nature cell biology

Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity

Ben D. MacArthur[1,2,3,10], Ana Sevilla[4,5,10], Michel Lenz[6], Franz-Josef Müller[7], Berhard M. Schuldt[6], Andreas A. Schuppert[6], Sonya J. Ridden[2,8], Patrick S. Stumpf[1], Miguel Fidalgo[4,5], Avi Ma'ayan[9], Jianlong Wang[4,5] and Ihor R. Lemischka[4,5,9,11]

A number of key regulators of mouse embryonic stem (ES) cell identity, including the transcription factor Nanog, show strong expression fluctuations at the single-cell level. The molecular basis for these fluctuations is unknown. Here we used a genetic complementation strategy to investigate expression changes during transient periods of Nanog downregulation. Employing an integrated approach that includes high-throughput single-cell transcriptional profiling and mathematical modelling, we found that early molecular changes subsequent to Nanog loss are stochastic and reversible. However, analysis also revealed that Nanog loss severely compromises the self-sustaining feedback structure of the ES cell regulatory network. Consequently, these nascent changes soon become consolidated to committed fate decisions in the prolonged absence of Nanog. Consistent with this, we found that exogenous regulation of Nanog-dependent feedback control mechanisms produced a more homogeneous ES cell population. Taken together our results indicate that Nanog-dependent feedback loops have a role in controlling both ES cell fate decisions and population variability.

Several important regulators of ES cell identity, including the homeodomain transcription factor Nanog[1–3], show significant temporal expression fluctuations at the single-cell level[4–15]. Such fluctuations give rise to robust functional heterogeneity within ES cell populations, profoundly affecting their long-term regenerative potency[9,16,17]. In the case of Nanog, apparently stochastic transitions between Nanog-high and Nanog-low states occur within individual Oct4-positive ES cells[13]. These fluctuations transiently prime individual ES cells for differentiation without marking definitive commitment[4]. Thus, Nanog seems to act as a molecular gatekeeper: suppressing adverse spontaneous differentiation events in fluctuating environments while ensuring robust differentiation in the presence of appropriate and persistent stimuli. However, the molecular basis for this mechanism remains unclear.

To investigate this issue we developed a time-course strategy designed to controllably reproduce the Nanog expression level fluctuations observed in wild-type ES cells[7,17]. To accurately regulate Nanog levels we used the doxycycline (dox)-dependent inducible system previously described[18,19] (Fig. 1a). In this system a short hairpin RNA (shRNA) depletes endogenous *Nanog* messenger RNA, and normal levels of Nanog are restored by a dox-inducible shRNA immune mRNA[18,19]. In the presence of dox, this engineered rescue mouse ES cell line (NanogR) expresses Nanog homogeneously (Fig. 1b) and is fully pluripotent both *in vitro* and *in vivo*[18,19]. On removal of dox, *Nanog* mRNA and protein levels sharply decline and pluripotency and self-renewal capacities are progressively lost[18,19]. Cell samples were collected at day 0 (dox present, Nanog-expressing) and at days 1, 3 and 5 days after dox withdrawal (Fig. 1c). Additionally, at each time point a set of samples was further treated with a 12 h pulse of dox before being collected and compared with untreated control samples collected at the same time. Thus, cells were exposed to transient periods (24, 72 and 120 h) of Nanog removal. In essence, this strategy mimics the reported temporal fluctuations of endogenous Nanog expression levels[4,13]. Gene expression microarrays were performed in triplicate at each time point and for each culture condition to determine the effects of Nanog fluctuations on global mRNA levels (Fig. 2).

[1]Centre for Human Development, Stem Cells and Regeneration, Institute of Developmental Sciences, University of Southampton, Southampton SO17 1BJ, UK. [2]School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK. [3]Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK. [4]Department of Developmental and Regenerative Biology, Mount Sinai School of Medicine, New York, New York 10029, USA. [5]Black Family Stem Cell Institute, Mount Sinai School of Medicine, New York, New York 10029, USA. [6]Aachen Institute for Advanced Study in Computational Engineering Science, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen 52062, Germany. [7]Zentrum für Integrative Psychiatrie, Kiel 24105, Germany. [8]Institute for Complex Systems Simulation, University of Southampton, SO17 1BJ, UK. [9]Department of Pharmacology and Systems Therapeutics, Mount Sinai School of Medicine, New York, New York 10029, USA. [10]These authors contributed equally to this work.
[11]Correspondence should be addressed to I.R.L. (e-mail: ihor.lemischka@mssm.edu)

Received 9 June 2012; accepted 17 September 2012; published online 28 October 2012; DOI: 10.1038/ncb2603

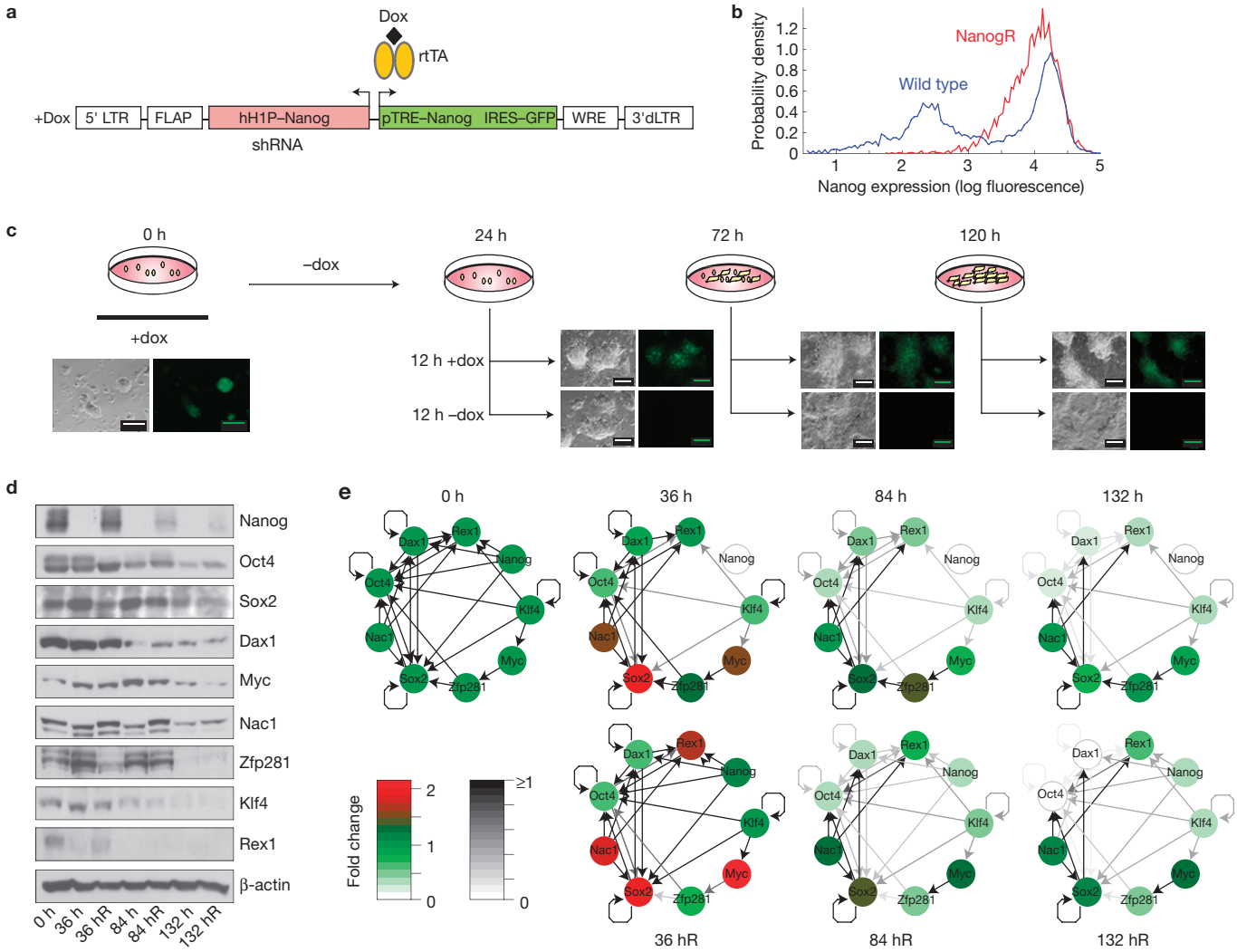

Figure 1 Quantifying the molecular effects of Nanog fluctuations. (**a**) The lentiviral vector construct to conditionally regulate Nanog expression levels[19]: dLTR, deleted long-terminal repeat; FLAP, sequence element that improves transduction efficiency; rtTA, a TetOn tetracycline (doxycycline)-controlled transcriptional activator; WRE, woodchuck hepatitis virus post-transcriptional regulatory element. (**b**) Flow-cytometric comparison of the distribution of Nanog expression levels in wild-type Nanog GFP (ref. 47) and NanogR (lentiviral-vector-regulated Nanog expression; **a**) (ref. 19) ES cells. In both cases, GFP levels reflect Nanog levels. (**c**) Experimental design. Scale bars, 100 μm. (**d**) Effect of Nanog downregulation and rescue on protein expression levels in the ES cell transcriptional regulatory network (TRN) as measured by western blot. Full scans are given in Supplementary Fig. S1. (**e**) Decomposition of the extended ES cell TRN after Nanog depletion. Colours and greyscale denote relative expression levels measured by quantitative PCR. The bottom row indicates expression levels subsequent to treatment with a 12 h pulse of dox at 24 h (denoted 36 hR), 72 h (84 hR) and 120 h (132 hR).

