(fold change)| > 2.

qPCR validation of miRNA expression shown in Extended Data Fig. 9 was performed on a Fluidigm BioMark using a 48 × 48 expression array. Starting with purified miRNA from cell samples (300 ng total RNA per sample), miRNAs were reverse transcribed in the presence of a pool of TaqMan miRNA RT primers according to manufacturer’s instructions using the TaqMan MicroRNA Reverse Transcription kit (Applied Biosystems) in a 15 μl reaction volume. 2.5 μl of the reverse transcription reaction was then input into a 25 μl pre-amplification reaction performed using TaqMan PreAmp Master Mix and the corresponding pool of TaqMan miRNA PCR primers, and 14 cycles of pre-amplification were performed according to Fluidigm protocols. This amplification product was diluted 1:10 with TE(8.0), and 5 μl was submitted for expression profiling on a Fluidigm BioMark. Ct values were determined with Fluidigm Real-time PCR Analysis software, applying the auto detectors method and using a linear (derivative) baseline correction with a quality threshold of 0.65. Expression differences were determined using the ΔΔCT method, normalizing to the geometric mean of four reference small RNAs (snRNA 142, snRNA 202, snRNA 234 and U6 RNA) and comparing to v6.5 mESCs cultured in serum + lIF as the reference sample. Error bars represent the standard deviations of three technical replicate PCR reactions.

**Identification of predicted and experimentally determined miRNA target gene sets.** Predicted targets of ESCC miRNAs, the let-7 family of miRNAs, and miR-152 were determined using the programs TargetScan35 and mirRnap44. For TargetScan, we considered as predicted targets of these miRNAs families the set of all genes targeted by all predicted sites of conservation. mirmap considers conservation in its prediction algorithm. The $P$ values for changes in expression of predicted let-7 and mir-152 targets in the section ‘miRNA balance and cell state’ were calculated using mirmap predicted targets, while the plots shown in Fig. 5b show expression of TargetScan predicted target genes. Selected experimentally validated targets of let-7 or miR-152 that are not predicted by TargetScan (Myc for let-7, Dnmt3b for miR-152) are indicated in yellow in Fig. 5b.

**Self-renewal assays.** For the clonal self-renewal assays described in Fig. 5, cells were cultured as described above, plated at a density of 200 cells per well of a matrigel-coated Nunc tissue culture plate, and allowed to form colonies over the course of 5 days. Colonies were then fixed using the Vector Red Alkaline Phosphatase Substrate Staining kit (Vector Labs) using 200 mM Tris-HCl (8.5) plus 150 mM NaCl as the staining buffer, and fixed with 4% paraformaldehyde for 10 min at room temperature. Colonies were imaged on an Axio Zoom V16 (Zeiss) and assigned as showing either solid or mixed alkaline phosphatase staining by visual inspection.
We observe a strong correlation between uniformity of alkaline phosphatase staining within individual colonies and the uniformity of Nanog protein expression they exhibit, and therefore interpret this assay as reflective of the self-renewal capacity of individual mESCs.

**miRNA perturbation experiments.** For the Lin28a knockdown experiments shown in Fig. 5d, shLin28a cells were either treated with 1 μg ml⁻¹ doxycycline or left untreated and maintained in serum + LIF media for 3 days. After 3 days of doxycycline treatment, cells were harvested and plated for self-renewal assays as described above, except that doxycycline was added to treated cells during self-renewal assays. For the forced let-7 experiments shown in Fig. 5e, let-7 cells were treated with 10 μg ml⁻¹ doxycycline or left untreated and cultured for 5 days, passing once in between. By day 5 of doxycycline treatment, colony morphology showed a noticeable shift towards smaller, compact colonies. Cells were then harvested and plated for self-renewal or RNA FISH assays as described above, and doxycycline induction was continued during growth for these assays. For the miRNA inhibition experiments shown in Fig. 5f, miR-23b and miR-24 cells were transfected with miRNA inhibitors (Hsa-miR-16-5p and Hsa-miR-122-3p) using DharmaFect 1 (Dharmacon) and were maintained off of feeders at a concentration of 5,000 cells per well of a 24-well plate, and were transfected the following day with miRIDIAN miRNA mimics (Dharmacon) at a concentration of 50 nM each using Dharmafect 1. Media was changed after 2 days, and cells were harvested 3 days after transfection and plated for self-renewal assays.

