ISSCR

Single-Cell Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal

Dmitri Papatsenko,[1,2,7,*] Henia Darr,[1,2,7] Ivan V. Kulakovskiy,[3,4] Avinash Waghray,[1,2] Vsevolod J. Makeev,[3,4] Ben D. MacArthur,[5] and Ihor R. Lemischka[1,2,6,*]

[1]Department of Regenerative and Developmental Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA
[2]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA
[3]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilova Strasse 32, Moscow 119991, Russia
[4]Department of Computational Systems Biology, Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkina Strasse 3, Moscow 119991, Russia
[5]Centre for Human Development, Stem Cells, and Regeneration, Institute of Developmental Sciences, University of Southampton, Southampton SO17 1BJ, UK
[6]Department of Pharmacology and System Therapeutics, Icahn School of Medicine at Mount Sinai, Systems Biology Center New York, One Gustave L. Levy Place, New York, NY 10029, USA
[7]Co-first author
*Correspondence: dmitri.papatsenko@mssm.edu (D.P.), ihor.lemischka@mssm.edu (I.R.L.)
 http://dx.doi.org/10.1016/j.stemcr.2015.07.004

SUMMARY

Analyses of gene expression in single mouse embryonic stem cells (mESCs) cultured in serum and LIF revealed the presence of two distinct cell subpopulations with individual gene expression signatures. Comparisons with published data revealed that cells in the first subpopulation are phenotypically similar to cells isolated from the inner cell mass (ICM). In contrast, cells in the second subpopulation appear to be more mature. Pluripotency Gene Regulatory Network (PGRN) reconstruction based on single-cell data and published data suggested antagonistic roles for Oct4 and Nanog in the maintenance of pluripotency states. Integrated analyses of published genomic binding (ChIP) data strongly supported this observation. Certain target genes alternatively regulated by OCT4 and NANOG, such as *Sall4* and *Zscan10*, feed back into the top hierarchical regulator *Oct4*. Analyses of such incoherent feedforward loops with feedback (iFFL-FB) suggest a dynamic model for the maintenance of mESC pluripotency and self-renewal.

INTRODUCTION

Mechanisms that control self-renewal and the ways of manipulating them in predictive manner are among the key current problems in regenerative biology. Addressing these issues is necessary to realize the biomedical potential of both embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs).

The stabilization of mouse ESCs (mESCs) is achieved by supplementing the culture medium with the cytokine leukemia inhibitory factor (LIF), which activates pluripotency genes via the Jak-Stat pathway and relieves their potential repression by Gsk3 (Niwa et al., 2009). LIF-mediated external signaling becomes dispensable in mESCs when the major mediators of signaling pathways ERK1/2 and GSK3 are inhibited by small molecules ("2i" conditions) (Martello et al., 2012). Under 2i conditions (Marks et al., 2012), the stability of the pluripotent state is not associated with extrinsic signals such as LIF, suggesting the presence of a self-sustainable mechanism of pluripotency supported by an intrinsic pluripotency gene regulatory network (PGRN).

Molecular ESC analyses and the discovery of iPSC reprogramming attributed pluripotency and self-renewal functions to the transcriptional regulators OCT4, SOX2, NANOG, and others (Ivanova et al., 2006; Loh et al., 2006; Takahashi and Yamanaka, 2006). Since then, numerous PGRN versions have been proposed (Festuccia et al., 2013; Yang et al., 2010), and the list of factors continues to grow. However, the hierarchy of the PGRN, the order of regulatory links, and the principles of PRGN function remain largely elusive.

In our studies (see flowchart in Figure S1), hundreds of single mESCs grown either under serum + LIF or serum-free 2i + LIF conditions have been collected, and their expression signatures with respect to 46 pluripotency genes were retrieved using high-throughput microfluidic single-cell qRT-PCR (White et al., 2011). Clustering individual cells based on their gene expression profiles revealed the presence of two major cell subpopulations in cells grown under the serum + LIF condition. In contrast, under 2i conditions, the two populations collapsed into one, which is in agreement with recent data suggesting a reduction in gene expression heterogeneity in 2i versus LIF alone (Marks et al., 2012). Comparison of our single-cell data with published single-cell data (Kumar et al., 2014; Tang et al., 2010) established that one subpopulation detected under the LIF condition has a gene expression signature similar to that of the inner cell mass (ICM) (Boroviak et al., 2014), whereas the other subpopulation resembles

more mature epiblast cells from the embryo. Detection of subpopulations became possible here because of the large number of analyzed cells (96 cells on each chip, seven chips in total).