RESULTS

Identifying a critical point-of-no-return in the ES cell fate switch

Expression of pluripotency-associated transcripts was progressively downregulated on Nanog removal (Figs 1d–e and 2b). To provide context to these changes we considered them in light of two previously published regulatory networks for ES cell pluripotency: a transcriptional regulatory network (TRN; see Fig. 1d; as detailed in ref. 20) and an extended ES cell regulatory network (as detailed in ref. 21 and updated in Supplementary Table S1). Although Nanog was robustly downregulated within 24 h of dox removal (without dox Nanog is almost undetectable after 1 day, see Figs 1d–e and 2b), most elements of both the TRN and the extended network did not show significant changes in expression until at least 3 days after Nanog depletion (Figs 1d–e and Fig. 2b and Supplementary Fig. S2). This indicates that loss of pluripotency occurs on a timescale significantly longer than that of Nanog loss. Indeed, consistent with previous observations[4,22], full decomposition of the ES cell TRN was observed only after 5 days (Fig. 1d–e), indicating that this network remains essentially active in the temporary absence of Nanog[4]. Once significant expression changes had occurred (day 3 onwards), reintroduction of Nanog did not have a significant rescue effect on most pluripotency markers (Figs 1d–e and 2b), suggesting that a critical point had been passed and that permanent changes in the TRN had occurred. To investigate this further we constructed a simple mathematical model of Nanog regulation of pluripotency. Analysis of this model suggests that the observed dynamics are due to a bistable switch in which Nanog plays a central role by positively reinforcing the pluripotent ground[3,23] state (see Supplementary Note S1 for full details).

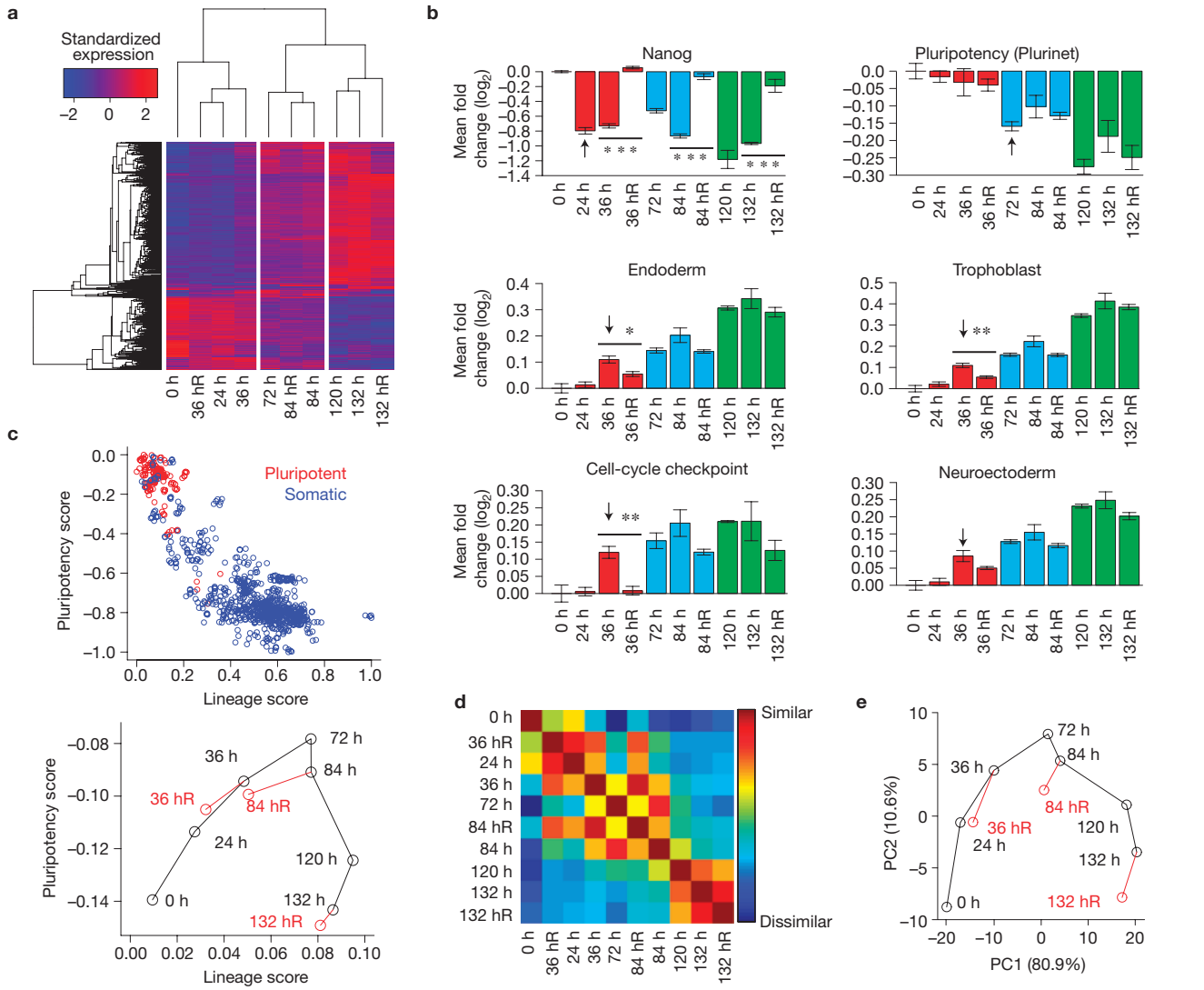

Figure 2 Transcriptome changes during periods of transient Nanog depletion. (**a**) Heat map of significant gene expression changes. (**b**) Mean fold expression changes for pluripotency- (elements of the extended ES cell regulatory network as detailed in ref. 21 and updated in Supplementary Table S1), cell-cycle- and lineage-associated gene sets. The asterisks indicate significance by 2-sample *t*-test with *P* values $^{*}<0.05$, $^{**}<0.01$, $^{***}<0.0001$; the arrows indicate the earliest time at which a significant expression change was observed ($P<0.05$). The error bars show ± one standard error, $n=3$. (**c**) Machine-learning classification of genome-wide expression patterns of 1,032 pluripotent and somatic cell samples (top) and Nanog downregulation time course samples (bottom). The large ranges in the pluripotency scores and lineage scores in the top panel reflect the wide variety of cell types used to construct the classifier. The Nanog depletion time series represent the earliest stages of ES cell differentiation, and therefore naturally show high pluripotency scores and low lineage scores. Nevertheless, a movement away from the pluripotent state is clearly detected using these classifiers; thus, highlighting their sensitivity and range. (**d**) Similarity matrix of samples in classification space. (**e**) PCA of the Nanog depletion time-course data (first two components are shown).

Lineage-associated gene expression changes are reversible

Markers of the earliest mammalian lineages (trophoblast, primitive endoderm and primitive ectoderm/neural ectoderm; as detailed in ref. 24 and given in Supplementary Table S1) as well as cell-cycle checkpoint-associated genes (Supplementary Table S1) showed significant upregulation within 36 h of Nanog removal, indicating that Nanog is a potent negative regulator of early lineage decisions[4,19] and cell-cycle checkpoint controls[25]. Furthermore, both lineage- and pluripotency-associated markers were significantly enriched within the set of genes that exhibited significant expression changes on Nanog removal (Supplementary Table S2). However, in contrast to expression changes of pluripotency-associated genes, changes in early lineage-associated genes were rapidly reversible on reintroduction of Nanog (Fig. 2b), indicating a gradual and revocable accumulation of lineage characteristics. A similar pattern of reversible expression changes was also observed in germ-cell-associated genes, in accordance with the central role that Nanog plays in primordial germ cell identity[4] (Supplementary Fig. S2).