**Polycomb target correlation with pluripotency factor expression.** Polycomb targets were identified as described above (‘Definition of Polycomb target genes for expression comparisons’). The proportion of Polycomb targets is the number of Polycomb targets that were detected (Lin1/TPM > 0) in each individual cell divided by the number of Polycomb targets that were detected in at least one cell profiled by RNA-seq (873). By this metric, any individual cell expressed between 6.0–21.5% of Polycomb target genes (Extended Data Fig. 8a). Pluripotency factor z-scores were calculated for each condition as:

\[
z = \frac{\text{score} - \mu(\text{expression})}{\sigma(\text{expression})}
\]

Pearson correlation coefficients were calculated on pluripotency factor z-scores and proportion of Polycomb targets expressed for each condition (Extended Data Fig. 8b).

**Statistics.** Below, we describe the statistical tests and/or sample sizes of all data comparisons shown in the main and Extended Data figures, or mentioned in the main text. Holm’s method corrected P-values are reported unless otherwise stated. No statistical method was used to predetermine sample size.

The P values for differences in H3K27me3 levels between detected and non-detected Polycomb target genes in ‘The landscape of gene expression variability in PSCs’ section were calculated using the Student’s t-test. P values for the significance of lower expression of let-7 and miR-152 target genes and higher expression of ESCC miRNA targets in 2i + LIF in the section ‘miRNA balance and cell state’ were calculated using the Mann–Whitney test. Calculation of the correlation in expression between predicted let-7 and miR-152 target genes is described in the ‘Single-cell correlation between expression of predicted let-7 or miR-152 target genes’ section of Supplementary Information. P values for significance of differences between high and low states shown in Fig. 2g were calculated using the Kolmogorov–Smirnov test.

For Fig. 1b, the number of colonies and cells examined was 318 cells from 5 colonies (Oct4), 1,226 from 10 colonies (Esrβ), and 551 cells from 5 colonies (Neurod1). For Fig. 1c, COV enrichment of biological pathway annotations. COV-based enrichment test.

For Fig. 4d, serum + LIF: 1,226 cells from 10 colonies (same data as in Figs 1b and 2c); 2i + LIF: 557 cells from 5 colonies; Dgcr8⁻/⁻: 187 cells from 5 colonies. For Fig. 4e, serum + LIF: 901 cells; 2i + LIF: 1,648 cells; Dgcr8⁻/⁻: 248 cells.

For Fig. 5a, miRNAs with log2(fold change) > 2 (mESC in 2i + LIF versus FBS + LIF) are shown.

For Extended Data Fig. 6b, number of cells in each condition: v6.5 on MEFs (75), v6.5 FF (83), E14 FF (77), v6.5 2i + LIF (77), E14 2i + LIF (87), Dgcr8⁻/⁻ (81), iPSCs (57), Hoechst (84), Dnmt1⁻/⁻ Dnmt3a⁻/⁻ Dnmt3b⁻/⁻ (34), Eed⁻/⁻ (54), BIX (30), TSA (30), VPA (30), Mbd3 (40), EpiSCs (41). Extended Data Fig. 6c, changes in the distributions of gene expression (by qPCR) were tested by two-sample Kolmogorov–Smirnov on the Ct values in each perturbation or condition versus the v6.5 serum + LIF sample, and P values were adjusted using the Holm’s method to correct for multiple tests.

For Extended Data Fig. 5b, NANOg and OCT4 in serum + LIF: 2,469 cells. Extended Data Fig. 5c, NROB1 and OCT4 in serum + LIF: 2,917 cells.

