ARTICLE

doi:10.1038/nature13920

Deconstructing transcriptional heterogeneity in pluripotent stem cells

Roshan M. Kumar[1,2]*, Patrick Cahan[3]*, Alex K. Shalek[4], Rahul Satija[5], AJay DaleyKeyser[1], Hu Li[6], Jin Zhang[3], Keith Pardee[1,2], David Gennert[5], John J. Trombetta[5], Thomas C. Ferrante[1], Aviv Regev[5,7], George Q. Daley[3] & James J. Collins[1,2]

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 circuitry 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 fates[1–5]. 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 pluripotency[5].

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 program[6–13], which may be fundamental to pluripotency[14–23]. 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 first used single-cell RNA-seq[24,25] 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 *Esrrb* (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 *Neurod1* (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 targets[26] 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 PSCs[26,27].

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA. [2]Howard Hughes Medical Institute, Department of Biomedical Engineering, Center of Synthetic Biology, Boston University, Boston, Massachusetts 02215, USA. [3]Stem Cell Transplantation Program, Division of Pediatric Hematology and Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Boston Children's Hospital and Dana Farber Cancer Institute, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. [4]Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. [5]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [6]Center for Individualized Medicine, Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. [7]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

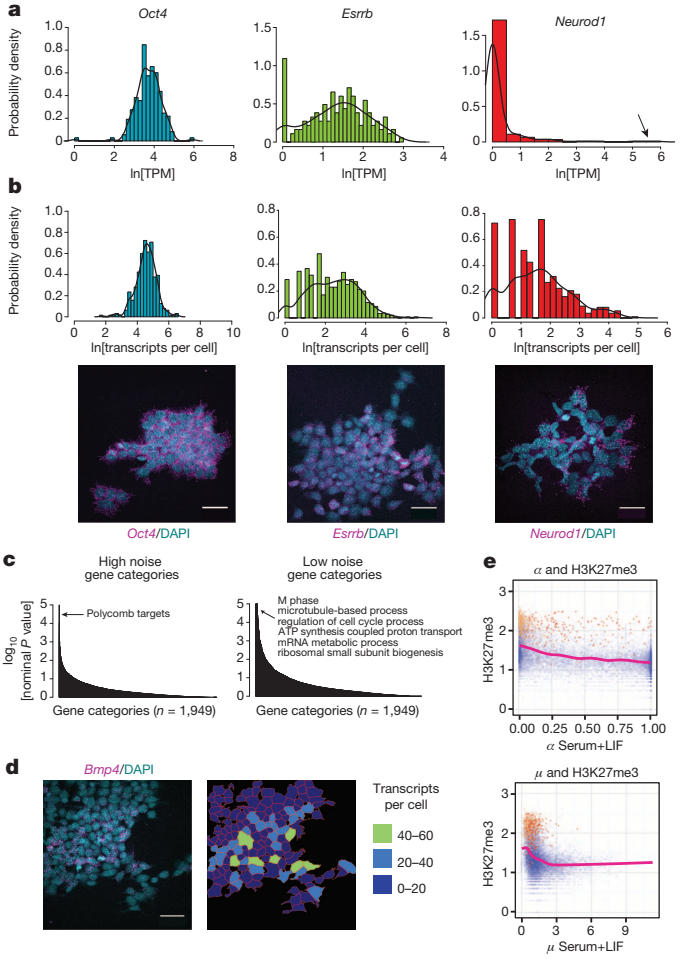


Figure 1 | Gene expression variability landscape of PSCs. **a**, Histograms of transcript distributions from single-cell RNA-seq of v6.5 mESCs cultured in serum+LIF. Arrow indicates high *Neurod1*-expressing cells. TPM, transcripts per million. **b**, Histograms and representative images of transcript 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. **e**, Relationship between population H3K27me3 levels, fraction of cells a gene is detected in (α, top) and average expression level when detected (μ, bottom). Detected (red) and undetected (orange) Polycomb target genes are highlighted. Overall trend lines are shown. All relevant statistical information can be found in the 'Statistics' section of Methods. Scale bar, 40 μm.

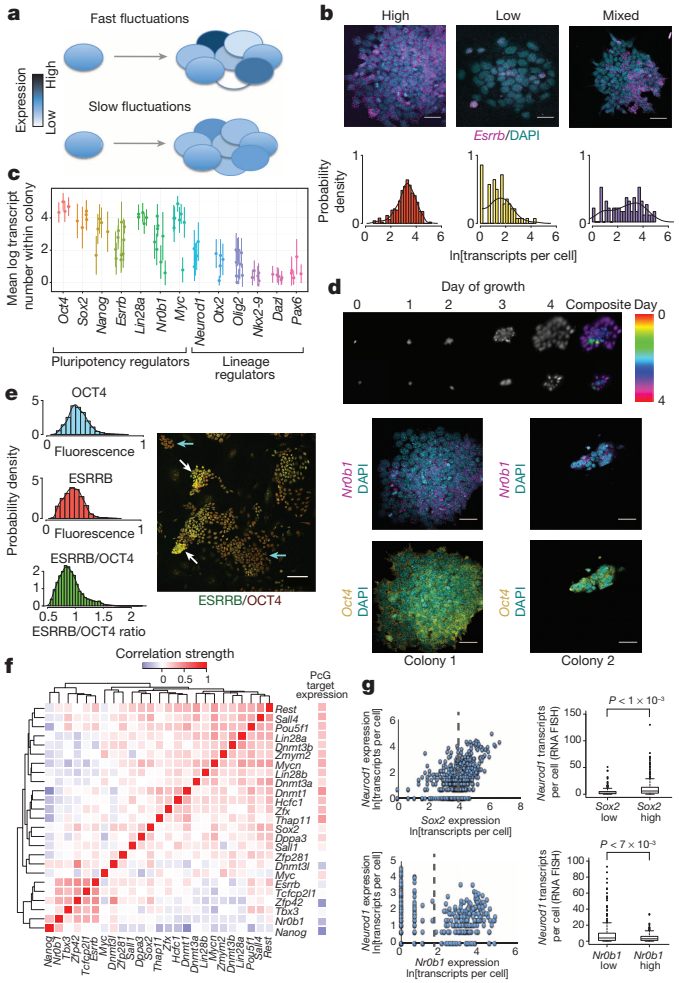


Figure 2 | Expression states of variable genes are coupled together and persist over multiple cell divisions. **a**, Slowly fluctuating genes show a high degree of intercolony variability. **b**, Expression of the pluripotency regulator *Esrrb* within individual colonies. **c**, Intra- and inter-colony variability in expression for selected pluripotency and lineage regulators. Average transcript number and standard deviation within colonies are indicated. **d**, Time-lapse imaging of colony formation from single cells, and *Nr0b1* and *Oct4* expression within these colonies. **e**, Relative OCT4 and ESRRB protein levels within mESCs cultured in serum+LIF. Groups of high and low ESRRB cells are indicated. **f**, Correlation of pluripotency regulator and Polycomb target gene expression between individual cells. PcG, Polycomb-group protein. **g**, Dependence of *Neurod1* expression on the level of *Sox2* and *Nr0b1* within individual cells. Dashed lines indicate high and low expression states, and *P* values for differences between states were calculated using the Kolmogorov–Smirnov test. Scale bars: **b**, **d**, 40 μm; **e**, 0.01 cm.

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 particularly 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*, *Otx2*, *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[6,9,28]. 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

©2014 Macmillan Publishers Limited. All rights reserved

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[29].

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 *Rex1*) 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, *Neurod1* 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 *Esrrb*, 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 conditions[13], treated with chemical inhibitors of epigenetic regulators, genetically modified to lack particular regulators[30–34], or in different states of pluripotency[35] (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 differentiation[13]. Notably, both culturing mESCs in 2i+LIF conditions[13] and impairing microRNA (miRNA) production[30,34] 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 *Dgcr8*$^{-/-}$ and *Dicer* knockout mESCs, which lack mature miRNAs due to loss-of-function of separate miRNA processing factors[30,34], 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