A bioinformatic classifier for pluripotency of mouse cells

To gain a better understanding of these early fate changes we developed a bioinformatic assay for pluripotency of mouse cells[26]. We first downloaded and manually curated a training set of 1,032 mouse microarray data sets from the Gene Expression Omnibus (GEO)

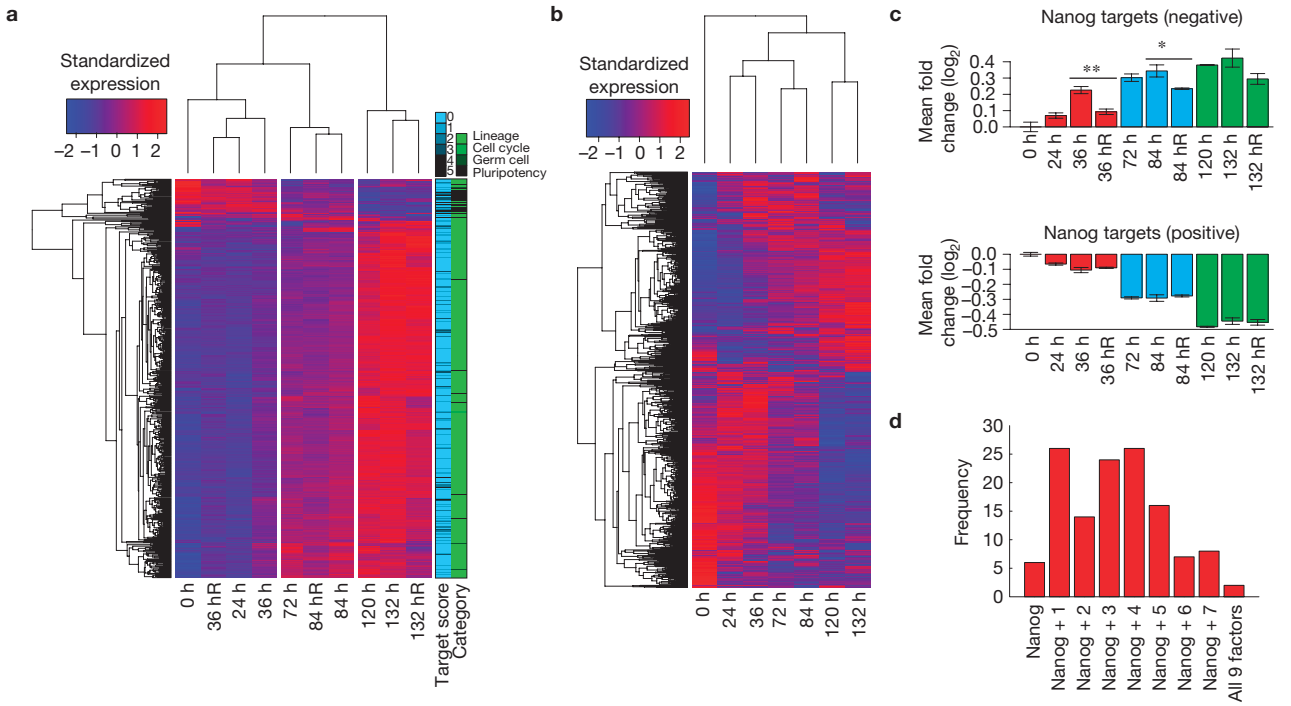

Figure 3 Expression changes and promoter occupancy by Nanog. (**a**) Hierarchical clustering of expression changes of pluripotency-, lineage- and cell-cycle-associated genes (see Supplementary Table S1). The blue bar shows the number of times each gene has been reported as a target of Nanog in six recent papers that examined promoter occupancy[20,48–52]. The green bar shows the category to which the genes belong. Pluripotency-associated genes are frequently high-confidence targets of Nanog. (**b**) Expression patterns for high-confidence direct targets of Nanog. Genes were selected as high-confidence Nanog targets if they were identified in at least three of six recent papers that examined Nanog target gene promoter occupancy[20,48–52]. (**c**) Mean fold changes for high-confidence direct targets of Nanog. Expression patterns are not uniform, so mean fold changes are shown separately for those genes that were upregulated and downregulated during the time course. The error bars show ± one standard error, $n = 3$. The asterisks indicate significance by 2-sample t-test with P values $^{*} < 0.05$, $^{**} < 0.01$. (**d**) Total number of Nanog target genes that changed significantly after depletion of Nanog grouped by the number of other ES cell TRN members that also regulate target gene expression. Most commonly, Nanog target gene expression is regulated by Nanog in concert with between one and five other transcription factors.

database (http://www.ncbi.nlm.nih.gov/geo/), including expression profiles of pluripotent cells (142 samples) and a variety of somatic cell samples (790 samples; Supplementary Table S3). We then developed two machine-learning classifiers (full details and code are given in Supplementary Note S2 and Supplementary Software) that, when taken together, were able to accurately distinguish pluripotent from non-pluripotent samples in our training data set (Fig. 2c, left panel). The first classifier, termed the pluripotency score, identifies patterns of gene expression specifically associated with pluripotency. The second, termed the lineage score, determines whether new expression patterns not observed in the pluripotent training samples are present. Thus, pluripotent cells have a high pluripotency score and a low lineage score; whereas somatic cells have the converse scores (Fig. 2c, left panel). In contrast to focused gene sets (Fig. 2b), the pluripotency and lineage scores are complex genome-wide biomarkers that measure global transcriptional patterns associated with pluripotent and somatic cells[27] and it is the combined use of these two scores that allows separation of pluripotent from somatic samples (see Supplementary Note S2 for further discussion). Application of these classifiers confirmed a gradual movement away from the (day 0) pluripotent state (Fig. 2c, right panel and Fig. 2d). However, whereas the lineage score progressively increased following Nanog removal, indicating a gradual increase in acquired lineage characteristics, the pluripotency score showed a transient increase, remaining high 3 days after removal of Nanog, before decreasing (Fig. 2c, right panel). Principal component analysis (PCA) of the time-course data also revealed a similar pattern (Fig. 2e). Comparable dynamics have previously been noted during neural differentiation of human ES cells and induced pluripotent stem cells[26]. To gain a better understanding of this transient increase in the pluripotency score, we reanalysed a previously published data set from an *in vitro* differentiation time course of murine ES cell cultures to pluripotent epiblast stem cells[28] and observed a similar increase in the pluripotency score (see Supplementary Note S2 for further details). Thus, a transient pluripotency score increase seems to be characteristic of movement from a relatively naive ES cell state[3,23] to a poised cellular intermediate in which early differentiation programs and pluripotency circuitry run in parallel.

Gene expression changes are regulated in a combinatorial manner

To better determine the molecular mechanisms underpinning these observations we compared target gene expression changes with previously published promoter occupancy data[20] for each of the elements of the ES cell TRN (Figs 3 and 4). We found that many genes with significant expression changes on Nanog removal are direct targets of Nanog (Fig. 3a–c) and other members of the extended ES cell TRN (Figs 3d, 4 and Supplementary Table S4). In accordance with previous observations[20] we found highly combinatorial regulation of

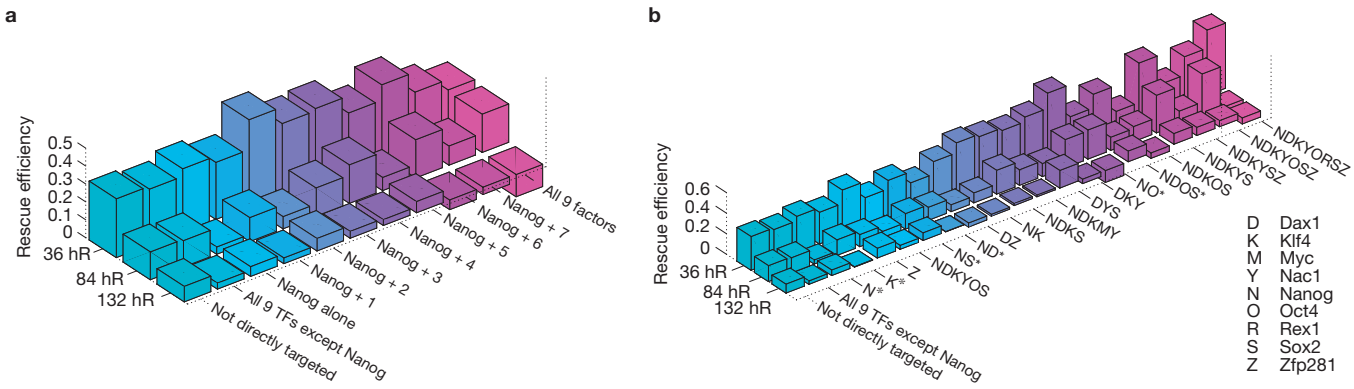

Figure 4 Gene expression changes are regulated in a highly combinatorial manner. (**a**,**b**) Rescue efficiency (see Methods) plotted against rescue time. Genes are grouped by the total number of factors in the ES TRN that directly regulate their expression (**a**) and the most significant regulatory combinations (odds ratio >1, Fisher's exact test P value <0.05, and 3 or more target genes; **b**). The asterisks highlight combinations that correspond to feedback loops in the ES cell TRN. A few combinations (Nanog + 6, Nanog + 7, All 9 factors, in **a**; DZ, DKY, NO, NDKOS, NDKORSZ, in **b**) have a negative rescue efficiency at the later time points, indicating that Nanog reintroduction resulted in further movement away from (rather than back towards) the initial state. Promoter occupancy data are from ref. 20.

target gene expression, with less than 1% of significantly changing genes being regulated by Nanog alone (Supplementary Table S4). In total, we identified 126 unique co-regulatory patterns (enumerated in full in Supplementary Table S4), indicating that the genome-wide changes that occur subsequent to Nanog removal are mediated through the combinatorial action of multiple factors. In accordance with this, we found that overall rescue efficiency (see Methods for details) also progressively diminished subsequent to Nanog removal as the core TRN shuts down (Fig. 4).