For Extended Data Fig. 7e, Oct4: serum + LIF (5 colonies, 318 cells (same data as in Figs 1b and 2c)), 2i + LIF (5 colonies, 485 cells), Dgcr8⁻/⁻ (5 colonies, 175 cells). Nanog: serum + LIF (5 colonies, 151 cells (same data as in Fig. 2c)), 2i + LIF (5 colonies, 398 cells), Dgcr8⁻/⁻ (5 colonies, 145 cells). Nr0b1: serum + LIF (8 colonies, 849 cells (same data as in Fig. 2c)), 2i + LIF (5 colonies, 429 cells), Dgcr8⁻/⁻ (5 colonies, 283 cells), Oct2: serum + LIF (5 colonies, 151 cells (same data as in Fig. 2c)), 2i + LIF (5 colonies, 398 cells), Dgcr8⁻/⁻ (5 colonies, 144 cells). Nanog and Oct2 were imaged together in the same set of cells using two colour RNA FISH. Extended Data Fig. 7g, average expression values in serum + LIF are the same shown in Extended Data Fig. 3a. For 2i + LIF, average expression values are given for Oct4, Nanog, Nr0b1 and Oct2 the distributions shown in Extended Data Fig. 7e (above). For the remaining 2i + LIF genes, numbers of cells are as follows: Myc and Isl1: 6 colonies, 785 cells; Lin28a and Olig2: 5 colonies, 354 cells; Esrβ: 10 colonies, 903 cells; Nkx2-9: 5 colonies, 345 cells; Sox2 and Neurod1: 5 colonies, 649 cells; Pax6: 5 colonies, 429 cells; and Dazl: 5 colonies, 485 cells. Oct4 and Dazl, Sox2 and Neurod1, Myc and Isl1, Lin28a and Olig2, and Nanog and Otx2 were each imaged by two colour RNA FISH in the same set of cells. Extended Data Fig. 7h, serum + LIF distributions are the same shown in Extended Data Fig. 5b, c. NANOg and OCT4 in 2i + LIF: 518 cells. NROB1 and OCT4 in 2i + LIF: 827 cells.

49. Sommer, C. A. et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells 27, 543–549 (2009).
50. Park, I. H., Lerou, P. H., Zhao, R., Huo, H. & Daley, G. Q. Generation of human-induced pluripotent stem cells. Nature 393, 877–879 (2008).
51. Dow, L. E. et al. A pipeline for the generation of shRNA transgenic mice. Nature Biotechnol. 18, 675–679 (2000).
52. Okabe, S., Forsberg-Nilsson, K., Spio, A. C., Segal, M. & McKay, R. D. Development of neural precursor cell cultures and functional postmitotic neurons from embryonic stem cell lines. J. Neurosci. 18, 4373–4382 (1998).
53. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20 (2005).
64. Vejnar, C. E. & Zdobnov, E. M. MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.* **40**, 11673–11683 (2012).
65. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
66. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
67. Matthews, L. et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, 619–622 (2009).
68. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **99**, 611–631 (2002).
69. Cahan, P. et al. CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).
Extended Data Figure 1 | Quality control of single-cell RNA-seq data. a, A combination of read alignment rate (y axis) and number of genes detected (ln(TPM) > 1) (x axis) was used to identify outlier cells (red circles) to remove from subsequent analysis, leaving 183 single mESCs cultured in serum + LIF, 94 mESCs cultured in 2i + LIF, and 84 Dgcr8-/- mESCs cultured in serum + LIF that were analysed by single-cell RNA-seq in this study (blue circles).
b, Correlation in mean expression in detected cells (µ) between replicate serum + LIF plates across a range of thresholds. µ was calculated separately for each plate. Correlations in µ were calculated after limiting genes to those with thresholds exceeding the specified threshold on the x axis. c, Maximum gene expression in replicate plates. Genes not reliably detected as defined in the text are coloured red.
d, Comparison of δ estimates in replicate plates, limited to reliably detected genes. Selected lineage regulators are denoted. e, Comparison of µ estimates in replicate plates, limited to reliably detected genes. Selected housekeeping, pluripotency and signalling genes are denoted. f, Top: relationship between estimates of δ (x axis) and µ (y axis) of all genes based on both plates of mESC in FBS + LIF. Undetected genes are coloured yellow. Bottom: same as above except only showing reliably detected genes, and overlaid with density contour (red lines).
Extended Data Figure 2 | Examination of mESCs cultured in serum+LIF for the presence of distinct subpopulations. a, Hierarchical clustering dendrogram of single-cell RNA-seq data for 183 mESCs cultured in serum+LIF. b, Principal component analysis of the 183 mESCs cultured in serum+LIF. Points coloured blue are those with PCA1 values ≤ − 25, which are classified as outlier cells. c, Box plots of expression of selected pluripotency regulators and the lineage regulator Pux3 (top), and genes associated with EpiSCs (bottom). ‘Normal’ indicates the majority of cells coloured as grey dots in Extended Data Fig. 2b; ‘Outliers’ indicates the distinct set of 14 cells coloured blue in Extended Data Fig. 2b. P values for statistically significant differences are shown. d, Histograms showing the expression distributions of pluripotency regulators previously found to fluctuate within mESC populations. Cutoffs to divide expression into high and low states to test for enrichment within outlier cells are indicated by dashed lines. e, Average expression of Polycomb target genes within outlier cells (left), and the majority of the mESCs cultured in serum+LIF (right).
**Pluripotency regulators**