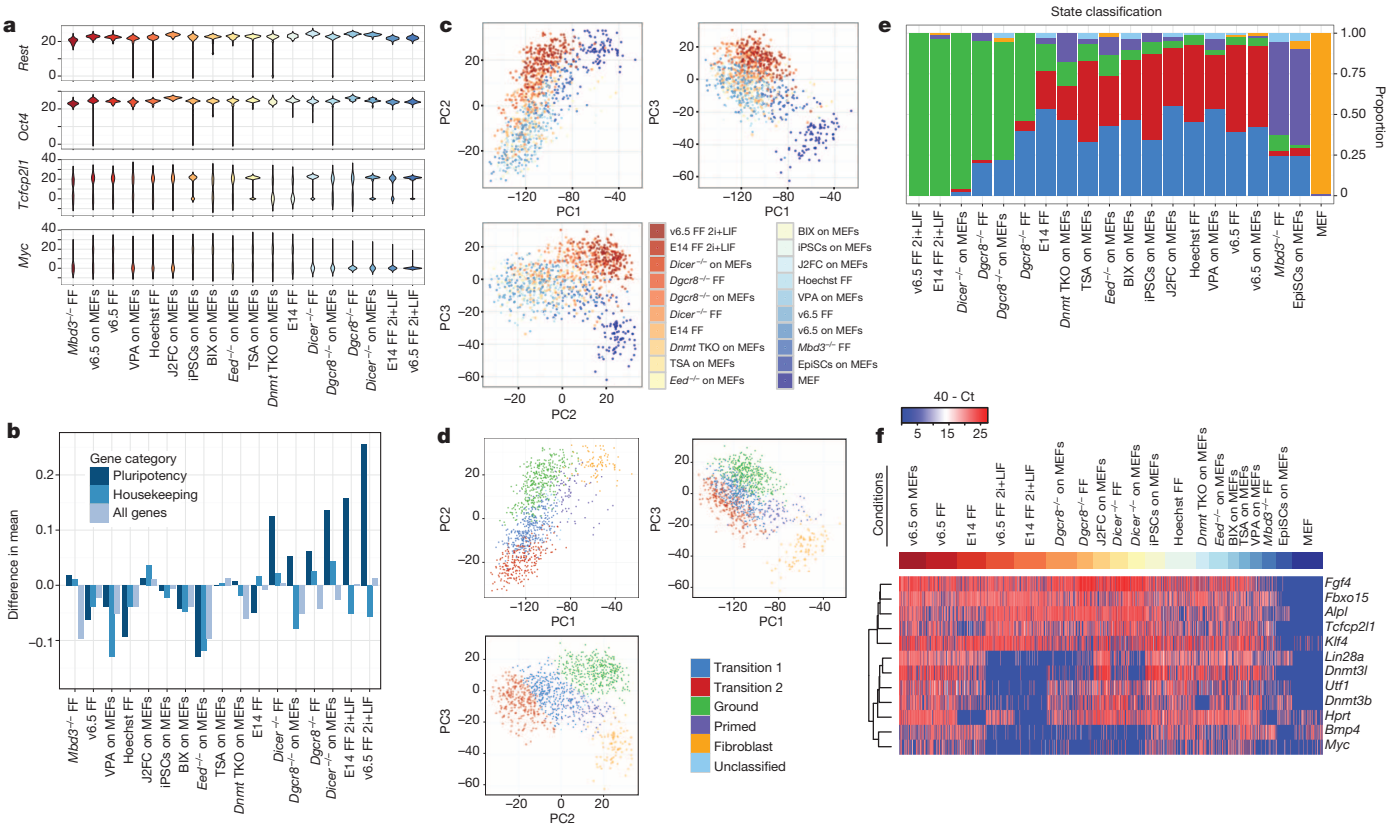


Figure 3 | Effect of perturbations on gene expression variability and cell state. **a**, Population distributions for the unimodally expressed genes *Oct4* and *Rest*, and the bimodally expressed genes *Tcfcp2l1* and *Myc*. iPSCs, induced pluripotent stem cells; MEFs, mouse embryonic fibroblasts; TKO, 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 heat map of genes contributing the most to the top three principal components, excluding housekeeping and fibroblast genes.

©2014 Macmillan Publishers Limited. All rights reserved

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[35]. Genes contributing the most to the principal components that distinguished *Dgcr8*$^{-/-}$ 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*$^{-/-}$ mESCs cultured in serum+LIF (*Dgcr8*$^{-/-}$ 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*$^{-/-}$ 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 low-noise ground state[20,28,37] (Fig. 4b and Extended Data Fig. 7). By contrast, *Dgcr8*$^{-/-}$ mESCs displayed increased population heterogeneity, consistent with a role for miRNAs in buffering gene expression noise[38]. When compared across conditions, however, individual *Dgcr8*$^{-/-}$ mESCs were more similar to 2i+LIF than serum+LIF mESCs, supporting a model in which removal of miRNAs drives a portion of *Dgcr8*$^{-/-}$ 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*$^{-/-}$ 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*$^{-/-}$ mESCs, suggesting that miRNA activity may be linked to the reduction in DNA methylation observed in the naive pluripotent state[28,39]. Regulators displaying bimodal expression patterns in serum+LIF mESCs, including the pluripotency factor Esrrb, showed altered distributions in both 2i+LIF and *Dgcr8*$^{-/-}$ 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[40]. 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*$^{-/-}$ 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[36], 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[41–43], 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[44] run contrary to our observed co-expression of ESCC and let-7 miRNAs in 2i+LIF mESCs. Analysing our own and published data[40], 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

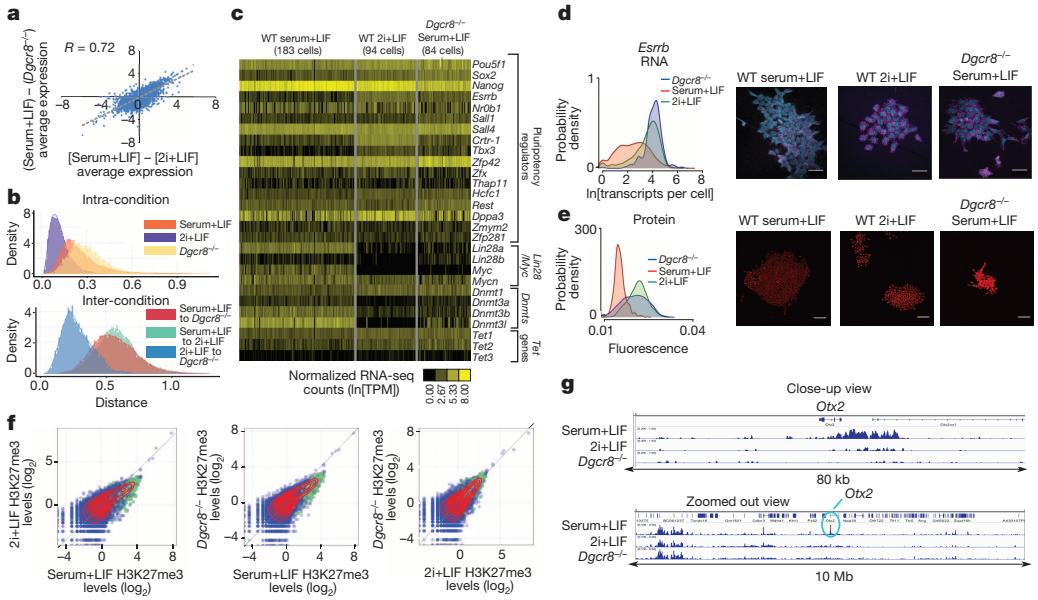


Figure 4 | *Dgcr8*$^{-/-}$ mESCs show evidence of ground-state self-renewal. **a**, Correlation between single-cell RNA-seq gene expression changes in different conditions. **b**, Distances between individual cells for pluripotency regulators shown in panel **c**. **c**, Heat map of single-cell RNA-seq data for selected pluripotency regulators. **d**, **e**, Single-molecule FISH (**d**) and quantitative immunofluorescence (**e**) showing a shift towards the high expression state of *Esrrb* in 2i+LIF and *Dgcr8*$^{-/-}$ mESCs. Scale bars: 40 μm (top row); 100 μm (bottom row). **f**, Promoter H3K27me3 levels in the three conditions examined by single-cell RNA-seq. Polycomb target genes are shown in green. **g**, H3K27me3 ChIP-seq tracks from the three conditions profiled showing the selective loss of H3K27me3 at the *Otx2* promoter in *Dgcr8*$^{-/-}$ and 2i+LIF mESCs.

©2014 Macmillan Publishers Limited. All rights reserved

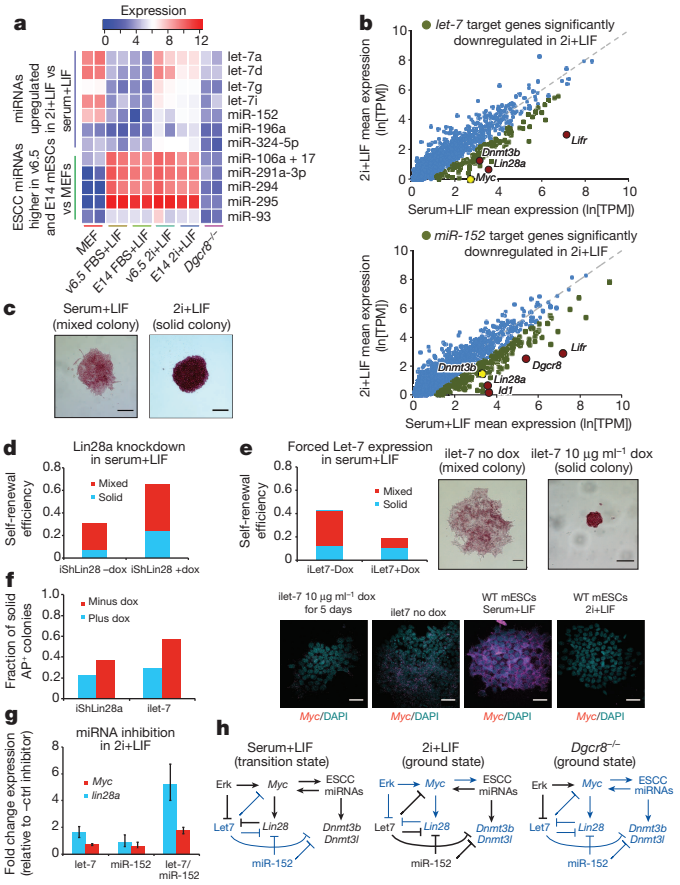


Figure 5 | miRNA balance controls transitions between ground and transition states. **a**, NanoString profiling of miRNAs expressed in mESCs or MEFs. **b**, Expression changes of predicted let-7 or miR-152 target genes between conditions. **c**, Representative colonies showing solid or mixed alkaline phosphatase staining. **d**, **e**, Self-renewal efficiency of mESCs bearing a doxycycline-inducible *Lin28a* shRNA construct (**d**) or doxycycline-inducible let-7 (**e**) in the presence or absence of doxycycline. Mean values from two replicate experiments are shown. *Myc* expression levels for representative ilet-7 colonies as measured by smFISH are shown in **e**. Scale bars: 100 μm (top row); 40 μm (bottom row). **f**, Fraction of uniformly alkaline-phosphatase-positive (AP) colonies in experiments shown in **d** and **e**. **g**, Expression changes in *Myc* and *Lin28a* induced by transfection of miRNA inhibitors into wild-type mESCs cultured in 2i+LIF. Error bars indicate standard deviations between triplicate transfection experiments. **h**, Model for interplay between Erk signalling, miRNAs and *Myc*/*Lin28*/let-7 axis in ground and transition states.

function to repress a set of target genes under ground-state conditions (Supplementary Table 9). Genes targeted by either or both let-7 and miR-152 that were downregulated in 2i+LIF as compared to serum+LIF included *Myc*, *Lin28a*, *Lin28b* and *Dnmt3b* (Fig. 5b). Furthermore, predicted let-7 and miR-152 targets were more highly correlated in expression at a single-cell level in 2i+LIF as compared to non-target genes (Supplementary Information), suggesting that these genes and miRNAs comprise a distinct regulatory module.