We integrated the single-cell data obtained in this study with the data available for knockdowns of major pluripotency transcription factors (Feng et al., 2009; Ivanova et al., 2006; Loh et al., 2006; Lu et al., 2009; Martello et al., 2012). PGRNs reconstructed based on the integrated data revealed network motifs such as incoherent feedforward loops (iFFL) (Goentoro et al., 2009; Milo et al., 2002; Papatsenko and Levine, 2011), linking OCT4 and NANOG with their target genes and suggesting an antagonistic interaction between OCT4 and NANOG. Certain genes alternatively regulated by OCT4 and NANOG (*Sall4* and *Zscan10*) appear to feed back to *Oct4* and *Sox2*. We discuss how these loops may stabilize OCT4 concentrations required for self-renewal.

RESULTS

ESCs Grown in Serum + LIF Contain Two Cell Subpopulations

Potential medical applications involving ESCs or iPSCs require reproducible behavior of the cultured cells. From this perspective, the presence of cell subpopulations and phenotypic variations in stem cell cultures represent obstacles. Previously, subpopulations have been identified in mESCs grown under LIF conditions based on the expression of NANOG, REX1, and other pluripotency factors (Toyooka et al., 2008). Recent technologies allow the assessment of many genes in statistically sound numbers of cells (96 × 96 Fluidigm chips) (White et al., 2011), pushing the limits of detection of phenotypic variations and subpopulations.

Here single cell gene expression data were collected for 48 genes in 480 individual mESCs grown under serum + LIF conditions and in 192 individual cells grown under 2i + LIF conditions. The list of 48 genes included controls, core pluripotency transcription factors (*Oct4*, *Sox2*, *Nanog*, etc.), epigenetic regulators (*Ctcf*, *Jarid2*, etc.), and differentiation markers (*Gata6*, *Pax6*, etc.). The collected single-cell data were normalized using a relative normalization method (Experimental Procedures; Figures S2A–S2G; Table S1).

To establish the subpopulation structure, the single-cell gene expression data were clustered based on gene expression. Figure S2H shows hierarchical clustering, and Figure 1 shows co-expression clustering and principal component analysis (PCA) (Experimental Procedures) for serum + LIF cells. Hierarchical as well as co-expression clustering revealed two major cell clusters in serum + LIF cells, corre-