- Pou5f1
- Tcfcp21
- Esrb
- Nanog
- Nr0b1
- Sox2
- Zfp42
- Myc
- Lin28a

**Lineage regulators / Polycomb targets**

- Dazl
- Nkx2.9
- Pax6
- Pax3
- Isl1
- Otx2
- Olig2
- NeuroD1
- Bmp4
Extended Data Figure 3 | Correlation of single-cell expression measurements between different technologies. a, Correlation between transcripts per million (TPM) measured by single-cell RNA-seq and transcripts per cell measured by single-molecule FISH for 14 selected genes in mESCs cultured in serum + LIF. Error bars represent standard deviations of measurements. b, Correlation between coefficients of variation measured by single-cell RNA-seq and FISH for the 14 genes shown in panel a. c, Correlation coefficients for $z$ (fraction detected, left) or $\mu$ (mean expression in detected cells, right) between single-molecule FISH and single-cell RNA-seq are plotted as a function of varying the threshold level for detection by RNA FISH ($x$ axis). An RNA FISH detection threshold of 10 indicates that genes expressed at $\leq$ 10 copies per cell would not be detected by RNA FISH. Correlation for $z$ between RNA-seq and RNA FISH peaked at an RNA FISH detection threshold of $>5$ transcripts per cell, giving an estimated single-cell RNA-seq detection efficiency of $\sim$20% (1 out of 5 transcripts detected, assuming single-molecule sensitivity for the RNA FISH method). d, Correlation in $z$ between single-cell RNA-seq and single-molecule FISH for 14 genes measured by both methods, assuming a single-molecule FISH detection threshold of $>5$ transcripts per cell. Dashed line shows linear fit to the data. The fraction of cells a gene is detected in shows good agreement between the two methods when taking the sensitivity of the RNA-seq into account. e, Comparison of the fraction of mESCs cultured in serum + LIF a gene was detected in by single-cell qPCR ($x$ axis) or single-cell RNA-seq ($y$ axis). Single-cell RNA-seq showed greater sensitivity overall as compared to single-cell qPCR, but a set of genes was sporadically expressed as measured by both methods. Trend line indicates linear fit to the data. f, Correlations of fraction detected between independent biological replicates for 96 genes profiled by single-cell qPCR. Trend line shows linear fit to the data, and indicates that the fraction of cells a gene is detected in remains consistent across independent biological replicates. g, Comparison of expression distributions measured by single-cell RNA-seq (light grey) and single-molecule FISH (darker grey) for pluripotency regulators (top) and Polycomb target genes (bottom).
Extended Data Figure 4  | Expression of Polycomb target genes in ESCs and neural precursor cells (NPCs).  

a, Polycomb target genes show highly variable expression in mESCs. Relationship between \( \mu \) (mean expression in detected cells, x axis) and coefficient of variation (standard deviation normalized by mean expression in all cells, y axis) is shown for Polycomb target genes (purple) and non-Polycomb target genes (grey). Distributions for \( \mu \) and coefficients of variation for the two gene sets are shown above and to the left of the graph, respectively. Polycomb target genes show pronounced variability in expression, even when controlling for expression level.

b, Expression of the neural regulators and Polycomb target genes Otx2 (top) and Neurod1 (bottom) measured by RNA FISH in mESCs cultured in serum + LIF. Overall distributions within the population (left) and representative colonies (right) are shown, along with gene tracks from the Broad Institute Integrated Genomics Viewer (IGV) showing ChIP-seq reads for H3K27me3 at the Otx2 and Neurod1 genes.

c, Enriched gene ontology categories among genes significantly upregulated (red) or downregulated (green) in neural precursor cells as compared to embryonic stem cells.