Enforced expression of ESCC miRNAs can correct cell cycle defects of *Dgcr8*$^{-/-}$ mESCs, promote rapid proliferation of these cells, suppress inhibitors of the G1 to S transition, and upregulate genes including *Myc* and *Lin28* through presumed indirect mechanisms (refs 36, 44). Transfection of the prototypical ESCC miRNA miR-294 into *Dgcr8*$^{-/-}$ mESCs has been shown to upregulate *Myc*, *Lin28*, *Dnmt3b* and *Dnmt3l* (ref. 44), which are let-7/miR-152 target genes suppressed in the ground state (Extended Data Fig. 9). This opposing effect of ESCC and let-7 miRNAs on a common set of target genes suggests a role for let-7 during differentiation, where it acts to repress pluripotency factors sustained by ESCC miRNAs in stem cells[44]. However, despite the continued high 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*, *Dnmt3b* and *Dnmt3l*.

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 ilet-7 mESCs grown in serum+LIF[45]. *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,

©2014 Macmillan Publishers Limited. All rights reserved

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[46–48]. 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 display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 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 Genet.* **38,** 431–440 (2006).
3. Loh, Y. H. *et al.* Genomic approaches to deconstruct pluripotency. *Annu. Rev. Genomics Hum. Genet.* **12,** 165–185 (2011).
4. MacArthur, B. D., Ma'ayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nature Rev. Mol. Cell Biol.* **10,** 672–681 (2009).
5. Young, R. A. Control of the embryonic stem cell state. *Cell* **144,** 940–954 (2011).
6. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
7. Hayashi, K., Lopes, S. M. C. S., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3,** 391–401 (2008).
8. Hong, S.-H. *et al.* Cell fate potential of human pluripotent stem cells is encoded by histone modifications. *Cell Stem Cell* **9,** 24–36 (2011).
9. Kalmar, T. *et al.* Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7,** e1000149 (2009).
10. Karwacki-Neisius, V. *et al.* Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell* **12,** 531–545 (2013).
11. MacArthur, B. D. *et al.* Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nature Cell Biol.* **14,** 1139–1147 (2012).
12. Reynolds, N. *et al.* NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10,** 583–594 (2012).
13. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
14. Arias, A. M. & Brickman, J. M. Gene expression heterogeneities in embryonic stem cell populations: Origin and function. *Curr. Opin. Cell Biol.* **23,** 1–7 (2011).
15. Cherry, A. & Daley, G. Q. Another horse in the meta-stable state of pluripotency. *Cell Stem Cell* **7,** 641–642 (2010).
16. Graf, T. & Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* **3,** 480–483 (2008).
17. Halley, J. D. *et al.* Self-organizing circuitry and emergent computation in mouse embryonic stem cells. *Stem Cell Res.* **8,** 324–333 (2012).
18. Loh, K. M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8,** 363–369 (2011).
19. MacArthur, B. D. & Lemischka, I. R. Statistical mechanics of pluripotency. *Cell* **154,** 484–489 (2013).
20. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132,** 532–536 (2008).
21. Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145,** 875–889 (2011).
22. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nature Rev. Mol. Cell Biol.* **14,** 357–368 (2013).
23. Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136,** 3853–3862 (2009).
24. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–240 (2013).
25. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510,** 363–369 (2014).
26. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4,** e1000242 (2008).
27. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–313 (2006).
28. Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55,** 319–331 (2014).
29. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).
30. Calabrese, J. M., Seila, A. C., 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. Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26,** 1496–1505 (2008).
32. Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8,** 285–292 (2006).
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 Cells* **11,** 805–814 (2006).
34. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genet.* **39,** 380–385 (2007).
35. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining 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., Kester, 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., Murmann, A. E. & Peter, M. E. The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer* **17,** F19–F36 (2010).
42. Tsuruta, T. *et al.* miR-152 is a tumor suppressor microRNA that is silenced by DNA hypermethylation in endometrial cancer. *Cancer Res.* **71,** 6450–6462 (2011).
43. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403,** 901–906 (2000).
44. Melton, C., Judson, R. L. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463,** 621–626 (2010).
45. Zhu, H. *et al.* The Lin28/let-7 axis regulates glucose metabolism. *Cell* **147,** 81–94 (2011).
46. Balázsi, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144,** 910–925 (2011).
47. Blake, W. J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24,** 853–865 (2006).
48. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422,** 633–637 (2003).

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. G.Q.D. is supported by grants from the NIH (R01GM107536, R24DK092760, P50HG005550) and is an affiliate member of the Broad Institute and an investigator of the Manton 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, an NIH CEGS (1P50HG006193- 01), an NIH Pioneer Award (DP1OD003958-01) and the HHMI. R.M.K. is supported by the Wyss Institute. P.C. is supported by NIDDK (K01DK096013) and received support from NHLBI (T32HL066987 and T32HL007623) and the Manton Center for Orphan Disease Research. R.S. was supported by an NIH Postdoctoral Fellowship (1F32HD075541-01). H.L. is supported by the Mayo Clinic Center for Individualized Medicine. Sequencing was performed by the Broad Institute Genomics Platform. Flow cytometry was performed in the Hematologic Neoplasia Flow Cytometry Facility at the Dana-Farber Cancer Institute and the BCH IDDRC Stem Cell Core Facility at Boston Children's Hospital supported by NIH-P30-HD18655. Single-cell qPCR experiments were performed at the BCH IDDRC Molecular Genetics Core Facility at Children's Hospital Boston supported by NIH-P30-HD18655. Fludigm C_1 experiments were performed at the Broad Institute and the Biopolymers Facility at Harvard Medical School.

Author Contributions R.M.K. designed and performed 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 analyse the ChIP-seq and single-cell RNA-seq data. A.D. performed experiments and helped to analyse the data. H.L. helped to analyse the data. J.Z. generated the iShLin28 mESCs. K.P. generated the iPSCs. T.C.F. assisted with imaging and wrote image analysis algorithms. A.R. oversaw single-cell RNA-seq experiments and helped to write the paper. J.J.C. and G.Q.D. oversaw the project and helped to design the study and write the paper.

Author Information Data are deposited in GEO under accession number GSE60749. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.Q.D. (george.daley@childrens.harvard.edu) or J.J.C. (jcollins@bu.edu).

©2014 Macmillan Publishers Limited. All rights reserved

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, Stasis stem cell-qualified, Gemini), 1× penicillin/streptomycin (Gibco), 1× non-essential amino acids (NEAA) (Gibco), 1× β-mercaptoethanol (β-Me) (EmbryoMax ES cell-qualified, Millipore), and 1,000 U ml^{-1} leukaemia inhibitory factor (LIF) (ESGRO, Millipore). Cells cultured in 2i+LIF were maintained in N2/B27 media, consisting of 50% DMEM/F12 (1:1) and 50% Neurobasal plus 0.5× N-2 supplement, 0.5× 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 PD0325901 (StemGent), 3 μM GSK3 inhibitor CHIR99021 (StemGent), and 1,000 U ml^{-1} LIF (Millipore). For the 1i experiments shown in Extended Data Fig. 9h, either the MEK inhibitor PD0325901 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 (R&D Systems), 0.1 mM 2-mercaptoethanol (Sigma), 2 mM L-glutamine (Gibco), and 1× NEAA (Gibco), and passaged using collagenase type IV (Life Technologies).

MEFs 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 mM valproic acid (Sigma), or 4 μM BIX01294 (Stemgent) for 6 days before analysis, passaging twice in the interim.

Origin of cell lines used in this study. V6.5 mESCs were a gift from R. Jaenisch. E14TG2a mESCs were obtained from ATCC (CRL-1821). *Eed*$^{-/-}$ mESCs were a gift from T. Magnuson. *Dnmt1*$^{-/-}$*Dnmt3a*$^{-/-}$*Dnmt3b*$^{-/-}$ triple knockout mESCs were a gift from M. Okano. *Dgcr8*$^{-/-}$ mESCs were obtained from Novus (NBA1-19349). *Mbd3*$^{-/-}$ mESCs were a gift from B. Hendrich. mEpiSCs were a gift from P. Tesar. *Dicer* knockout mESCs and their matched J1 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[49]. 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 then washed with PBS and transferred into 2 ml of retroviral infection medium[50] containing 25 μl of concentrated Tet-STEMCCA and rtTA lentiviruses and 5 μg ml^{-1} polybrene, and incubated for 24 h. The virus-containing medium was then removed and replaced with 2 ml of fresh ES cell medium containing 1 μg ml^{-1} doxycycline, which was changed daily until the formation of colonies. Colonies were then picked and characterized for pluripotency markers.

Generation of inducible shLin28a ES cell line. The miR-30-based shRNA for Lin28a was designed as described[51], and cloned into a cTGM targeting construct that is integrated at single copy into KH2 ES cells by Flp-driven recombinase-mediated cassette exchange. The KH2 cells contain a recipient locus downstream of *Col1a1* gene, and an rtTA expressed from the *Rosa26* promoter, allowing GFP expression and gene silencing upon addition of 1 μg ml^{-1} doxycycline.

Derivation of neural precursor cells from mESCs. V6.5 mESCs were differentiated to nestin-positive neural precursor cells by established methods[52,53]. Briefly, feeder-free mESCs were allowed to form embryoid bodies for 4 days in suspension culture by plating on a Petri dish in serum+LIF mESC media. At day 4, embryoid bodies were plated on to a tissue culture surface and transferred to DMEM/F12 medium supplemented by insulin, transferrin, selenium and fibronectin (ITSFn) the following day. Cells were maintained in ITSFn for 8 days to select for nestin-positive cells, and were then plated on to laminin and polyornithine-coated plates and shifted to N2 media supplemented with bFGF/laminin to expand nestin-positive neural precursor cells. Media was changed every 2 days, and cells were passaged with trypsin. Expanded neural precursor cells were harvested and processed for single-cell RNA-seq as described below.