sponding to two major mESC subpopulations (Figures 1A–1F; Figure S2H). Each cluster is characterized by a specific expression signature (Figure 1G). Cluster 1 cells occupy positions of ICM cells from published data (Tang et al., 2010; Figures 1H and 1I). Clustering data for 2i + LIF cells using the same set of methods showed that these cells occupy a position different from the two major LIF clusters (Figures 1B, 1D, 1F, 1J, and 1K). The detected variation of gene expression levels (Figures S3A–S3AB) was in agreement with previously reported reduced gene expression variations under 2i + LIF conditions (Wray et al., 2010). The fuzziness of the identified clusters may be attributed to the precision of the detection method and/or to intrinsic stochastic variation in expression levels not related to the deterministic expression signatures identified here (two signatures in serum + LIF and one in 2i + LIF).
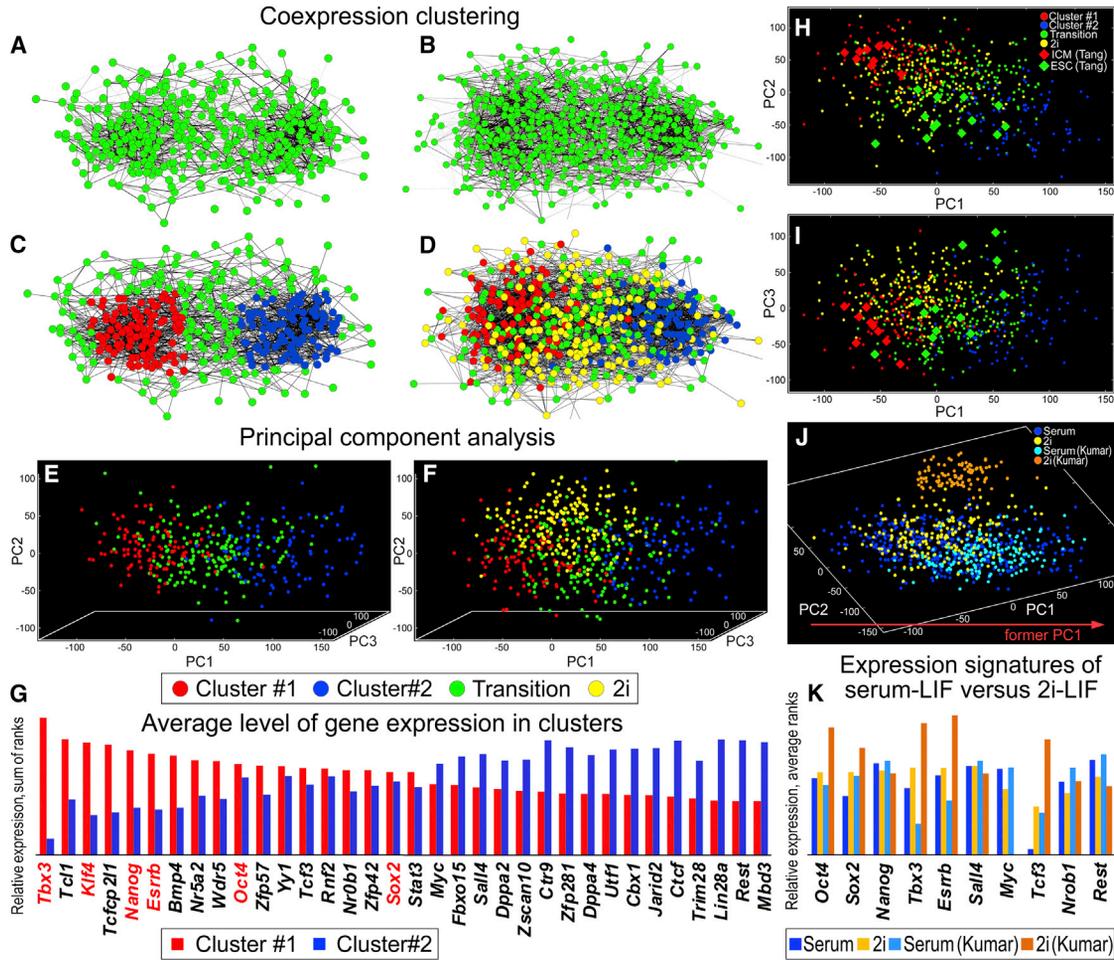
Identities of Subpopulations

To characterize cell identities in the two mESC subpopulations, the corresponding cell clusters were compared with published single-cell data obtained from the ICM, epiblast cells, and ESCs (Boroviak et al., 2014; Tang et al., 2010). These analyses have shown (Figures 1H and 1I and 2) that cells in cluster 1 (∼50% of cells grown in serum + LIF) have expression signatures similar to ICM from embryonic day 3.5 (E3.5) early blastocysts or even E2.5 morulae (Figures 2A–2C), whereas cluster 2 cells may be similar to more mature epiblast E5.5.cells. The previous view that mESCs have the identity of preimplantation epiblast cells (Plusa and Hadjantonakis, 2014) is not in contradiction with the current findings because the averaged expression signature for the two clusters (E2.5 and E5.5) may well resemble preimplantation epiblast cells (E4.0–E4.5).