Taken together, these results indicate that mouse ES cells adopt a reversible primed state during short-lived downregulation of Nanog, characterized by promiscuous co-expression of pluripotency and early lineage markers as well as nascent engagement of cell-cycle checkpoints. However, in the continued absence of Nanog these changes become consolidated into committed fate decisions with an irrevocable downregulation of pluripotency genes and a concomitant upregulation of differentiation genes.

Early fate changes are stochastic and reversible at the single-cell level

As a number of key ES cell genes, including members of the core ES cell TRN such as *Nanog*, *Rex1* and *Klf4*, are heterogeneously expressed at the single-cell level[4–15], we reasoned that population-based microarray data might mask important cell–cell variability. To gain better insight into the molecular changes accompanying transient Nanog removal we conducted high-throughput single-cell transcriptional profiling. Time-lapse microscopy of individual ES cells has previously shown that stochastic fluctuations into a Nanog-low state last approximately 24 h (ref. 13). This timescale is consistent both with our observation that Nanog protein levels fall markedly within 24 h of Nanog downregulation[18] and the relative instability of Nanog protein ($t_{1/2} \sim 2$ h; ref. 29). Therefore, we sought to further investigate the effects of Nanog fluctuations over this natural 24 h timescale. We used the BioMark 96.96 Dynamic Array platform (Fluidigm) to profile a panel of 77 genes (Supplementary Table S5), including housekeeping-, pluripotency-, early-lineage- and cell-cycle-associated markers, in cells collected at 0 (dox present, Nanog-expressing, denoted 0 h), 24 and 36 h after dox withdrawal (denoted 24 h and 36 h respectively) and treated with a 12 h pulse of dox after 24 h without dox (denoted 36 hR; Fig. 5). In total, 384 individual cells were profiled covering these different time points. Overall, expression patterns derived from single cells showed a good correspondence with microarray population-based profiles and exhibited a non-trivial covariance structure (for a full discussion see Supplementary Note S2). Flow-cytometric single-cell analysis confirmed both efficient, synchronous downregulation of Nanog on dox removal, and efficient, synchronous rescue of its expression on reintroduction of dox (Fig. 5b). Consistent with previous publications[4–7,13,30] we found that mouse ES cells are highly heterogeneous with respect to their overall expression profiles (Fig. 5a). Single-cell analysis confirmed transient upregulation of transcripts associated with early differentiation and cell-cycle checkpoints on Nanog removal (Fig. 5a). Moreover, unsupervised clustering failed to identify discrete subpopulations (Fig. 5a), indicating that the early stages of differentiation subsequent to Nanog depletion occur as a gradual stochastic population drift rather than a collective and synchronous transition. A similar phenomenon was recently observed using high-throughput transcriptional profiling of single cells during haematopoiesis[31], suggesting that stochasticity in commitment may be an inherent feature of mammalian development. Although distinct clusters were not identified using unsupervised approaches, a distinction was apparent using a Support Vector Machine (SVM) classifier[32] using the 0 and 36 h data sets to train the benchmark pluripotent and lineage primed classes, respectively (Fig. 5c). A training misclassification rate (MCR) of 4% was achieved, indicating the presence of different patterns of expression in the two training samples. However, using this SVM only 45% of the 24 h cells were classified as pluripotent (Fig. 5c left panel), indicating that early fate changes are stochastic at the single-cell level, whereas 72% of the 36 hR cells were classified as pluripotent (Fig. 5c right panel), indicating reversibility at this early stage.

Feedback loops and ES cell fate commitment

Feedback loops (which can be positive, negative or mixed) commonly regulate phenotypic variability in diverse organisms and contexts

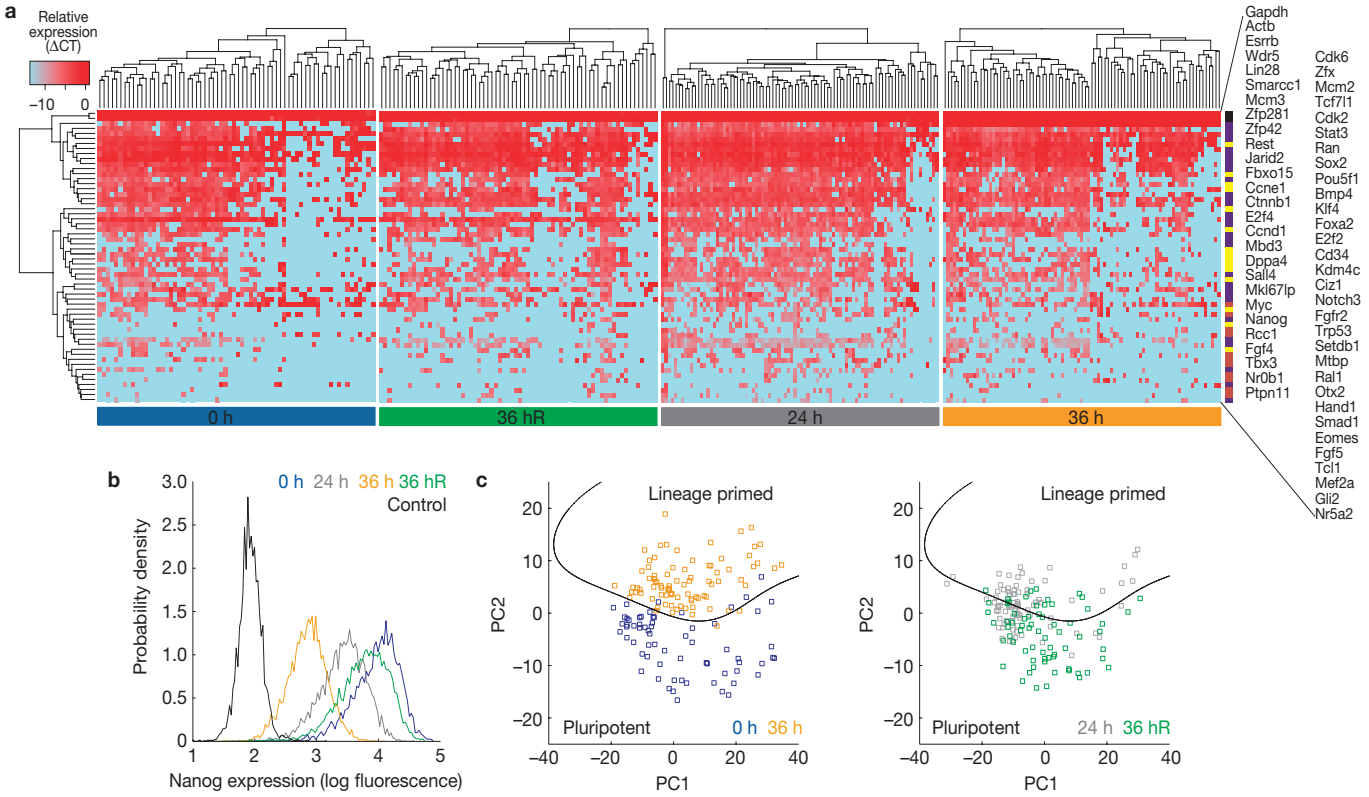

Figure 5 Single-cell gene expression patterns. (**a**) Heat maps of single-cell expression profiles. Highly expressed genes are in red; absent (not expressed) genes are in light blue. The coloured side bar identifies the following classes of genes: housekeeping genes (black); pluripotency-associated genes (purple); cell-cycle-associated genes (yellow); lineage-associated genes (brown). (**b**) Flow-cytometric analysis of Nanog expression levels during transient Nanog downregulation. (**c**) SVM classification of single-cell expression profiles plotted in the first two principal components. Training data sets (0 and 36 h) are in the left panel and the 24 and 36 hR data sets are in the right panel. In both panels, the black line separates the pluripotent and lineage primed classes.