Cell capture and whole transcriptome amplification. Single-cell whole transcriptome amplification (WTA) was performed using the Fluidigm C_1 Single-Cell Auto Prep System (C_1 System) as per the manufacturer's recommendations (full details available at http://www.fluidigm.com). Briefly, mES or neural precursor cells were taken directly from culture, trypsinized, washed in PBS, spun down and resuspended at a concentration of 3×10^5 cells ml^{-1} of complete media, mixed 7:3 with C_1 suspension reagent (Fluidigm), and loaded onto C_1 Single-Cell Auto Prep chips (C_1 chips; Fluidigm). After loading, each of the cell isolation chambers on the C_1 chip was optically inspected for the presence of a cell. Subsequently, these cells were lysed and SMART-Seq[54] whole transcriptome amplified (WTA) products were prepared with the C_1 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 C_1 chip, diluted to a concentration of 0.15 ng μl^{-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 s to 60 s. After the PCR step, all 96 samples were pooled without library normalization, cleaned twice with 0.9× AMPure XP SPRI beads (Beckman Coulter), and eluted in buffer TE. The pooled libraries were quantified using a Quant-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 sequencer (25-bp paired-end reads). Single-cell samples were sequenced to an average depth of more than ~1 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 (RNeasy plus Micro RNA kit; Qiagen), according to the manufacturer's recommendations. Subsequently, 1 μl of RNA in water was added to 2 μl of C_1 lysis reaction mix, thermocycled using cycling conditions I (as below). Next, 4 μl of the RT Reaction Mix were 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 above.

For the cell lysis mix we used the following composition and cycling conditions. C1 loading reagent (stock concentration 20×; volume 0.60 μl), SMARTer Kit RNase Inhibitor (40×; 0.30 μl), SMARTer Kit 3′ SMART CDS Primer II A (12 μM; 4.20 μl), SMARTer kit dilution buffer (1×; 6.90 μl); cycling conditions I: (a) 72 °C, 3 min, (b) 4 °C, 10 min, (c) 25 °C, 1 min.

For the reverse transcription (RT) reaction mix we used the following composition and cycling conditions. C_1 loading reagent (stock concentration 20.0×; 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 mM; 2.10 μl), SMARTer kit SMARTer II A oligonucleotide (12 μM; 2.10 μl), SMARTer kit RNase inhibitor (40×; 0.53 μl), SMARTer kit SMARTScribe reverse transcriptase (100.0×; 2.10 μl); cycling conditions II: (a) 42 °C, 90 min, (b) 70 °C, 10 min.

For the PCR mix we used the following composition and cycling conditions. PCR water (35.2 μl), 10× Advantage 2 PCR buffer (stock concentration 10.0×; volume 5.6 μl), 50× dNTP mix (10 mM; 2.2 μl), IS PCR primer (12 μM; 2.2 μl), 50× Advantage 2 Polymerase mix (50.0×; 2.2 μl), C1 loading reagent (20.0×; 2.5 μl); cycling conditions III: (a) 95 °C, 1 min; (b) 5 cycles of (i) 95 °C, 20 s, (ii) 58 °C, 4 min, (iii) 68 °C, 6 min; (c) 9 cycles of (i) 95 °C, 20 s, (ii) 64 °C, 30 s, (iii) 68 °C, 6 min; (d) 7 cycles of (i) 95 °C, 30 s, (ii) 64 °C, 30 s, (iii) 68 °C, 7 min; (e) 72 °C, 10 min.

Calculation of expression levels and statistics. Mapping of reads and calculation of expression values was performed as described[24]. Briefly, a Bowtie[55] index was created based on the UCSC knownGene[56] transcriptome, and paired-end reads were directly aligned to this index. RSEM[57] was run on these alignments to generate expression level estimates, and these estimates (τ) 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 ln[TPM]. Unless otherwise specified, all RNA-seq 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 (σ) 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 Sailfish to estimate the abundances of non-coding RNAs in the single-cell RNA-seq libraries[58]. Sailfish was executed using a transcript index file (generated as described below) and fastq files of the library reads to generate transcript level

©2014 Macmillan Publishers Limited. All rights reserved

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 lincRNA 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 5% 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. Those studies were performed in v6.5 mESCs grown under standard conditions in the presence of serum+LIF.

Identification of gene sets enriched for variable genes. We took 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 (COVs) 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 gene sets 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*, *Fgf4*, *Rex1*, *Sall4*, *Dnmt1*, *Tcfcp2l1*, *Zfx*, *Lin28a*, *Dnmt3l*, *Dnmt3b*, *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 Student's 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 $|\log_2(\text{ratios})| > 1$. 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, QuantiGene 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 488 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 package[59]. 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 CellProfiler[60], 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 low background of confocal microscopy (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 calcein 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 described[61].

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 overlapped with 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 (STA). Amplification products were diluted 1:5 with TE(8.0), and samples were

©2014 Macmillan Publishers Limited. All rights reserved

checked for the presence of amplified cDNA by duplex microlitre-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 standard deviation of the average 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 G_0/G_1 phase of the cell cycle were isolated for analysis.

To determine expression distributions changed 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. 3f. To identify distinct transcriptional states of the qPCR data, we modelled the first three principal components as multivariate normal mixtures, allowing the model selection of 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:

$$P(x \in \text{cluster A}) = \frac{P_A f_A(x)}{\sum_{i=1}^{C} P_i f_i(x)}$$

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[62]. 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. For the data shown in Fig. 4e and 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[62]. 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 $|\log_2(\text{fold change})| > 2$.

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 mRNA 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 (snoRNA 142, snoRNA 202, snoRNA 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 TargetScan[63] and miRmap[64]. For TargetScan, we considered as predicted targets of these miRNA families the set of all predicted target genes in mouse, irrespective of site 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 stained 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.

©2014 Macmillan Publishers Limited. All rights reserved

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, iShLin28a 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, ilet-7 cells[45] were treated with 10 μg ml^{-1} doxycycline or left untreated and cultured for 5 days, passaging 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. 5g, mESCs grown in 2i+LIF media were transfected with miRNA inhibitors (hsa-let-7 miRCURY LNA microRNA Power family inhibitor, Product 460006; mmu-miR-152-3p miRCURY LNA Power microRNA inhibitor, Product 4101871-101; or miRCURY LNA microRNA Power Inhibitor Negative Control A, 199020-04, all from Exiqon) at a concentration of 20 nM each using Dharmafect 1 (Dharmacon) one day after plating at a concentration of 2,000 cells per well of a 24-well plate. Media was changed after 24 h, and cells were harvested 3 days after transfection and RNA was extracted using a Qiagen miRNeasy Micro kit for qPCR analysis. miRNA expression experiments shown in Extended Data Fig. 9 were performed essentially as described[44]. *Dgcr8*$^{-/-}$ mESCs maintained off of feeders were plated 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 24 h, 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 (Ln[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:

$$z-\text{score}\left(\text{sample}_s,\ \text{gene}_g,\ \text{condition}_c\right)=\frac{\text{expression}_{s,g,c}-\mu(\text{expression})_{g,c}}{\sigma(\text{expression})_{g,c}}.$$

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[65] 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 (*Esrrb*), and 551 cells from 5 colonies (*Neurod1*). For Fig. 1c, COV enrichment of biological pathway annotations. COV-based enrichment analysis was performed by testing whether genes in each Gene Ontology Biological Process[66] and Reactome[67] annotation tended to have higher or lower COVs in serum+LIF cells than would be expected by chance given the distribution of μ in the gene set.

For Fig. 2b, cell colony numbers: 258 cells for the colony on the left in the high expression state, 70 cells for the colony in the middle in the low expression state, and 98 cells for the colony on the right in the mixed expression state. These three colonies are part of the 10 colonies examined for the overall *Esrrb* population distribution in serum+LIF shown in Figs 1b and 2c. Fig. 2c, *Oct4* and *Dazl*: 5 colonies, 318 cells; *Sox2* and *Neurod1*: 5 colonies, 551 cells; *Nanog* and *Otx2*: 5 colonies, 151 cells; *Esrrb*: 10 colonies, 1,226 cells; *Lin28a*: 5 colonies, 472 cells; *Nr0b1*: 8 colonies, 849 cells; *Myc*, 8 colonies, 867 cells; *Olig2*, 10 colonies, 1,097 cells; *Nkx2-9*: 5 colonies, 524 cells; and *Pax6*: 5 colonies, 417 cells. *Oct4* and *Dazl*, *Sox2* and *Neurod1*, and *Nanog* and *Otx2* were each imaged by two-colour RNA FISH in the same set of cells. Average expression values and standard deviations for these data are shown in Extended Data Fig. 3a. *Esrrb* and *Neurod1* individual colony data comprise the population distributions shown in Fig. 1b.

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. Fig. 4e, serum+LIF: 901 cells; 2i+LIF: 1,648 cells; *Dgcr8*$^{-/-}$: 248 cells.

For Fig. 5a, miRNAs with $|\log2(\text{fold change})| > 2$ (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, NR0B1 and OCT4 in serum+LIF: 2,917 cells.