d, Expression changes of selected genes in neural precursor cells as compared to the embryonic stem cells they were derived from. As expected, neural regulators and ES Polycomb target genes Pax6, Prrx1, Hes1, Msx1, Pbx3, Nes and Runx1 were upregulated in NPCs, while the pluripotency regulators Esrrb, Nr0b1, Pou5f1 (Oct4), Zfp42 (Rex1) and Nanog were downregulated. Expression of the housekeeping genes Gapdh and Actb, and the pluripotency and neural regulator Sox2, were relatively unchanged between the two cell types.

e, Expression comparison of ES Polycomb target genes that are detected in either mESCs or NPCs. Many Polycomb target genes that are neuronal regulators are detected in a higher fraction of NPCs than ESCs, while certain Polycomb targets such as Pax3 (a regulator of musculoskeletal development) are detected in a smaller fraction of cells. Genes are ordered in ascending order of ESC to NPC average expression change.

f, Histograms showing distributions of expression levels for selected housekeeping genes (grey) and neuronal regulators (green and red) in NPCs. The neuronal regulators Msx1 and Runx1 (red) show bimodal expression in NPCs.

g, Distance distributions within ESC (red) and NPC (blue) populations. NPCs show more population heterogeneity than ESCs.

h, State classification based on principal component analysis of single-cell RNA-seq of NPCs. Four distinct states are identified.
Extended Data Figure 5 | Fluctuations in pluripotency and lineage regulator expression. a, RNA FISH images showing expression of Nanog (top), Nr0b1 (middle) and Esrrb (bottom) in individual colonies or regions of cells. b, c, Histograms show distributions of fluorescence intensities within individual cells from quantitative immunofluorescence of OCT4 and (b) NANOG, or (c) NR0B1, along with NANOG/OCT4 or NR0B1/OCT4 ratios as indicated. For NANOG/OCT4 images (b), NANOG staining is coloured green while OCT4 staining is coloured red. In the panel on the left a cluster of low NANOG cells is indicated by a blue arrow, while a cluster of high NANOG cells is indicated by a white arrow. In the panel on the right, the same image is shown with DAPI staining coloured blue, and groups of OCT4 negative/NANOG negative differentiated cells that may have arisen from the low NANOG cells are indicated with grey arrows. For NR0B1/OCT4 images (c), NR0B1 is coloured green while OCT4 staining is coloured red. A relatively high NR0B1 colony is shown in the panel on the left, while a region of low NR0B1 cells is displayed in the panel on the right. d, Quantitative immunofluorescence images showing expression of OCT4 (red) and NR0B1 (green, top row), NANOG (green, middle row), or ESRRB (green, bottom row) within individual colonies of mESCs grown in serum + LIF. OCT4 is used as an internal reference as it shows relatively invariant expression within mESCs. e, Single-cell RNA-seq relationships for gene pairs shown in Fig. 2g.