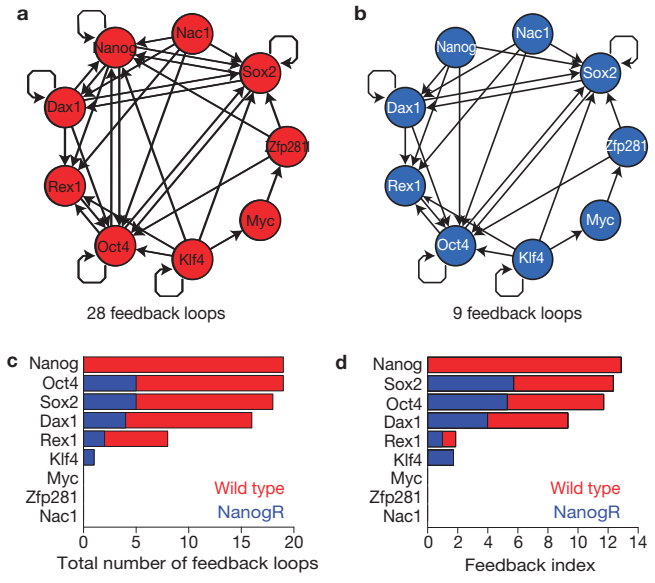

Figure 6 Feedback in the ES cell TRN (**a**) The feedback-rich wild-type TRN. (**b**) The feedback-depleted NanogR TRN (+dox, Nanog active). (**c**) The total number of feedback loops that each transcription factor participates in for the wild-type ES cell TRN (red) and NanogR TRN (blue). (**d**) Feedback centrality (as detailed in refs 45,46) for the wild-type ES cell TRN (red) and NanogR TRN (blue).

by generating complex dynamics[33], such as multi-stability[34–39], excitability[13] and oscillations[40–42], and by modulating molecular noise[43,44]. Accordingly, we reasoned that Nanog fluctuations might regulate early cell fate decisions and population variability by controlling feedback mechanisms in the ES cell TRN. To investigate this possibility we analysed the feedback structure of the extended ES cell TRN (ref. 20; Fig. 6). We found that this network is rich in feedback, containing a total of 28 distinct feedback loops (full details in Supplementary Table S6). Furthermore, these feedback loops are not evenly distributed (Fig. 6c). Rather, the global feedback structure of this network is highly nested and is critically dependent on Nanog, Oct4 and Sox2, which participate in 68% (19/28), 68% (19/28) and 64% (18/28) of all feedback loops, respectively (see Supplementary Table S6). Calculation of a simple returnability index[45,46] (see Methods for details), which takes into account both the total number and the lengths of all closed paths present in the extended TRN, identified Nanog as the most central element in the global feedback structure (Fig. 6d). Removal of Nanog therefore severely compromises overall feedback structure, leaving only 32% (9/28) of the feedback loops intact. Consequently, fluctuations in Nanog expression levels transiently activate different subnetworks in the ES cell TRN (ref. 30), driving transitions between a (Nanog-expressing) feedback-rich, robust and self-perpetuating pluripotent state and a (Nanog-diminished), feedback-sparse and differentiation-sensitive state.

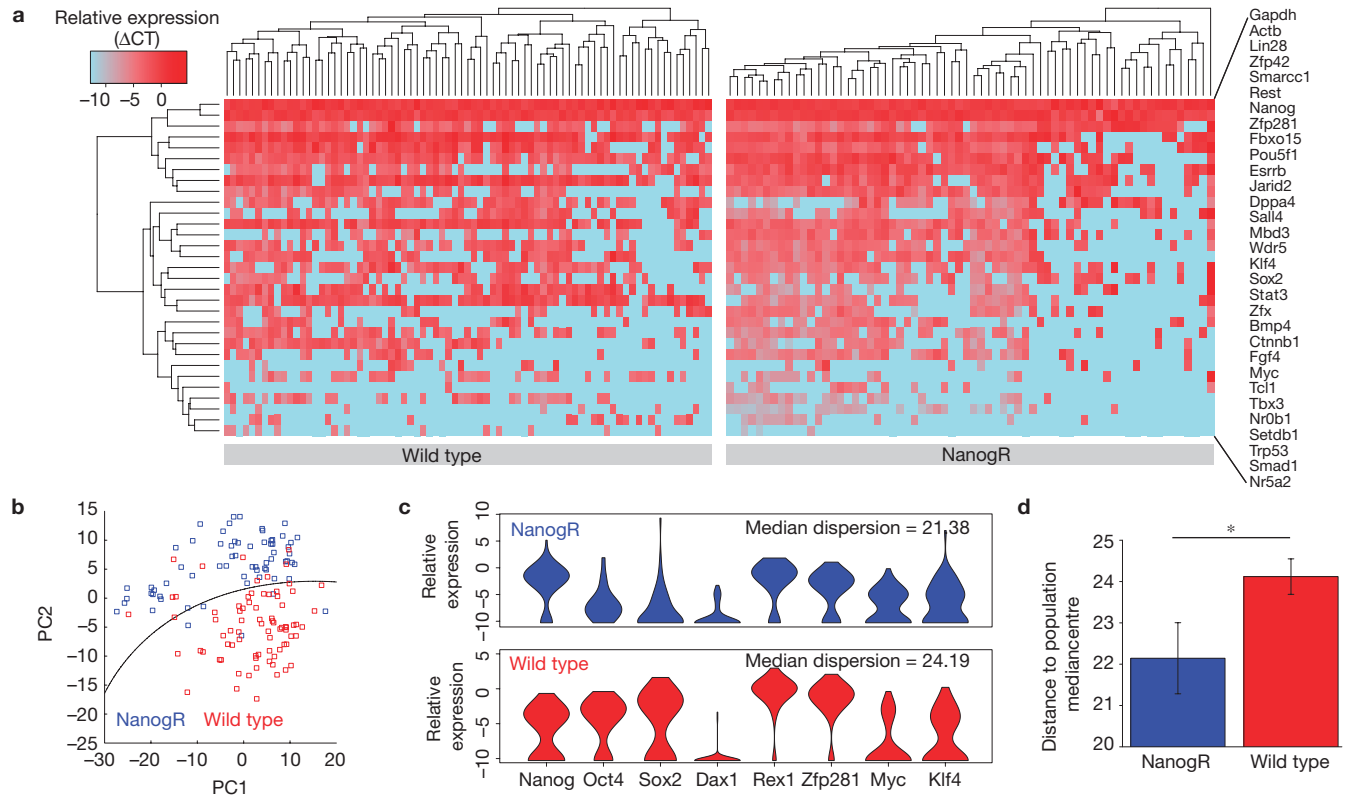

Figure 7 Cell–cell variability in wild-type and NanogR populations. (**a**) Heat maps of single-cell expression profiles in wild-type CCE and NanogR ES cell populations. Highly expressed genes are in red; absent (not expressed) genes are in light blue. (**b**) SVM classification of NanogR and wild-type CCE mouse ES cells plotted in the first two principal components. The black line separates the NanogR and wild-type classes. (**c**) Violin plots of single-cell relative expression levels in NanogR (blue) and wild-type CCE ES cells (red) for each factor in the extended ES cell TRN. Note, these plots show expression variation but not covariance. To quantify overall (multivariate) variability, the median dispersion of the populations was calculated (see Methods). Expression of Nac1 is not shown because it was not detected in sufficient numbers of cells in either NanogR or CCE ES cell populations to estimate its distribution. (**d**) The distance to the mediancentre may be used as a test statistic to assess significant differences in overall (multivariate) variability in NanogR and wild-type CCE cells. The asterisk indicates significance by a multivariate analogue of Levene's test with $P < 0.05$. The error bars show ± one standard error, $n = 66$ (NanogR) and $n = 77$ (wild-type).

We note that although the feedback structure of the extended TRN is severely compromised on removal of Nanog, it is not entirely destroyed: a small number of key feedback loops still remain, most notably those involving Oct4, Sox2, Dax1 and Rex1 (but not Nanog, see Supplementary Table S6). This may explain why, although they are prone to differentiate, ES cells can be maintained in a self-renewing state in the absence of Nanog[4]. In this situation self-renewing ES cells may adapt to rely on a compromised feedback structure. This also underscores the remarkably robust nature of the pluripotency TRN, and it will be interesting to determine whether ES cells can similarly adapt to loss of other network components; specifically, those with similar fluctuation properties.

Feedback loops and ES cell heterogeneity

To further assess the role of feedback in population heterogeneity we compared single-cell expression patterns in NanogR cells and in CCE wild-type ES cells. The wild-type ES cell TRN is self-perpetuating when shielded from differentiation-inducing stimuli[23]. However, in the NanogR cell line endogenous regulation of the *Nanog* gene does not contribute to Nanog protein levels. Consequently, all feedback loops that involve Nanog in the wild-type TRN are absent in the NanogR cells. In these cells the ES cell TRN is therefore effectively held in a feedback-depleted state (Fig. 6b) and maintenance of pluripotency is dependent on continued exogenous expression of Nanog rather than activation of self-perpetuating feedback loops. Importantly, NanogR cells are fully pluripotent and capable of producing germ-line chimaeras[19]. This highlights the largely dispensable nature of the complex endogenous feedback architecture in regulating pluripotency.

To investigate the effect of these changes to the TRN on cell–cell variability we compared single-cell expression patterns of 31 key pluripotency markers (Supplementary Table S7) in NanogR cells (grown in dox) with those in wild-type CCE ES cells. Overall, NanogR and CCE cells exhibited similar levels of marker expression (Fig. 7a,c), although they could be separated with a SVM classifier (9% MCR) (Fig. 7b), indicating some differences in expression patterns. However, we found that NanogR cells are less variable than CCE cells in overall marker expression patterns ($P = 0.03$ by a multivariate analogue of Levene's test for equality of variances, see Methods for details; Fig. 7d). Although it is only a comparison of two cell lines, this suggests that the feedback architecture of the ES cell TRN may play a role in controlling cell–cell variability.