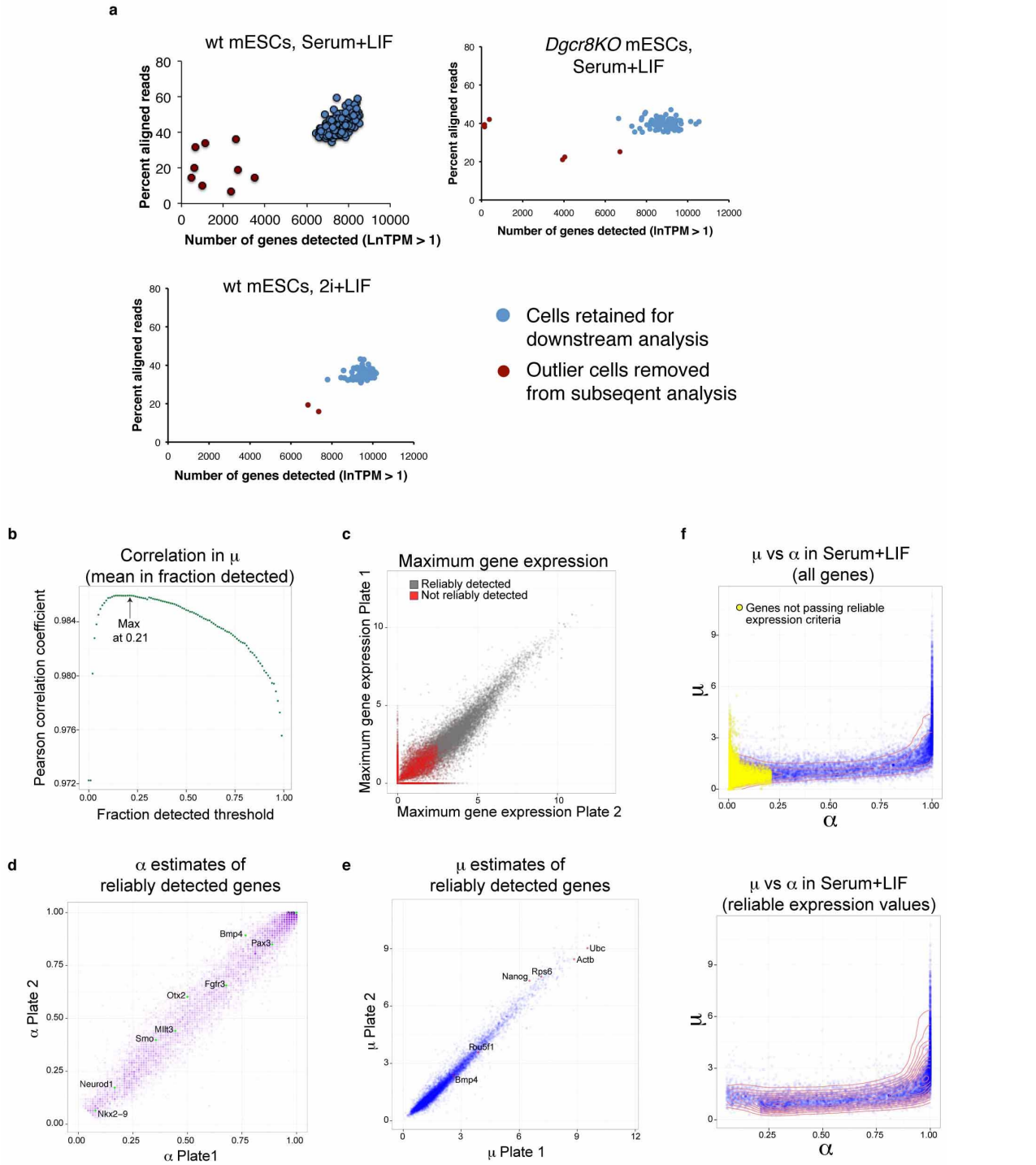
For Extended Data Fig. 7e, *Oct4*: serum+LIF (5 colonies, 318 cells (same data as in Figs 1b and 2c)), 2i+LIF (5 colonies, 485 cells), *Dgcr8*$^{-/-}$ (5 colonies, 175 cells). *Nanog*: serum+LIF (5 colonies, 151 cells (same data as in Fig. 2c)), 2i+LIF (5 colonies, 398 cells), *Dgcr8*$^{-/-}$ (5 colonies, 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). *Otx2*: 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 *Otx2* 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 *Otx2* for 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; *Esrrb*: 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. NR0B1 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 Protocols* **3,** 1180–1186 (2008).
51. Dow, L. E. *et al.* A pipeline for the generation of shRNA transgenic mice. *Nature Protocols* **7,** 374–393 (2012).
52. Lee, S. H., Lumelsky, N., Studer, L., Auerbach, J. M. & McKay, R. D. Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. *Nature Biotechnol.* **18,** 675–679 (2000).
53. Okabe, S., Forsberg-Nilsson, K., Spiro, A. C., Segal, M. & McKay, R. D. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech. Dev.* **59,** 89–102 (1996).
54. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnol.* **30,** 777–782 (2012).
55. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
56. Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39,** D876–D882 (2010).
57. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
58. Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnol.* **32,** 462–464 (2014).
59. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5,** 877–879 (2008).
60. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).
61. Garber, M. *et al.* A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* **47,** 810–822 (2012).
62. Spector, D. L. & Smith, H. C. Redistribution of U-snRNPs during mitosis. *Exp. Cell Res.* **163,** 87–94 (1986).
63. 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).

©2014 Macmillan Publishers Limited. All rights reserved

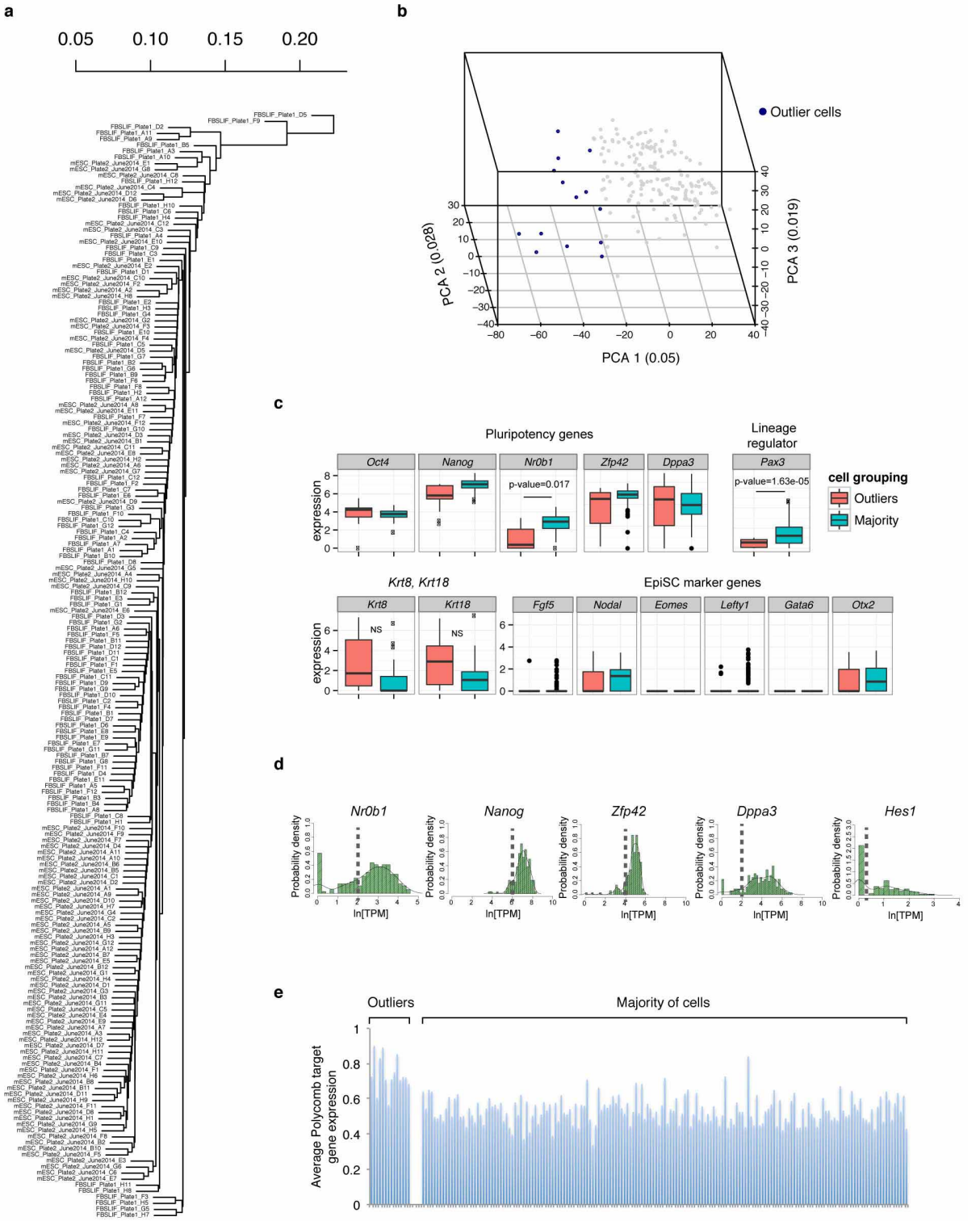
64. Vejnar, C. E. & Zdobnov, E. M. MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.* **40,** 11673–11683 (2012).
65. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6,** 65–70 (1979).
66. Ashburner, M. *et al.* Gene ontology: 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.* **458,** 611–631 (2002).
69. Cahan, P. *et al.* CellNet: network biology applied to stem cell engineering. *Cell* **158,** 903–915 (2014).