To further characterize the cell identities in clusters 1 and 2, temporal expression profiles for genes explored in this study were retrieved from available ESC differentiation data (Hailesellasse Sene et al., 2007; Lu et al., 2009). Figure 2D shows that a majority of genes highly expressed in cluster 1 have peak expression in undifferentiated mESCs (day 0). In contrast, cluster 2 genes have peak expression at the onset of differentiation (12–24 hr, Figure 2E).

Together, our analyses suggest that serum + LIF cells contain two subpopulations or major cell types: cluster 1, matching E2.5/E3.5 ICM cells, and cluster 2, matching E5.5 epiblast cells. Cluster 2 cells also express higher levels of genes associated with the onset of differentiation (*Jarid2*, *Ctcf*, *Zscan10*, and *Lin28a*). Cells in both clusters express similar levels of Oct4 and similar levels of differentiation markers (Figure S3D).

Pathways and Factors Stabilizing mESC States
Although the presence of an E2.5 morula/E3.5 ICM state (cluster 1) in ESCs is expected (ICM is the source of ESCs),
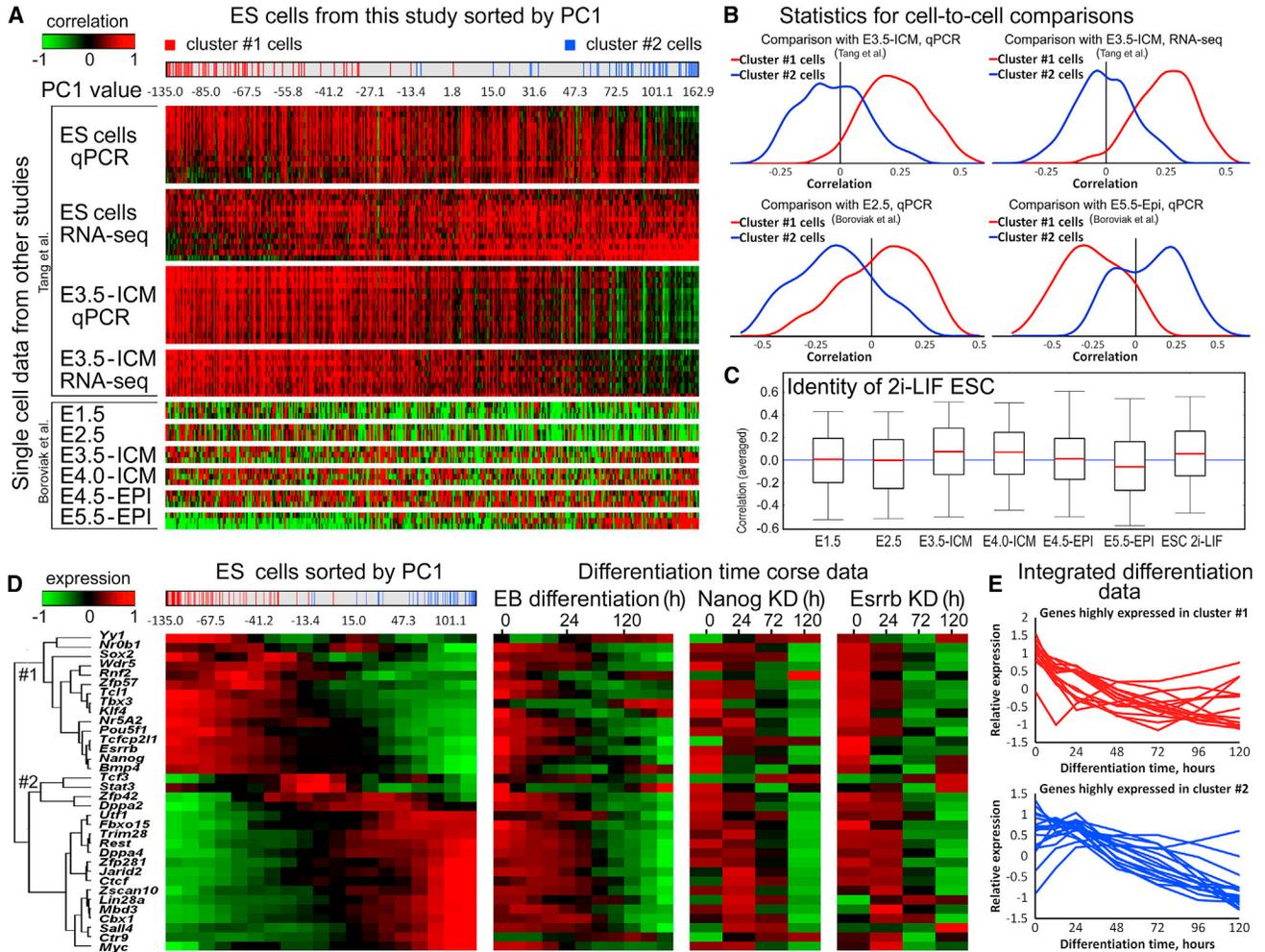
Figure 1. Identified ESC Subpopulations