DISCUSSION

Previously we have shown that removing Nanog results in a complex mixture of lineages[18,19]. The precise single-gene perturbation we have used in this study does not therefore reflect the full complexity of ES cell differentiation *in vivo*. Nevertheless, by initiating differentiation in a precise and tightly controlled manner, this model system provides a powerful means to study the early stages of differentiation. High-throughput single-cell profiling techniques are in their infancy and inevitably exhibit technical variation. Nevertheless, taken together our results indicate that Nanog-dependent feedback loops in the ES cell TRN play a role in controlling early fate changes at the single-cell level and heterogeneity at the population level. It remains to be determined whether feedback-controlled population heterogeneity has a role *in vivo*. We suggest that distinct individual states of a fluctuating TRN may reflect a variety of coexisting lineage primed differentiation tendencies that can respond to the presence of diverse stimuli. This remains to be investigated; however, a better understanding of the role of feedback in controlling ES cells will facilitate the maintenance of more defined pluripotent populations and the development of more robust differentiation protocols.

Accession codes

Primary accessions. All microarray data are deposited at the GEO database repository under accession number GSE40335.

Referenced accessions. GSE34243 (GEO). Further accession codes are given in Supplementary Table S3. □

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank K. Hochedlinger for the Nanog–GFP mouse ES cells[53]. We gratefully acknowledge funding support by the NIH (GM078465) and NYSTEM (C024410) to I.R.L. and by the NIH (GM095942) and NYSTEM (C026420) to J.W. This work was also supported by an EPSRC Doctoral Training Centre grant (EP/G03690X/1) and an EPSRC 2011/12 Institutional Sponsorship Award (EP/J501530/1).

AUTHOR CONTRIBUTIONS

B.D.M., A.S. and I.R.L. designed the project and prepared the manuscript. A.S., M.F. and J.W. performed the experiments. B.D.M., M.L., F.J.M., B.M.S., A.A.S., S.J.R., P.S.S. and A.M. performed the bioinformatic analyses and mathematical modelling. All authors reviewed and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at www.nature.com/doifinder/10.1038/ncb2603
Reprints and permissions information is available online at www.nature.com/reprints

1. Chambers, I. *et al*. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643–655 (2003).
2. Mitsui, K. *et al*. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).
3. Silva, J. *et al*. Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
4. Chambers, I. *et al*. Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230-U1238 (2007).
5. Canham, M. A., Sharov, A. A., Ko, M. S. H. & Brickman, J. M. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* **8,** e1000379 (2010).
6. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135,** 909–918 (2008).
7. Hayashi, K., Lopes, S. M. C. D., 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. Trott, J., Hayashi, K., Surani, A., Babu, M. M. & Martinez-Arias, A. Dissecting ensemble networks in ES cell populations reveals micro-heterogeneity underlying pluripotency. *Mol. Biosyst.* **8,** 744–752 (2012).
9. Macfarlan, T. S. *et al*. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
10. Niakan, K. K. *et al*. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev.* **24,** 312–326 (2010).
11. Singh, A. M., Hamazaki, T., Hankowski, K. E. & Terada, N. A heterogeneous expression pattern for nanog in embryonic stem cells. *Stem. Cells* **25,** 2534–2542 (2007).
12. Zalzman, M. *et al*. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464,** 858-U866 (2010).
13. Kalmar, T. *et al*. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7,** e1000149 (2009).
14. Kobayashi, T. *et al*. The cyclic gene Hes1 contributes to diverse differentiation responses of embryonic stem cells. *Genes Dev.* 1870–1875 (2009).
15. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460,** 118–122 (2009).
16. Arias, A. M. & Brickman, J. M. Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Curr. Opin. Cell Biol.* **23,** 650–656 (2011).
17. Stewart, M. H., Bendall, S. C., Levadoux-Martin, M. & Bhatia, M. Clonal tracking of hESCs reveals differential contribution to functional assays. *Nat. Methods* **7,** 917-U975 (2010).
18. Lu, R. *et al*. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462,** 358–U126 (2009).
19. Ivanova, N. *et al*. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442,** 533–538 (2006).
20. Kim, J., Chu, J. L., Shen, X. H., Wang, J. L. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049–1061 (2008).
21. Muller, F. J. *et al*. Regulatory networks define phenotypic classes of human stem cell lines. *Nature* **455,** 401–U455 (2008).
22. Ramirez, J. M. *et al*. Brief report: Benchmarking human pluripotent stem cell markers during differentiation into the three germ layers unveils a striking heterogeneity: All markers are not equal. *Stem Cells* **29,** 1469–1474 (2011).
23. Ying, Q. L. *et al*. The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–U515 (2008).
24. Aiba, L. *et al*. Defining developmental potency and cell lineage trajectories by expression profiling of differentiating mouse embryonic stem cells. *DNA Res.* **16,** 73–80 (2009).
25. Zhang, X. *et al*. A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. *J. Cell Biol.* **184,** 67–82 (2009).
26. Muller, F. J. *et al*. A bioinformatic assay for pluripotency in human cells. *Nat. Methods* **8,** 315–U354 (2011).
27. Williams, R., Schuldt, B. & Muller, F. J. A guide to stem cell identification: progress and challenges in system-wide predictive testing with complex biomarkers. *Bioessays* **33,** 880–890 (2011).
28. Nora, E. P. *et al*. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
29. Ramakrishna, S. *et al*. PEST motif sequence regulating human NANOG for proteasomal degradation. *Stem Cells Dev.* **20,** 1512–1520 (2011).
30. Xiong, W. & Ferrell, J. E. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426,** 460–465 (2003).
31. Pina, C. *et al*. Inferring rules of lineage commitment in haematopoiesis. *Nat. Cell Biol.* **14,** 287–294 (2012).
32. Cristianini, N. & Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (Cambridge Univ. Press, 2000).
33. Tyson, J. J., Chen, K. C. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15,** 221–231 (2003).
34. Smits, W. K., Kuipers, O. P. & Veening, J. W. Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* **4,** 259–271 (2006).
35. Ferrell, J. E. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14,** 140–148 (2002).
36. Becskei, A., Seraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20,** 2528–2535 (2001).
37. MacArthur, B. D., Ma'ayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* **10,** 672–681 (2009).
38. MacArthur, B. D., Please, C. P. & Oreffo, R. O. C. Stochasticity and the molecular mechanisms of induced pluripotency. *PLoS ONE* **3,** e3086 (2008).

39. MacArthur, B. D., Ma'ayan, A. & Lemischka, I. R. Toward stem cell systems biology: From molecules to networks and landscapes. *Cold Spring Harb. Symp. Quant. Biol.* **73,** 211–215 (2008).
40. Tigges, M., Marquez-Lago, T. T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457,** 309–312 (2009).
41. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).
42. Glauche, I., Herberg, M. & Roeder, I. Nanog variability and pluripotency regulation of embryonic stem cells—Insights from a mathematical model analysis. *PLoS ONE* **5,** e11238 (2010).
43. Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439,** 608–611 (2006).
44. Paulsson, J. Summing up the noise in gene networks. *Nature* **427,** 415–418 (2004).
45. Estrada, E. & Rodriguez-Velazquez, J. A. Subgraph centrality in complex networks. *Phys. Rev. E* **71,** 056103 (2005).
46. Estrada, E. & Hatano, N. Returnability in complex directed networks (digraphs). *Linear Algebra Appl.* **430,** 1886–1896 (2009).
47. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).
48. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38,** 431–440 (2006).
49. Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev.* **22,** 746–755 (2008).
50. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
51. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134,** 521–533 (2008).
52. Mathur, D. *et al.* Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET. *Genome Biol.* **9** (2008).
53. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1,** 55–70 (2007).

DOI: 10.1038/ncb2603

METHODS

ES cell culture. Mouse ES cells were cultured as previously described[19]. Briefly, NanogR ES cells were cultured in dox (1 μg ml^{-1} Sigma) on 0.1% gelatin-coated tissue culture plates without feeder cells for all experiments. Routine media includes: D-MEM high glucose (Dulbecco's modified Eagle's medium 1× high glucose), 15% FBS (fetal bovine serum; Hyclone), 100 mM MEM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 1 mM L-glutamine and penicillin/streptomycin (Invitrogen) and 10^3 units ml^{-1} LIF (Chemicon). To induce differentiation, we withdrew dox from the media, but still maintained all other routine ES cell nutrients. All cell cultures were maintained at 37 °C with 5% CO_2 and were plated at a density of 3×10^5 cells per 10 cm dish.

Real-time quantitative PCR. Cells were trypsinized and collected at specific time points. Total RNA was extracted using Trizol Reagent (Invitrogen), column-purified with RNeasy kit (Qiagen) and treated with RNase-free DNase (Qiagen). Total RNA (1 μg) was reverse transcribed using a high-capacity reverse transcription kit (Applied Biosystems). All quantitative PCR analyses were performed using the Fast SYBR Green Master Mix (Applied Biosystems) following the manufacturer's protocols on the Light Cycler 480 Real-Time PCR System (Roche). All measurements were performed in technical triplicate. The strategy for primer design to discriminate between endogenous and exogenous *Nanog* is depicted in Supplementary Fig. S3 as previously described[54]. Primer sequences are listed in Supplementary Table S8.