©2014 **Macmillan Publishers Limited. All rights reserved**



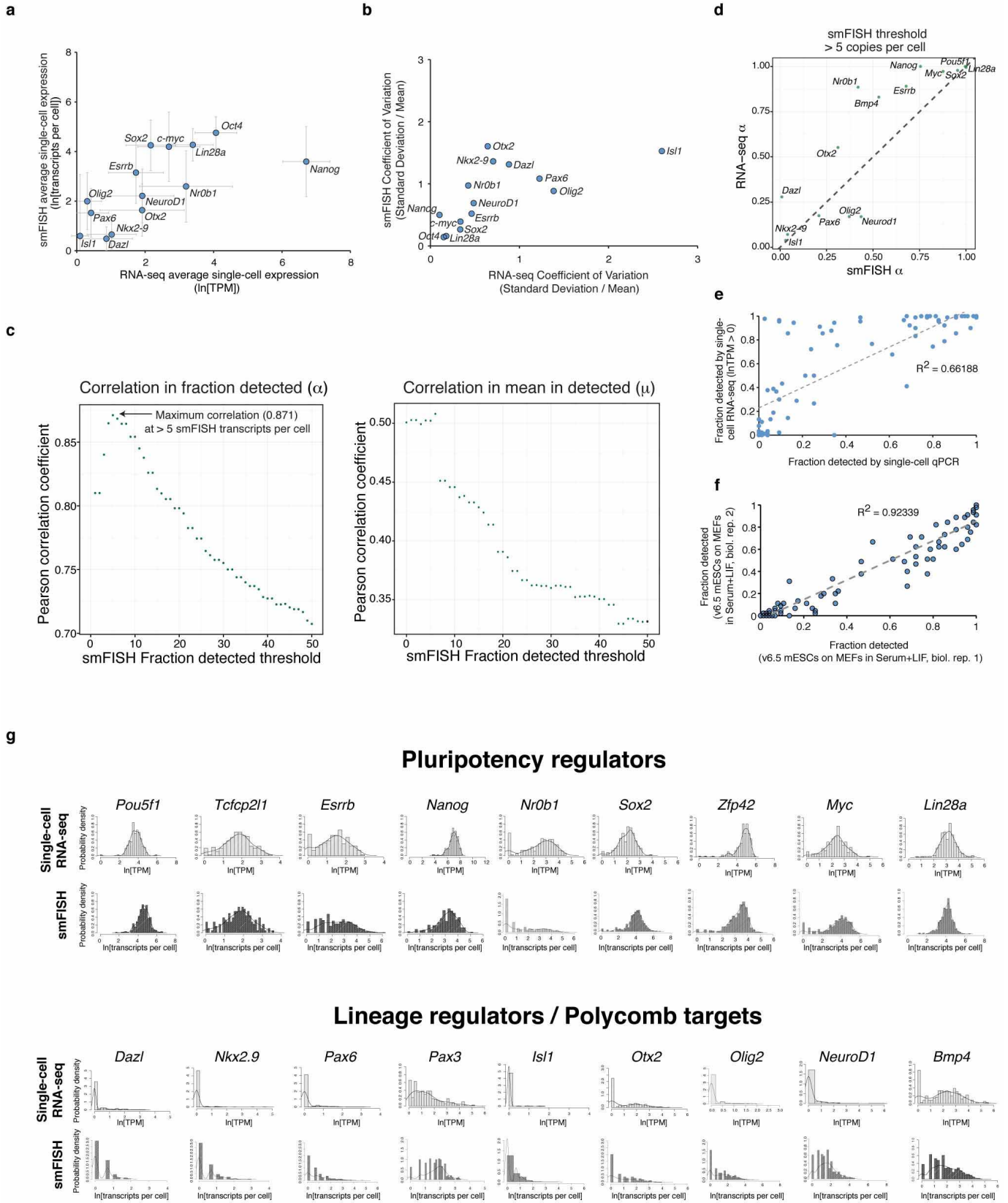
Extended Data Figure 1 | Quality control of single-cell RNA-seq data. **a**, A combination of read alignment rate (*y* axis) and number of genes detected (ln[TPM] >1) (*x axis*) was used to identify outlier cells (red circles) to remove from subsequent analysis, leaving 183 single mESCs cultured in serum+LIF, 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 α (fraction of cells a gene was detected in) thresholds. μ was calculated separately for each plate. Correlations in μ were calculated after limiting genes to those with α 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 α 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 α (*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).

©2014 Macmillan Publishers Limited. All rights reserved



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 *Pax3* (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).

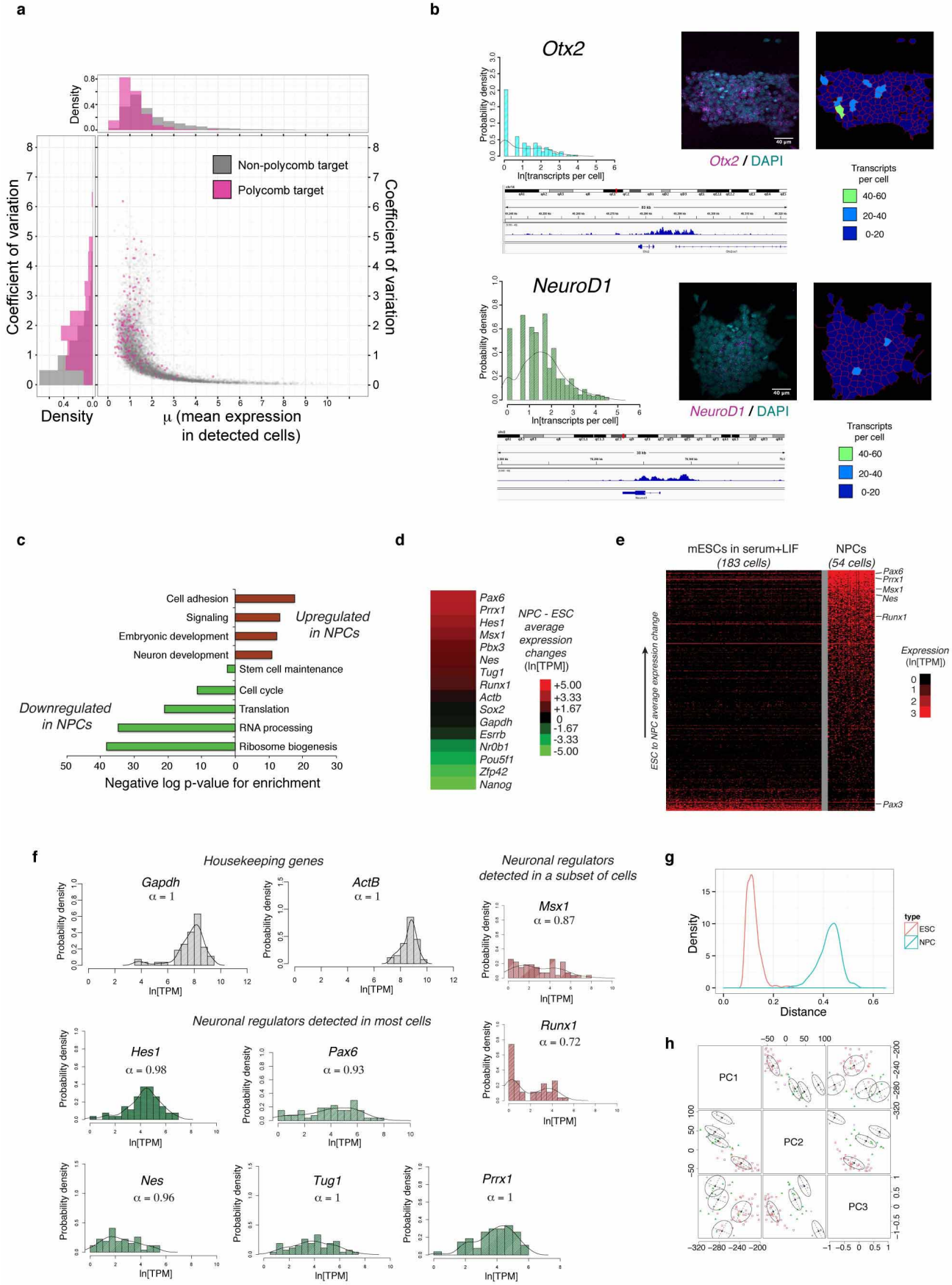
©2014 Macmillan Publishers Limited. All rights reserved



©2014 Macmillan Publishers Limited. All rights reserved

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 α (fraction detected, left) or μ (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 $<$10 copies per cell would not be detected by RNA FISH. Correlation for α 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 α 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).

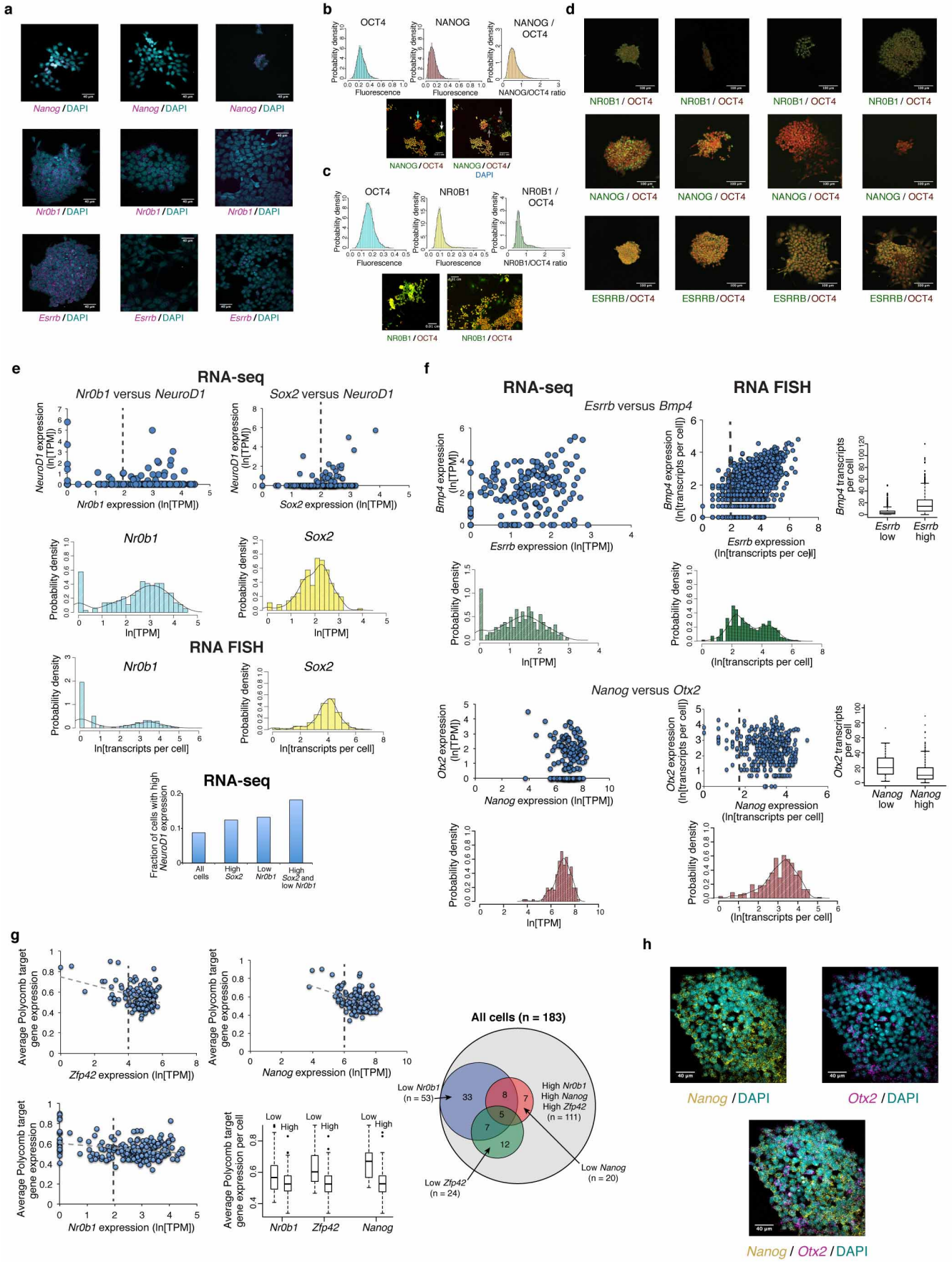
©2014 Macmillan Publishers Limited. All rights reserved