(A) Co-expression clustering of cells grown under LIF conditions based on gene expression.

(B) Co-expression clustering of the integrated dataset containing cells grown under LIF and 2i conditions.

(C) Same as (A), with the exception that cells in the identified clusters are shown in red and blue.

(D) Clustering as in (B). LIF cells are labeled as in (C). Cells grown under 2i conditions are shown in yellow.

(E) PCA of cells grown on LIF.

(F) PCA of the integrated LIF and 2i datasets. Cells are highlighted as in (C) and (D).

(G) Gene expression signatures of the identified LIF clusters.

(H and I) PCA of data from this study and published data for dissociated embryos and ESCs (Tang et al., 2010). The positions of ICM cells coincide with cluster 1.

(J) PCA of data from this study and published single-cell RNA-seq data (Kumar et al., 2014). The RNA-seq data better capture the absolute expression levels; therefore, the 2i RNA-seq cluster was observed separately. However, the positions of both 2i datasets along the coordinate corresponding to PC1 in (E) and (F) (former PC1 in J) in both datasets are similar.

(K) Comparison of the relative expression of major pluripotency factors under 2i versus serum conditions demonstrating the presence of a common signature similar to both 2i datasets.

the presence and origin of the second subpopulation (cluster 2) is more obscure. To explore both states on a genome-wide level, differentiation time series datasets (Hailesellasse Sene et al., 2007; Lu et al., 2009) were searched for genes with a peak of expression in undifferentiated ESCs (0 hr), matching cluster 1 genes, and with expression peaks at the onset of differentiation (12–24 hr), matching genes in

cluster 2. The results were validated computationally using published single-cell RNA sequencing (RNA-seq) data (Tang et al., 2010; Figure 3A). Five ICM-derived cells were selected, displaying the highest similarity to cells in cluster 1, and five ESC cells were selected, displaying the highest similarity with cells in cluster 2 (Figures 3B and 3C). Strikingly, 37 of 53 genes (69.8% of genes matching

Figure 2. Cell Identities in ESC Subpopulations

(A) A cell-to-cell comparison of data from this study sorted along PC1 (Figure 1) and published single-cell data for embryo development (Boroviak et al., 2014; Tang et al., 2010). The color bar at the top left shows correlation based on gene expression. The bar at the top shows positions of cells from cluster 1 (red) and cluster 2 (blue) on PC1.

(B) Distributions of correlation values based on cell-to cell comparisons for cluster 1 (red) and cluster 2 (blue) cells. The expression signatures of cluster 1 cells are similar to cells isolated from the E3.5 ICM (Tang et al., 2010) or even E2.5 morula (Boroviak et al., 2014). Cluster 2 cells are similar to epiblast cells at E5.5.

(C) 2i cells produced in this study are similar to E3.5-ICM/E4.0-ICM cells.

(D) Left: the transition of gene expression from cluster 1 to cluster 2. The cells were sorted by PC1 and binned. The average normalized expression in each bin is shown. Right: published data for differentiation time courses. Most genes highly expressed in cluster 1 have peak expression before differentiation (0 hr), whereas most genes from cluster 2 have peak expression between 12 and 24 hr. EB, embryoid body; KD, knockdown.