Distributions of gene expression from RNA-seq and RNA FISH experiments are shown. Dashed lines indicate divisions between high and low states for box plot shown in Fig. 2g. Single-cell RNA-seq data show that the subset of cells in both a high Sox2 and low Nr0b1 state show an increased probability of expressing Neurod1 [ln(TPM) >1] as compared to all cells (bottom). f, RNA-seq (left) and RNA FISH (right) correlations between the pluripotency regulator Esrrb and the signalling factor Bmp4 (top), and the pluripotency regulator Nanog and the neural regulator Otx2 (bottom). Dashed lines indicate divisions between high and low states for the box plots shown on the right. g, Correlations from single-cell RNA-seq data between average Polycomb target gene expression and Zfp42 (Rex1), Nanog and Nr0b1. Dashed lines indicate divisions between high and low states for the box plot comparing Polycomb target expression with expression of the three regulators, which are all negatively associated with expression of Polycomb targets. The Venn diagram shows coupling between high and low states of the three regulators. Low Nr0b1 cells are more likely to be in a low Zfp42 and/or low Nanog state as compared to high Nr0b1 cells, suggesting that Nr0b1 functions to maintain Zfp42 and Nanog expression and repress Polycomb target genes. h, RNA FISH images of an ESC colony hybridized with probes against Nanog (yellow) and Otx2 (magenta), showing inverse relationship between Nanog and Otx2 expression.
Extended Data Figure 6 | Single-cell qPCR of mESCs exposed to chemical and genetic perturbations. a, Shown are the average Ct values and standard deviations for technical triplicates (error bars on x axis) or biological triplicates (error bars on y axis) across 96 genes for pools of 100 or 10 cells or single sorted cells. Cells were sorted into PCR strips containing RT–PCR reagents and primer pools, reverse transcription and pre-amplification was performed, and cDNA was quantified on a Fluidigm BioMark PCR system. b, Heat map of single-cell qPCR data for 84 genes examined across 19 different PSC perturbations and in MEFs (n = 1,144 single cells). Unsupervised hierarchical clustering grouped genes into three clusters: bimodally expressed genes (right group), ubiquitously expressed genes (left group), and sporadically expressed genes (middle group). c, Numbers of genes showing significant changes in expression distributions as compared to the reference conditions of v6.5 mESCs cultured in serum + LIF on MEFs. Significance of changes was determined by the two-sample Kolmogorov–Smirnov test, correcting P values for multiple tests using Holm’s method. d, Selection of the state classification model that maximizes the Bayesian Information Criteria (BIC, y axis). Mclust was used to generate multivariate Gaussian mixture models of the first three principal components of the Fluidigm qPCR-based expression values of individual mESCs. The models vary in the number of components (one to nine) and the following geometric characteristics: volume, shape and orientation as described. The best model was used to classify cells into states.
Extended Data Figure 7 | Gene expression changes in mESCs upon culture in 2i or removal of mature miRNAs. a, Changes in expression of the pluripotency regulators shown in Fig. 4c going from serum + LIF to 2i + LIF culture (x axis), or between wild-type and Dgcr8⁻/⁻ cells cultured in serum + LIF (y axis), as measured by single-cell RNA-seq. Selected genes are highlighted. b, Changes in expression of 18 commonly used housekeeping genes (ActB, Aip, Cxxc1, Gapdh, Gusb, Hmbs, Hprt, Ip08, Mrpl48, Mtcp1, Pk1, Ppia, Rpl13a, Rplp2, Rp56, Tbp, Ubc and Ywhaz) between the same conditions as in panel a. c, Intra-condition (left) and inter-condition (right) distances between individual cells based on single-cell RNA-seq data for all genes (top), 219 transcription factors that regulate pluripotent cells as determined by CellNet69 (middle), or lineage regulators, defined as the 256 previously determined Polycomb target genes in mESCs that are transcription factors (bottom). d, Comparison of single-cell average expression changes going from serum + LIF to 2i + LIF culture in the present study (x axis) against population-level expression changes between mESCs cultured in serum + LIF versus 2i + LIF measured by ref. 40 (y axis). Trend line from linear fit to the data are shown, and selected genes that show lower expression in 2i + LIF culture in both studies are highlighted. e, Single-molecule FISH showing shifts in expression of Oct4, Nanog, Nr0b1 and Otx2 at the RNA level between wild-type mESCs in serum + LIF and 2i + LIF culture conditions and Dgcr8⁻/⁻ mESCs cultured in serum + LIF. f, Representative RNA FISH images showing expression of the Polycomb target gene and neural regulator Otx2 in individual mESC colonies under the three conditions examined. g, Correlation between expression shifts between serum + LIF and 2i + LIF culture observed by single-cell RNA-seq (x axis) and RNA FISH (y axis) for the 14 genes shown in Extended Data Fig. 3a. Trendline indicates linear fit to the data. h, Quantitative immunofluorescence showing changes in NANOG/OCT4 (left) and NR0B1/OCT4 (right) ratios between serum + LIF and 2i + LIF culture. Serum + LIF data are the same shown in Extended Data Fig. 5. i, RNA FISH images of E14 mESC colonies cultured in serum + LIF (left) or 2i + LIF (right) media and probed for Nanog (top) or Esrrb (bottom) expression. Both Nanog and Esrrb show bimodal expression patterns in E14 mESCs grown in serum + LIF, and shift towards the high expression state in 2i + LIF culture. White arrows indicate regions of low Nanog or Esrrb expression in mESCs grown in serum + LIF.