©2014 Macmillan Publishers Limited. All rights reserved

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 μ (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 μ 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.

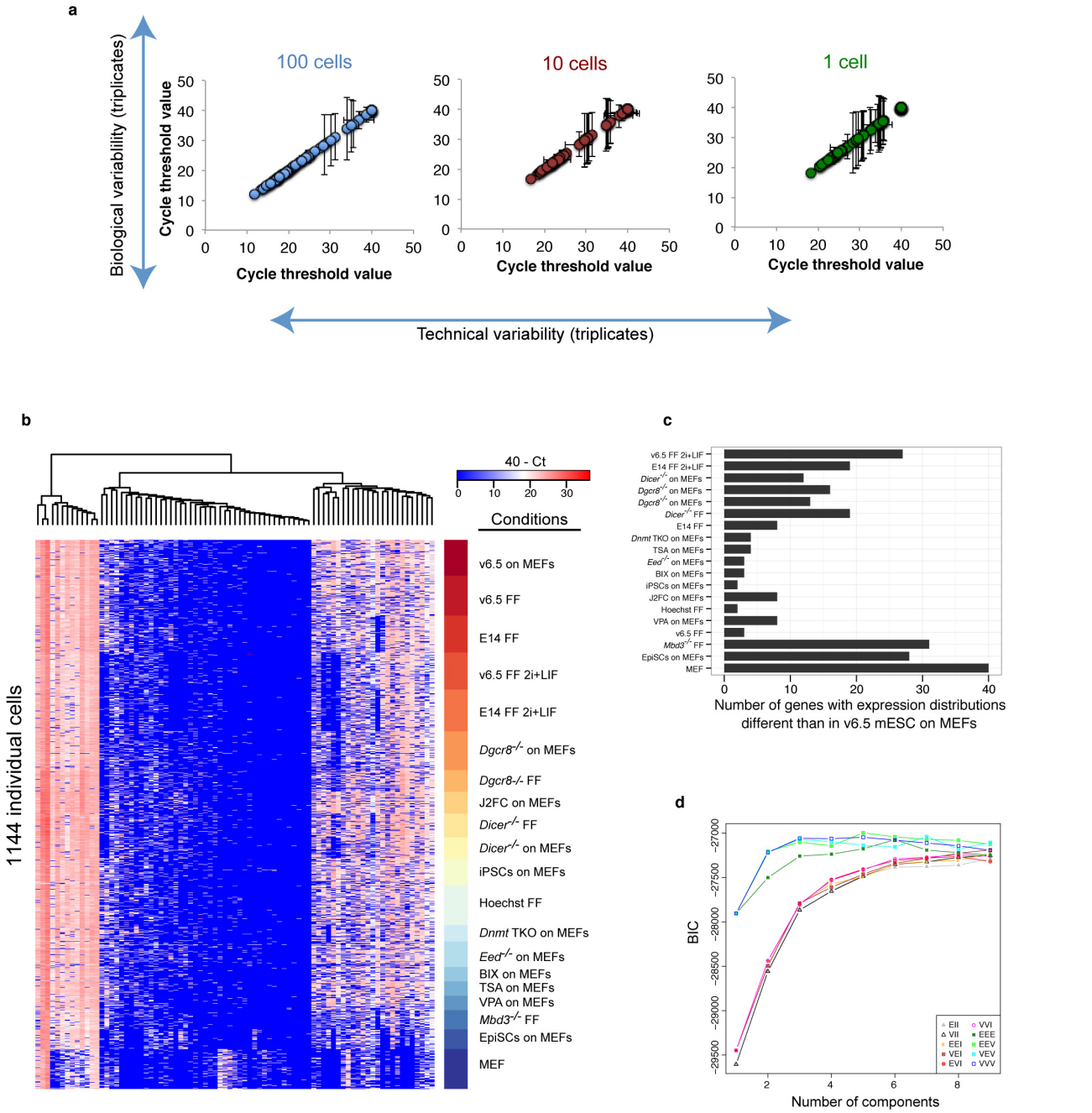
©2014 Macmillan Publishers Limited. All rights reserved



©2014 Macmillan Publishers Limited. All rights reserved

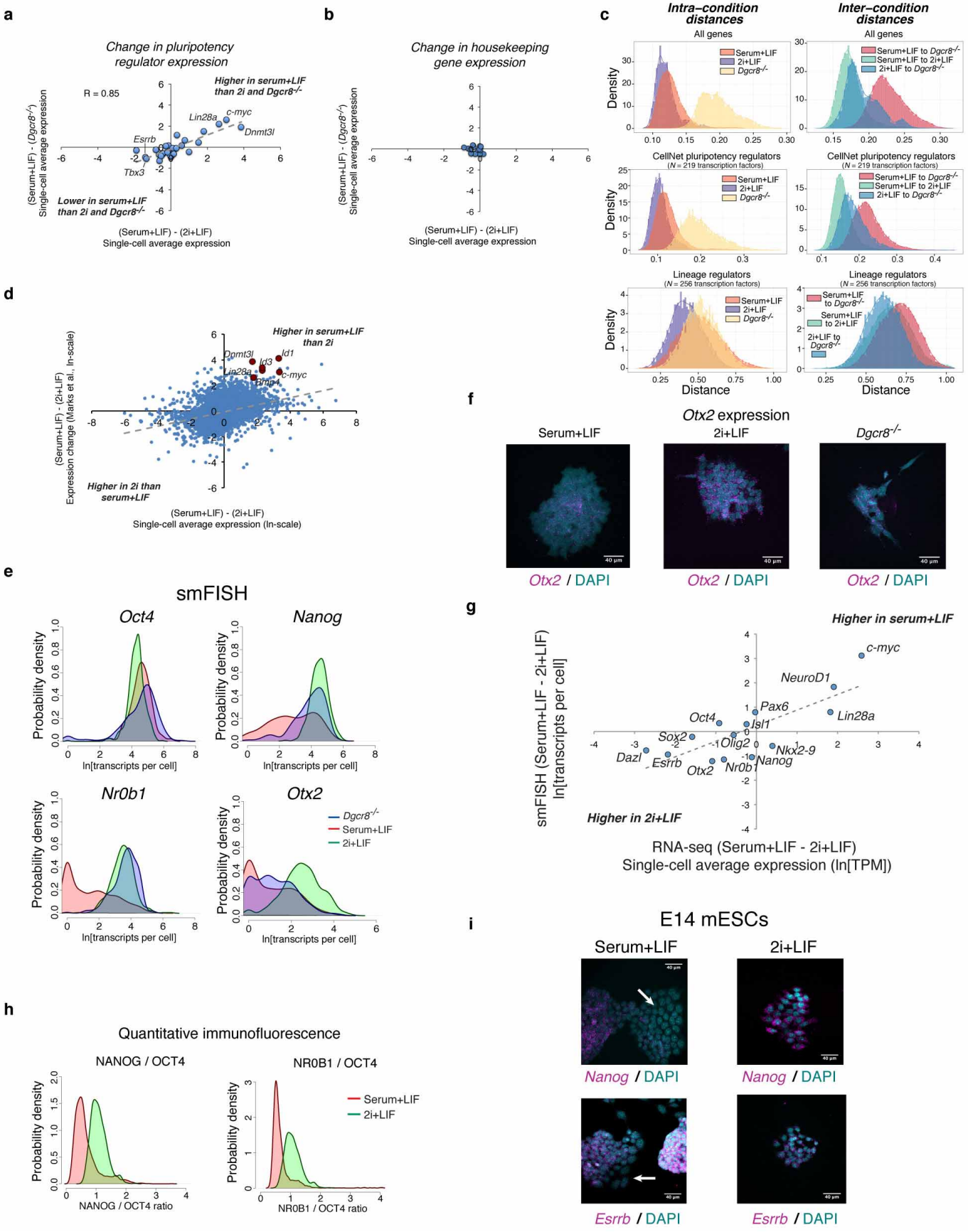
Extended Data Figure 5 | Fluctuations in pluripotency and lineage regulator expression. **a**, RNA FISH images showing expression of *Nanog* (top), *Nr0b1* (middle) and *Esrrb* (bottom) in individual colonies or regions of cells. **b**, **c**, Histograms show distributions of fluorescence intensities 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.