Antibodies. The following commercially available antibodies were used at the indicated concentrations for western blot: α–β-actin (Sigma, catalogue number A5441, clone AC-15, 1:1,000), Nanog (ReproCell, catalogue number RCAB0002P-F, 1:1,000), Oct3/4(N-19) (Santa Cruz, sc-8628, 1:1,000), Sox2 (Y-17) (Santa Cruz, sc-17320, 1:2,000), Dax1(K-17) (Santa Cruz, sc-841, 1:1,000), c-Myc (N-262) (Santa Cruz, sc-764, 1:1,000), Nac-1 (Abcam, ab29047, 1:250), Zfp-281 (Abcam, ab112047, 1:5,000), GKLF(H-180) (Santa Cruz, sc-20691, 1:1,000), Rex-1 (Abcam, ab 28141, 1:1,000).

Microarray gene expression profiling. RNA probes from each time point were hybridized to Affymetrix Gene Chip Mouse Gene 1.0 ST microarrays (3 biological replicates; 30 arrays in total) according to the manufacturer's protocols by the Genomics Core Laboratory in The Institute for Personalized Medicine at Mount Sinai Medical Center. Data were normalized using the Robust Multichip Average (RMA) method in the Affymetrix Expression Console software. Expression measurements were obtained by taking the mean readings for gene-specific probe sets and the data were $\log_2$ normalized.

Microarray data analysis. Significant genes for each time series were found using one-way analysis of variance at a Bonferroni-corrected P value of 0.05 (uncorrected P value of 1.9×10^{-6}) and a total fold change of greater than 2. In total, 1,132 significant genes were identified. All subsequent clustering and dimensionality reduction analyses were performed in R and Matlab, using the Bioinformatics and Statistics Toolboxes. Data were row standardized to avoid bias towards highly expressed genes and hierarchical clustering and PCA were performed on standard scores. Hierarchical clustering was performed using the Euclidean distance metric and the average linkage function.

Let x_t denote the microarray expression profile obtained at time $t > 0$ and let x_t^R and denote the corresponding microarray expression profile subsequent to Nanog rescue. Let $d(.,.)$ denote the Euclidean distance. The ability of Nanog reintroduction (at time t) to reverse the differentiation trajectory is measured by the rescue efficiency, which is defined as $1-(d(x_t^R,x_0))/(d(x_t,x_0))$. A rescue efficiency of 1 indicates perfect reversion back to the initial state; a rescue efficiency of zero indicates no reversion back to the initial state; a negative rescue efficiency indicates further movement away from the initial state on Nanog reintroduction.

Bioinformatic assay for mouse cell pluripotency. To construct the classifiers we modified the PluriTest algorithm[26], a recently developed bioinformatic assay for pluripotency in human cells[26]. Briefly, probe identifiers from the Affymetrix GeneChip Mouse Gene 1.0 ST array were matched by homology to the probe identifiers in the Illumina Human HT 12 bead array, the platform used for the original PluriTest. Homology mapping was conducted using the getLDS function in the biomaRt package for R (ref. 47). To calibrate the algorithm we collated and manually curated a reference data set consisting of 1062 Affymetrix Mouse Gene 1.0 ST array samples representing cells of known phenotypes (Supplementary Table S2) from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). This reference data set was normalized together with the NanogR time series data using the RMA-sketch algorithm as part of the Affymetrix Power Tools software. Pluripotency scores and lineage scores were then calculated using the metagenes method[26]. The complete R code is included in the Supplementary Software folder.

Single-cell gene expression profiling. Single cells were sorted directly into a mixture of CellsDirect 2× Reaction Mix (component of CellsDirect One-Step qRT PCR Kits, Invitrogen), 0.2× TaqMan Assay Mix (Applied Biosystems) and SuperScript III RT/Platinum Taq Mix (Invitrogen). Reverse transcription (RT) and specific target amplification were performed sequentially. Conditions for reverse transcription were 15 min at 50 °C. Samples were then incubated for 2 min at 95 °C to inactivate the reverse transcription enzyme and activate Taq polymerase. Samples were pre-amplified for 22 cycles of 95 °C for 15 s and 60 °C for 4 min. Pre-amplified complementary DNA was diluted with TE buffer (1:5) and was used for the real-time quantitative PCR. Single-cell gene expression was analysed using BioMark 96-96 Dynamic Arrays (Fluidigm) according to the manufacturer's protocols. The PCR profile included a 10 min, 95 °C hot-start to activate Taq polymerase, followed by 40 cycles of a two-step programme: 15 s at 95 °C and 60 s at 60 °C. Data were processed using BioMark Real-Time PCR Analysis Software version 2.0 (Fluidigm) to obtain Ct values.

Single-cell gene expression data analysis. All analyses were conducted in R and Matlab using the Bioinformatics and Statistics Toolboxes. Ct values were converted to relative expression levels using a variation of a previously described method[55]. A maximum Ct value of 28 was assumed. Ct values were normalized to endogenous controls by subtracting the average of *Actb* and *Gapdh* expression levels. Cells that did not express both *Actb* and *Gapdh* were excluded from analysis. Contaminated channels, as assessed by a H_20 control, were also excluded. Similarly, genes that were expressed in less than 10% of cells from each sample were also excluded from analysis. Genes that were not expressed were set to an assumed minimum value 10% lower than the lowest recorded reading. Hierarchical clustering was conducted using the Euclidean distance metric and complete linkage function. PCA was conducted on mean-centred data to avoid bias towards highly expressed genes. SVM classifiers were constructed using a Gaussian radial basis function kernel and model parameters were chosen to minimize the MCR using 10-fold cross-validation. Minimization of MCR was performed using the Nelder–Mead simplex algorithm starting from 1,000 random initial conditions. To provide easily interpretable figures SVM classification was also performed on the projection of the full data set onto the first two principal components (see Figs 5c and 7c). To ensure that the observed results were not due to changes in Nanog expression alone, we also conducted SVM classification excluding Nanog expression patterns from the analysis. SVM classification subsequent to PCA exhibited a similar trend to classification on the full data set, as did SVM classification excluding Nanog. Figures quoted in the main text refer to classification performed on the full data set. Comparison of gene expression variability in NanogR and wild-type ES cell populations was conducted using a multivariate analogue of Levene's test[56,57]. Briefly, let x_{ij} be the relative expression of the jth gene in the ith cell in the NanogR cell population and let y_{ij} be the relative expression of the jth gene in the ith cell in the wild-type CCE cell population. Nanog is, by design, more homogeneously expressed in the NanogR cell line than the wild-type CCE line (see Fig. 1b). Therefore, to accurately compare expression variability in the two populations we excluded Nanog expression patterns from the analysis. Let $d(.,.)$ denote the Euclidean distance. The mediancentre of a group of points is that point in the multidimensional space for which the sum of Euclidean distances from each observation in the group to m is minimized[58]. The mediancentre is a multivariate analogue of the univariate median and is a measure of the central tendency of a multivariate sample that is robust to outliers. Let m_j^{Ngr} be the mediancentre of the NanogR population and m_j^{wt} be the mediancentre of the wild-type CCE population. The variables $s_i = d(x_{ij}, m_j^{Ngr})$ and $t_i = d(y_{ij}, m_j^{wt})$, which give the Euclidean distance of the ith cell from its population mediancentre, were then compared using a one-sided bootstrap hypothesis test (2,500,000 replicates) with hypotheses H_0 (the NanogR and wild-type populations have the same variability) and H_1 (the wild-type population is more variable than the NanogR population). The median values of s_i and t_i were used to quantify the total dispersion of the two populations.

FACS analyses and sorting. Wild-type Nanog–GFP ES cells were treated with puromycin (1 μg ml^{-1}) for three successive passages and then three additional passages without drug selection to obtain a comparable Nanog–GFP distribution to that previously reported[4]. These cells were sorted as the high-Nanog subpopulation using a BD Bioesciences Influx cell sorter, and were then cultured for 14 days before being evaluated on a BD Bioesciences LSRII FACs analyser.

Quantification of network feedback loops. The total numbers of feedback loops of each length were enumerated using the adjacency matrix method of ref. 59. The longest feedback loop in this network has a length of 5, because only 5 nodes in the extended ES cell TRN (Nanog, Oct4, Sox2, Dax1 and Rex1) have both incoming and outgoing edges. As the ES cell TRN is small, specific feedback loops were found by exhaustive enumeration within a few seconds using a bench-top PC. The returnability index given in Fig. 6 is that of refs 45,46. The index for the ith node in the network is $B_{ii} - 1$, where $B_{ij} = e^{A_{ij}}$ is the matrix exponential of the network

adjacency matrix. This is a measure of returnability (rather than direct feedback) because it represents a weighted sum of all closed walks in the network. Thus, it provides a convenient index of node involvement in both direct (non-intersecting) and indirect (self-intersecting) feedback. The −1 term is included for convenience to ensure that nodes that do not participate in any closed walks have an index of zero.