Extended Data Figure 8 | Dependence of Polycomb target gene expression on culture conditions and miRNAs. a, Fraction of Polycomb target genes detected in wild-type mESCs cultured in serum+LIF (red) or 2i+LIF (green), or Dgcr8^−/− mESCs cultured in serum+LIF (blue). b, Correlation between Polycomb target gene expression and pluripotency regulator expression in different conditions (serum+LIF in blue, Dgcr8^−/− in orange, and 2i+LIF in red). Displayed are the Pearson correlation coefficients (PCC) between pluripotency-related regulator z-score and proportion of Polycomb targets detected, computed across all single cells. The z-score is defined as the number of standard deviations that a sample exceeds (z-score >0) or is less than (<0) the mean value. z-scores for pluripotency regulators were computed for each condition separately. A low PCC indicates that a lower factor expression (relative to its mean in the condition) increases the likelihood that Polycomb targets will be detected as expressed (for example, Zfp42 in FBS+Lif). c, Scatter plots comparing amount of H3K9me3 (top), H3K27ac (middle) and RNA polymerase II (bottom) at promoter regions in wild-type mESCs cultured in serum+LIF versus 2i+LIF conditions, wild-type mESCs versus Dgcr8^−/− mESCs cultured in serum+LIF, or Dgcr8^−/− mESCs cultured in serum+LIF versus wild-type mESCs cultured in 2i+LIF as indicated. ChIP-seq reads at gene promoters were median normalized for comparison, and Polycomb target genes are indicated in green. Unlike H3K27me3, levels of these three factors do not show a strong decrease at Polycomb target genes under 2i+LIF conditions and in Dgcr8^−/− mESCs (compare to Fig. 4f).
Extended Data Figure 9 | Perturbing miRNA balance and the Myc/Lin28/let-7 axis. a, Purified RNA from v6.5 mESCs and E14 mESCs cultured in either serum+LIF or 2i+LIF conditions was extracted and reverse-transcribed with TaqMan primers specific to the indicated miRNAs, and then expression was profiled by qPCR. Error bars represent standard deviation from technical triplicate PCR reactions, and samples are normalized to a basket of reference small noncoding RNAs and the decrease in Ct values compared to v6.5 mESCs grown in serum+LIF is shown. See Methods for full details. b, Fold-change in expression of the indicated miRNAs in wild-type mESCs cultured in serum+LIF or 2i+LIF, induced and uninduced ilet-7 mESCs cultured in serum+LIF, and MEFs grown under standard conditions. Changes are shown relative to v6.5 mESCs grown in serum+LIF. c, Comparison of genes that change in expression upon introduction of the ESCC miRNA miR-294 to Dgcr8−/− mESCs in ref. 44 (y axis) to genes that change in expression between serum+LIF and 2i+LIF culture in single-cell RNA-seq data (x axis). Genes that are significantly upregulated in Dgcr8−/− mESCs upon miR-294 introduction are indicated in blue, and those that are significantly downregulated are indicated in red. Selected genes upregulated by miR-294 and downregulated in 2i+LIF culture as compared to serum+LIF are highlighted. As a group, genes downregulated by miR-294 show higher expression in 2i+LIF than in serum+LIF. d, mESC colonies staining uniformly positive for alkaline phosphatase show reduced levels of Myc. Overall and colony-specific Myc distribution in serum+LIF culture as measured by RNA FISH, showing uniformly positive (top), mixed (middle), or negative (bottom) alkaline phosphatase staining. e, Western blot showing reduction of Lin28a protein levels in iShLin28a cells upon addition of doxycycline. f, let-7 expression changes in doxycycline-inducible ilet-7 mESCs grown in serum+LIF upon induction. RT–qPCR was performed as in panel a, and Ct changes are shown relative to v6.5 mESCs in serum+LIF in a separate experiment from panel a. The inducible let-7 construct is detected by the let-7g probe. g, Effect of miRNA transfection on self-renewal efficiency of Dgcr8−/− mESCs. miRNA mimics were transfected into Dgcr8−/− mESCs, and self-renewal efficiency was measured. Error bars indicate standard deviations between triplicate transfection experiments. Co-transfection of the ESCC miRNA miR-294 with a let-7 miRNA results in enhanced self-renewal efficiency as compared to miR-294 alone. h, Expression changes of selected genes measured by qPCR upon culture of v6.5 mESCs in serum+LIF, 2i+LIF, or treatment with only Erk or GSK3β inhibitors.
Extended Data Table 1 | List of genetic and chemical perturbations whose effects were profiled by single-cell qPCR