(E) Temporal expression profiles for genes representing the two clusters (integrated data).

cluster 1, 0 hr) and 135 of 189 genes (71.4% of genes matching cluster 2, 12–24 hr) were evaluated as true positives based on RNA-seq data (Figure 3D; Tables S2 and S3).

Among the genes predicted to be highly expressed in cluster 1 are self-renewal factors such as *Klf5*, *Prdm14*, and *Tet2* (Table S3). Ectopic expression of KLF5 is sufficient to maintain ESCs in the absence of LIF (Parisi and Russo, 2011), and PRDM14 blocks differentiation toward extraem-

bryonic endodermal fates (Ma et al., 2011). Genes predicted to be highly expressed in cluster 2 include *p53*, *Pml*, *Tcea3*, *Nodal*, and others. P53 may interfere the WNT signaling pathway and inhibit differentiation (Abdelalim and Tooyama, 2014). overexpression of Tcea3 in the presence of differentiation signals blocks in vitro differentiation. TCEA3 appears to work via the LEFTY1/NODAL pathway (Park et al., 2013), and all three genes (*Tcea3*, *Lefty1*, and
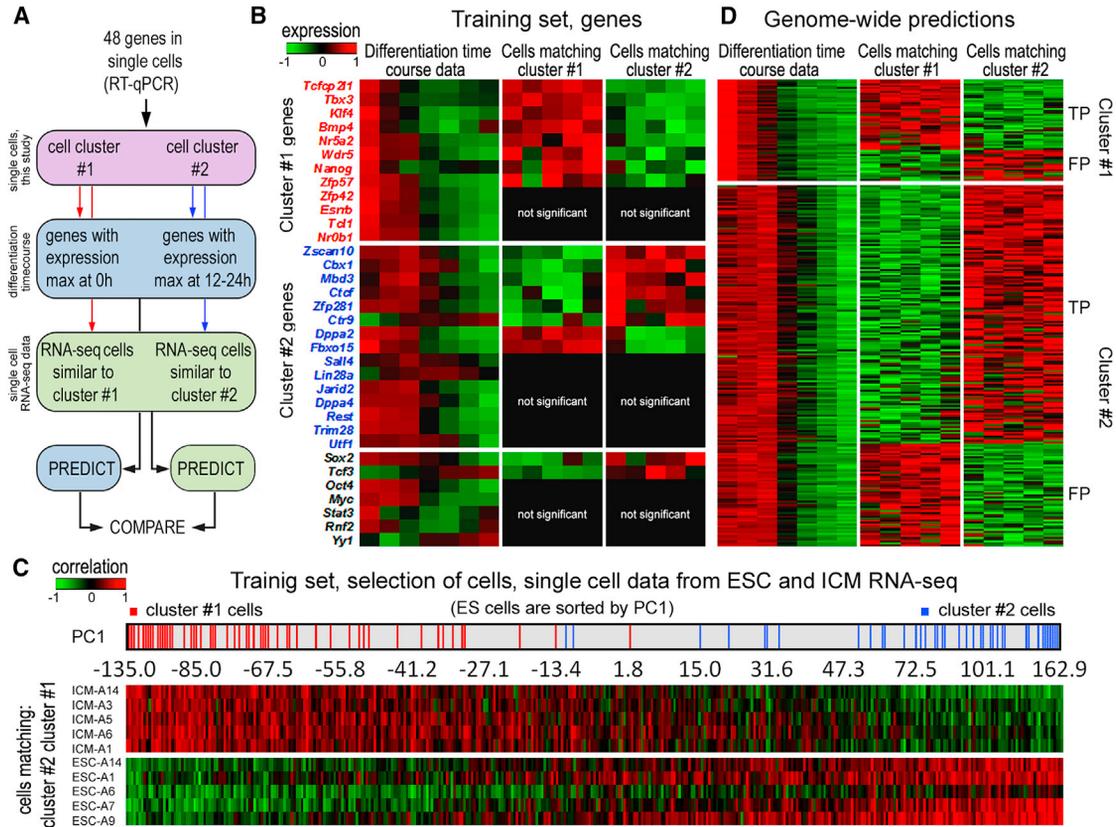
Figure 3. Genome-wide Predictions for Identities of LIF Cell Subpopulations

(A) Strategy of analysis.