©2014 Macmillan Publishers Limited. All rights reserved



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[68]. The best model was used to classify cells into states.

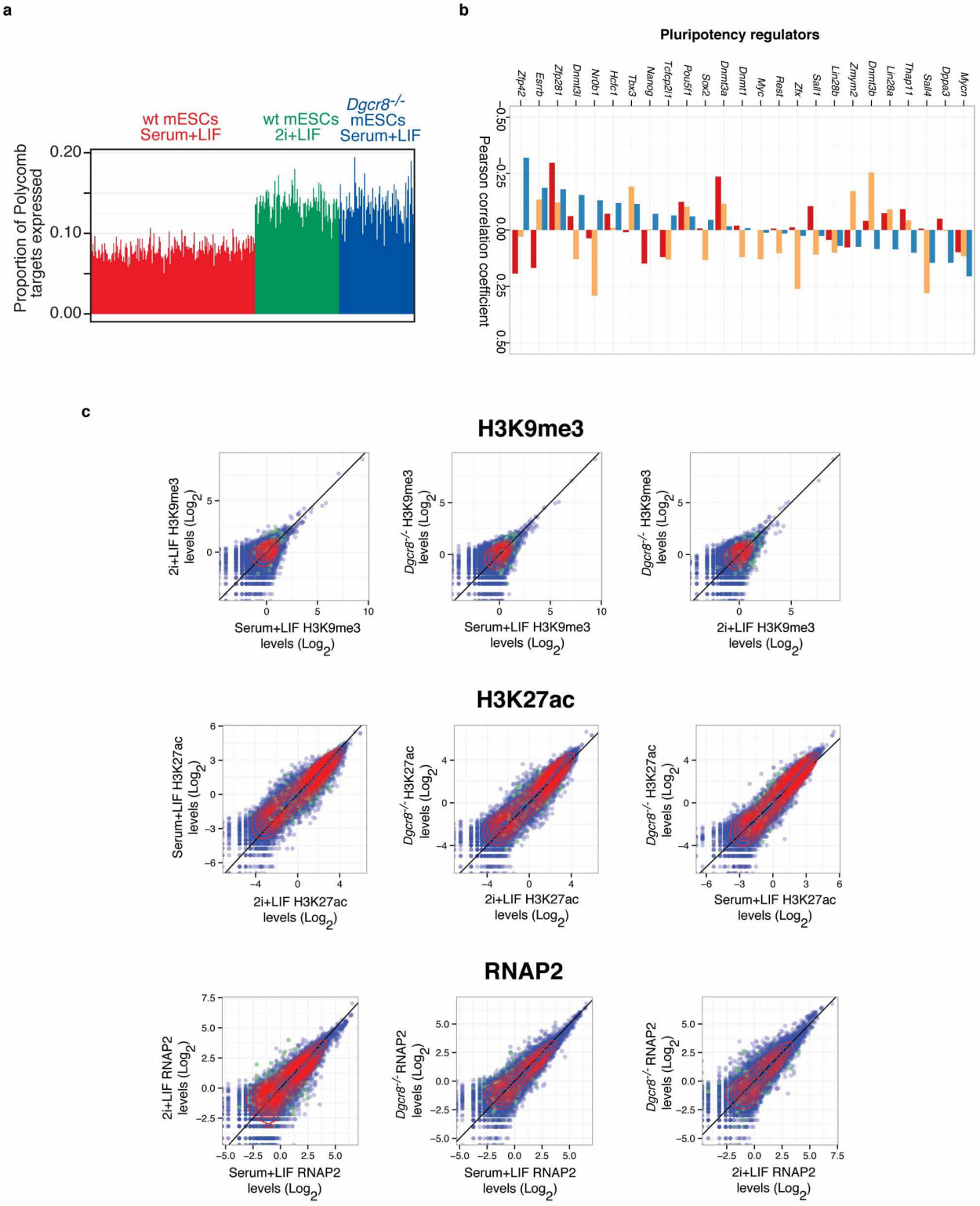
©2014 Macmillan Publishers Limited. All rights reserved



©2014 Macmillan Publishers Limited. All rights reserved

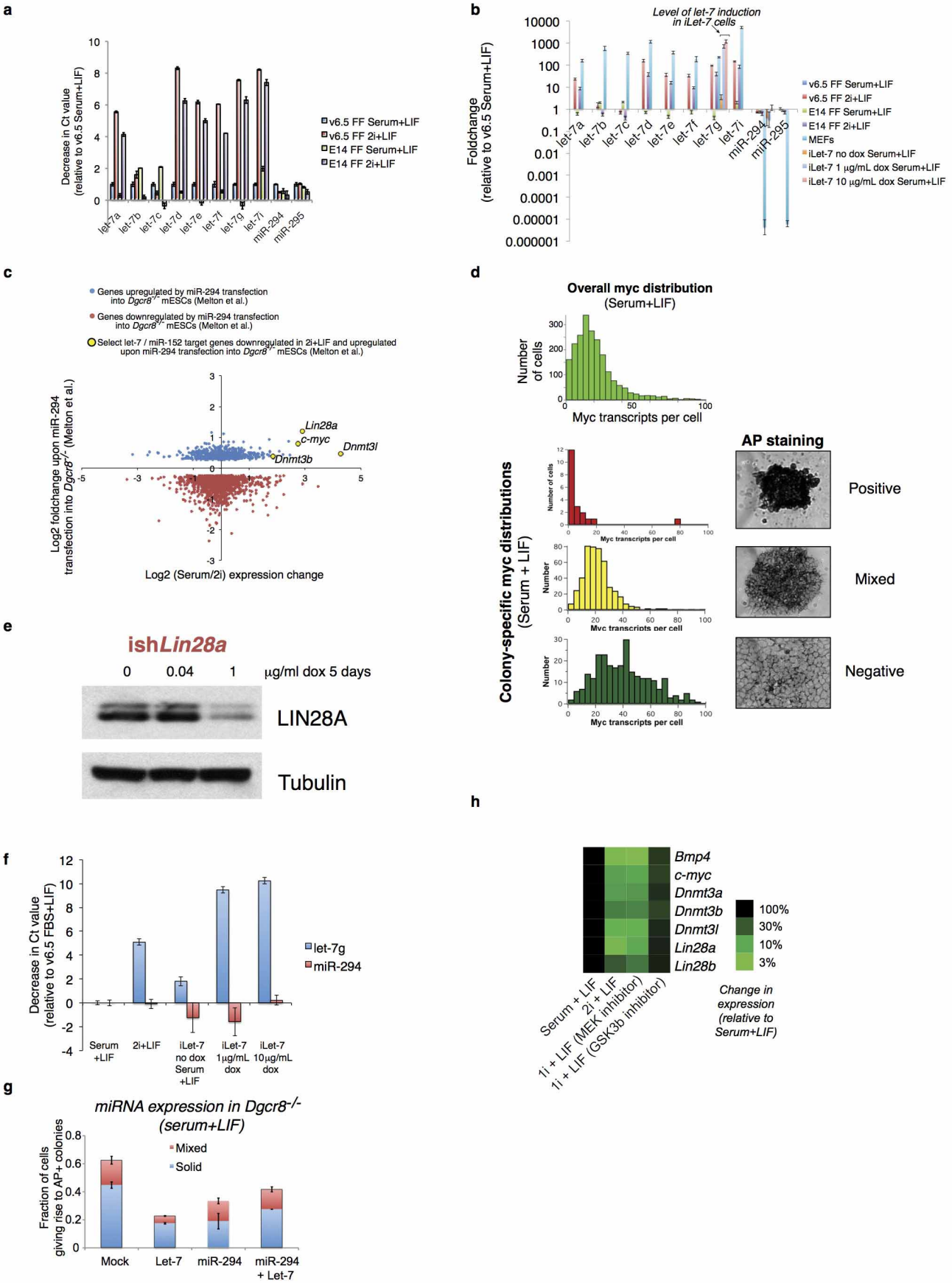
Extended Data Figure 7 | Gene expression changes in mESCs upon culture in 2i or removal of mature miRNAs. **a**, Changes in expression of the pluripotency regulators shown in Fig. 4c going from serum+LIF to 2i+LIF culture (*x* axis), or between wild-type and *Dgcr8*$^{-/-}$ cells cultured in serum+LIF (*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*, *Ipo8*, *Mrpl48*, *Mtcp1*, *Pgk1*, *Ppia*, *Rpl13a*, *Rplp2*, *Rps6*, *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 CellNet[69] (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.

©2014 Macmillan Publishers Limited. All rights reserved



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).

©2014 Macmillan Publishers Limited. All rights reserved



©2014 Macmillan Publishers Limited. All rights reserved

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.

©2014 Macmillan Publishers Limited. All rights reserved

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*$^{-/-}$	Serum+LIF	Yes	*Eed*$^{-/-}$	31
DNA methyltransferases	Loss of DNA methylation	*Dnmt* TKO	Serum+LIF	Yes	*Dnmt* TKO	33
MicroRNA processing	Loss of mature miRNAs (Knockout of *Dgcr8*)	*Dgrc8*$^{-/-}$	Serum+LIF	Yes	*Dgrc8*$^{-/-}$ on MEFs	34
				No	*Dgrc8*$^{-/-}$ FF	
	Loss of mature miRNAs (Knockout of *Dicer*)	*Dicer*$^{-/-}$	Serum+LIF	Yes	*Dicer*$^{-/-}$ on MEFs	30
				No	*Dicer*$^{-/-}$ FF	
NuRD complex	Disruption of the NuRD complex	*Mbd3*$^{-/-}$	Serum+LIF	Yes	*Mbd3*$^{-/-}$	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_0/G_1 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