| Perturbation                        | Effect                                      | Cell line | Growth media | Feeders | Abbreviation        | Ref. |
|-------------------------------------|---------------------------------------------|-----------|--------------|---------|---------------------|------|
| **Culture conditions**              |                                             |           |              |         |                     |      |
| Standard (v6.5 MEFs)                | Baseline conditions                         | v6.5 mESCs | Serum+LIF    | Yes     | v6.5 on MEFs        |      |
| Feeder-free                         | Increased spontaneous differentiation         | v6.5 mESCs | Serum+LIF    | No      | v6.5 FF             |      |
| Feeder-free (E14)                   | Increased spontaneous differentiation         | E14Tg2A   | Serum+LIF    | No      | E14 FF              |      |
| Ground state                        | Increased self-renewal efficiency            | v6.5 mESCs | 2i+LIF       | No      | V6.5 2i+LIF         | 13   |
| Ground state (E14)                  | Increased self-renewal efficiency            | E14Tg2A   | 2i+LIF       | No      | E14 2i+LIF          | 13   |
| **Genetic knockouts**               |                                             |           |              |         |                     |      |
| Polycomb repressive complex 2       | Disruption of the PRC2 complex, loss of H3K27me3 | Eed<sup>-/-</sup> | Serum+LIF | Yes     | Eed<sup>-/-</sup>   | 31   |
| DNA methyltransferases              | Loss of DNA methylation                      | Dnmt TKO | Serum+LIF    | Yes     | Dnmt TKO            | 33   |
| MicroRNA processing                 | Loss of mature miRNAs (Knockout of Dgcr8)   | Dgcr8<sup>-/-</sup> | Serum+LIF | Yes     | Dgcr8<sup>-/-</sup> on MEFs | 34   |
|                                    | Loss of mature miRNAs (Knockout of Dicer)   | Dicer<sup>-/-</sup> | Serum+LIF | No      | Dicer<sup>-/-</sup> on MEFs | 30   |
| NuRD complex                        | Disruption of the NuRD complex              | Mbd3<sup>-/-</sup> | Serum+LIF | Yes     | Mbd3<sup>-/-</sup>  | 32   |
| **Chemical treatments**             |                                             |           |              |         |                     |      |
| Trichostatin A                      | HDAC inhibition                              | v6.5 mESCs | Serum+LIF    | Yes     | TSA                 |      |
| Valproic acid                       | HDAC inhibition                              | v6.5 mESCs | Serum+LIF    | Yes     | VPA                 |      |
| BIX01294                            | Inhibition of G9a histone methyltransferase  | v6.5 mESCs | Serum+LIF    | Yes     | BIX                 |      |
| Hoechst 33342                       | Cell cycle analysis (cells in G<sub>0</sub>/G<sub>1</sub> were isolated) | v6.5 mESCs | Serum+LIF    | Yes     | Hoechst             |      |
| **Alternate cell types**            |                                             |           |              |         |                     |      |
| Induced pluripotent stem cells      | Reprogrammed from somatic cells              | iPSCs     | Serum+LIF    | Yes     | iPSCs               |      |
| Epiblast stem cells                 | Primed state of pluripotency                | EpiSCs    | EpiSC growth media | Yes | EpiSCs | 35 |

©2014 Macmillan Publishers Limited. All rights reserved
Erratum: Deconstructing transcriptional heterogeneity in pluripotent stem cells

Roshan M. Kumar, Patrick Cahan, Alex K. Shalek, Rahul Satija, A. Jay Daley Keyser, Hu Li, Jin Zhang, Keith Pardee, David Gennert, John J. Trombetta, Thomas C. Ferrante, Aviv Regev, George Q. Daley & James J. Collins

Nature 516, 56–61 (2014); doi:10.1038/nature13920

In this Letter, owing to a production error, in Fig. 3e the sixth bar from the left was inadvertently labelled ‘Dgcr8−/− FF’ instead of ‘Dicer−/− FF’. This has now been corrected in the online versions of the paper.