(B) Left: integrated differentiation time course data for genes highly expressed in clusters 1 and 2. Center: published data for ICM cells, selected based on similarity to cluster 1. Right: published data for ESCs, selected based on similarity to cluster 2.

(C) A cell-to-cell comparison of data from this study sorted by PC1 and published single-cell data selected for computational validation.

(D) Left: genome-wide predictions for cluster identities based on the timing of gene expression during differentiation. Center and right: predictions based on data from the selected single cells with signatures similar to cluster 1 or 2. The two methods agree by >60%. TP, true positives; FP, false positives.

Nodal) were among the 135 genes predicted to be highly expressed in cluster 2. NODAL is essential for human ESCs (hESCs), which, phenotypically, are more mature than mESCs (James et al., 2005), suggesting a link between hESCs and the mature state (cluster 2) detected in mESCs in this study.

OCT4-NANOG Antagonism and Feedback Control of Pluripotency

Identification of cell clusters 1 and 2 under serum + LIF conditions suggested the presence of at least two groups of coordinately acting gene products. Co-expression clustering of the single-cell data with respect to genes supported this view (Figure S4A). Co-expression clustering and Bayesian network inference (Figures 4A–4C; Figures S4B–S4F) revealed that gene interactions within the gene clusters were mainly positive, whereas gene interac-

tions between the gene clusters were mainly negative (Figure 4C).

To reconstruct a signed and hierarchical PGRN, additional expression data from published knockdown studies targeting *Oct4*, *Nanog*, *Sox2*, and *Esrrb* were taken into consideration (Feng et al., 2009; Ivanova et al., 2006; Loh et al., 2006). Data from single-cell co-expression analyses, Bayesian inference based on the same studies, and published RNAi knockdown studies were integrated (Table S4). The resulting PGRN was searched for network motifs connecting up to three nodes and comprising iFFLs (Milo et al., 2002). The information-processing potential of iFFLs is of great interest because these motifs combine both positive and negative regulatory links, creating opportunities for alternative output solutions. This search returned 41 network motifs that were integrated into the network shown in Figure 4A. Several iFFLs integrated OCT4 and
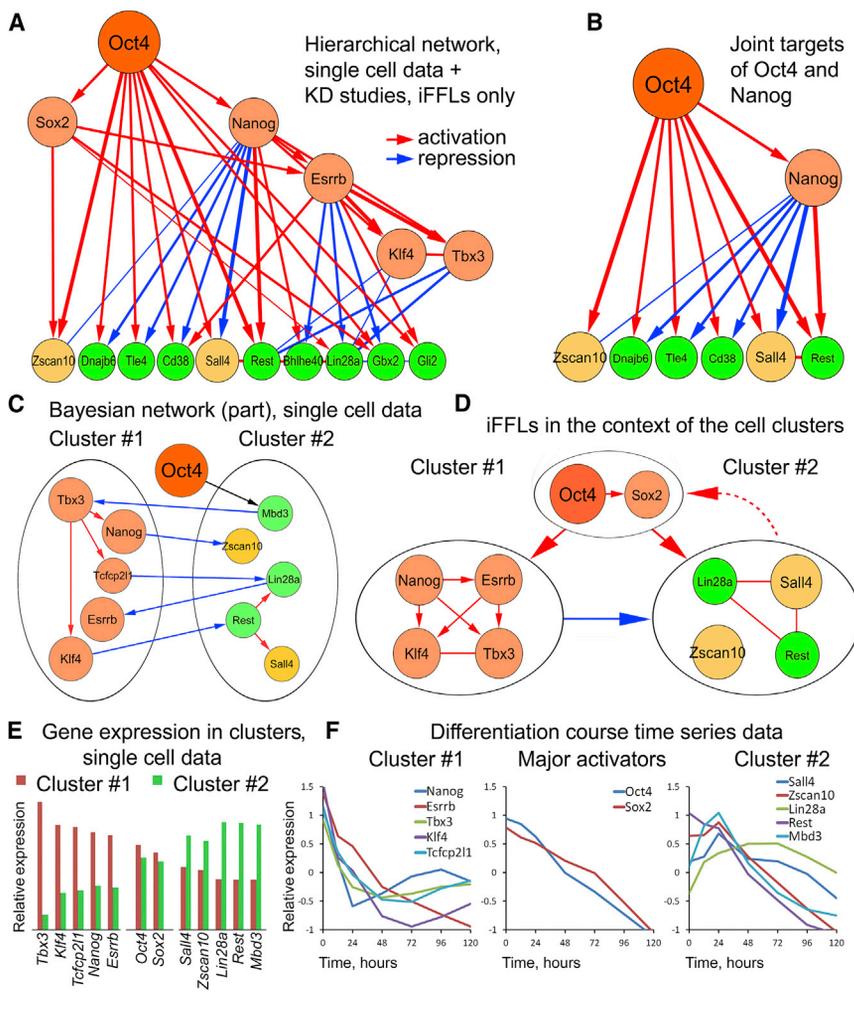
Figure 4. Regulatory Gene Network and Feedback Control of Pluripotency