54. Lee, D. F. *et al*. Combining competition assays with genetic complementation strategies to dissect mouse embryonic stem cell self-renewal and pluripotency. *Nat. Protoc.* **7,** 729–748 (2012).
55. Guo, G. J. *et al*. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18,** 675–685 (2010).
56. O' Brien, P. C. Robust procedures for testing equality of covariance matrices. *Biometrics* **48,** 819–827 (1992).
57. Manly, B. F. J. *Multivariate Statistic Methods: A Primer* (Chapman and Hall/CRC, 2005).
58. Gower, J. C. Algorithm AS 78: The mediancentre. *J. R. Stat. Soc. Ser. C* **23,** 466–470 (1974).
59. Harary, F. & Manvel, B. On the number of cycles in a graph. *Math. Slovaca* **21,** 55–63 (1971).

nature cell biology

DOI: 10.1038/ncb2603

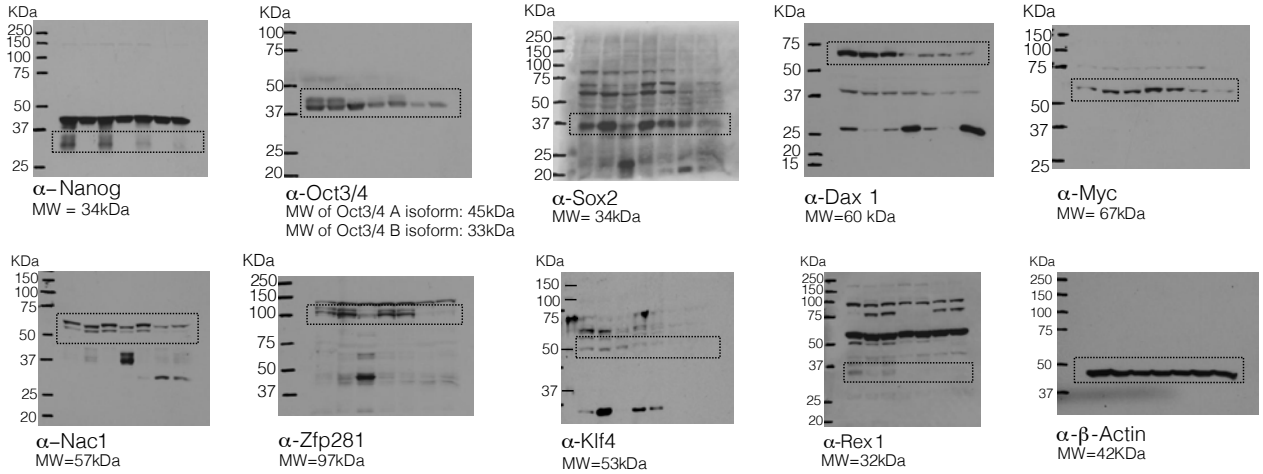

Figure S1 Full Scans from **Fig. 1d**.

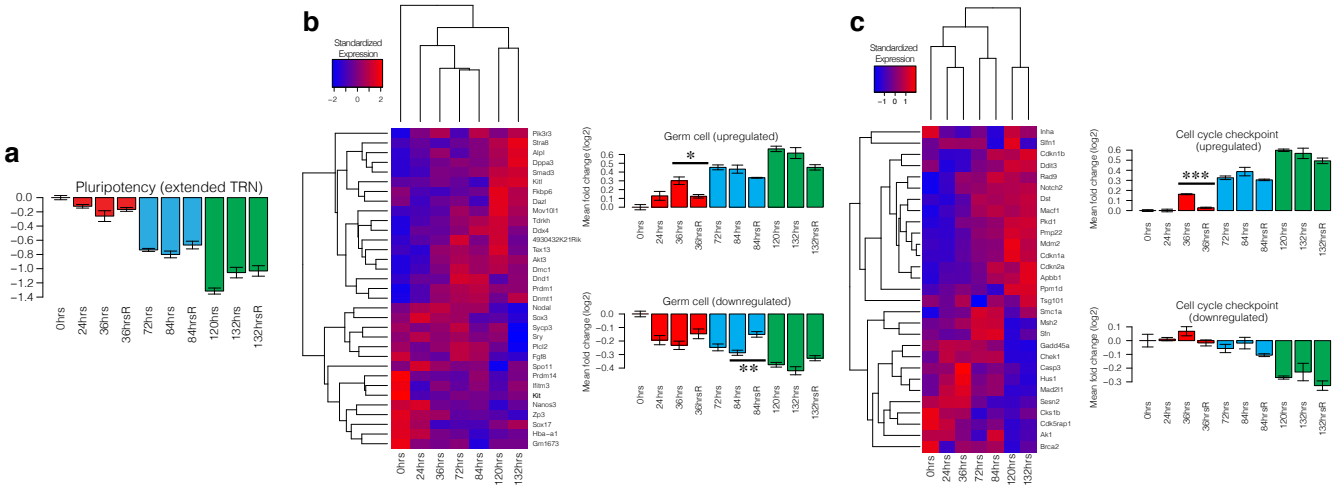

Figure S2 (a) Mean fold changes in expression for elements of the ES cell TRN. **(b)** Expression patterns and mean fold changes in expression for germ cell associated genes (**Supplementary Table S1**). Expression patterns are not uniform, so mean fold changes are shown separately for those genes that are upregulated and downregulated during the time-course. **(c)** Expression patterns and mean fold changes in expression for cell cycle checkpoint associated genes (**Supplementary Table S1**). Expression patterns are not uniform, so mean fold changes are shown separately for those genes that are upregulated and downregulated during the time-course. In all panels error bars show ± one standard error, $n = 3$.

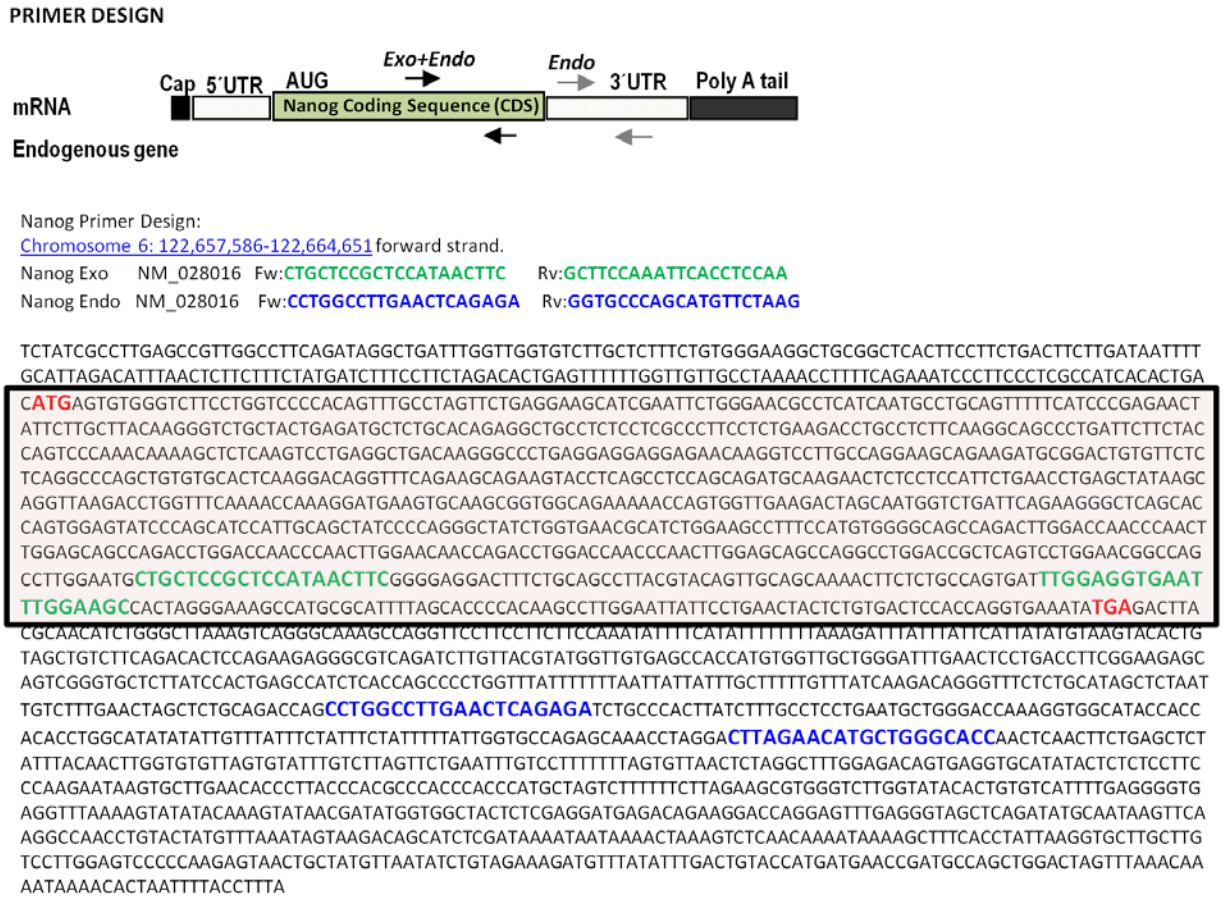

Figure S3 Primer design and sequences to discriminate between endogenous and exogenous *Nanog*.