(A) Part of the hierarchical gene network, including all identified iFFLs.

(B) Connectivity pattern emerging from iFFL analyses. Oct4 activates genes repressed by Nanog.

(C) Antagonistic relationships between gene clusters 1 and 2. Edge directions were predicted using Bayesian inference (Figure S4). Signs (activation/repression) were assigned based on coexpression clustering.

(D) Generalized representation of the network shown in (C) with feedback control of *Oct4* by cluster 2 factors such as SALL4.

(E) Gene expression signature (as in Figure 1G) for the genes shown in (C) and (D).

(F) Differentiation expression profiles of the genes shown in (C) and (D) (integrated data).

NANOG and their shared target genes. Strikingly, in many such motifs, OCT4 acts as an activator and NANOG as a repressor (Figure 4B), suggesting a high significance for OCT4-NANOG antagonism. A number of shared OCT4-NANOG target genes were found in gene cluster 2 (Figure 4C). A few of these (*Sall4* and *Zscan10*) have been suggested as *Oct4* regulators (Tan et al., 2013; Yang et al., 2010; Yu et al., 2009), providing a feedback-based control system to the PGRN (Figure 4D).

In Vivo Binding Data Suggest the Existence of an OCT4-SOX2/NANOG Composite Element

To identify a potential link between OCT4 and NANOG, available in vivo binding data (chromatin immunoprecipitation [ChIP]) for OCT4, SOX2, and NANOG and their binding motifs were retrieved and analyzed (Figure S5). Comparison of all published binding motifs revealed a secondary motif component present adjacent to the core of the published OCT4, NANOG, and SOX2 motifs (Figures 5A–5F). In the case of the OCT4 and SOX2 motifs, the sec-

ondary motif indicates the presence of a well-known OCT4-SOX2 element (Yuan et al., 1995). Most of the published NANOG motifs performed very poorly in the computational validation test (Figure 5D), with the exception of two motifs (shown in the figure as CHEN2008 and ChIPMunk) that contained a secondary motif component resembling the consensus OCT4 binding site.

De novo motif reconstruction based on ChIP data from multiple sources (ChIPMunk; Table S5; Kulakovskiy et al., 2013a) produced a similar bipartite consensus sequence for all three binding motifs, suggesting the presence of a composite OCT4-SOX2/NANOG element. In this element, SOX2 and Nanog binding sites overlap (Figure 5D), suggesting competition between Nanog and the OCT4-SOX2 complex. The competition may prevent activation of the target genes by OCT4 and appear as their repression by Nanog (Figure 5G). Distribution of the composite element in the *Oct4*, *Nanog*, *Sall4*, and *Zscan10* loci is shown in Figures S6A–S6D.

Distributions of the OCT4-SOX2/NANOG element and ChIP-seq peaks were analyzed in the loci of genes highly
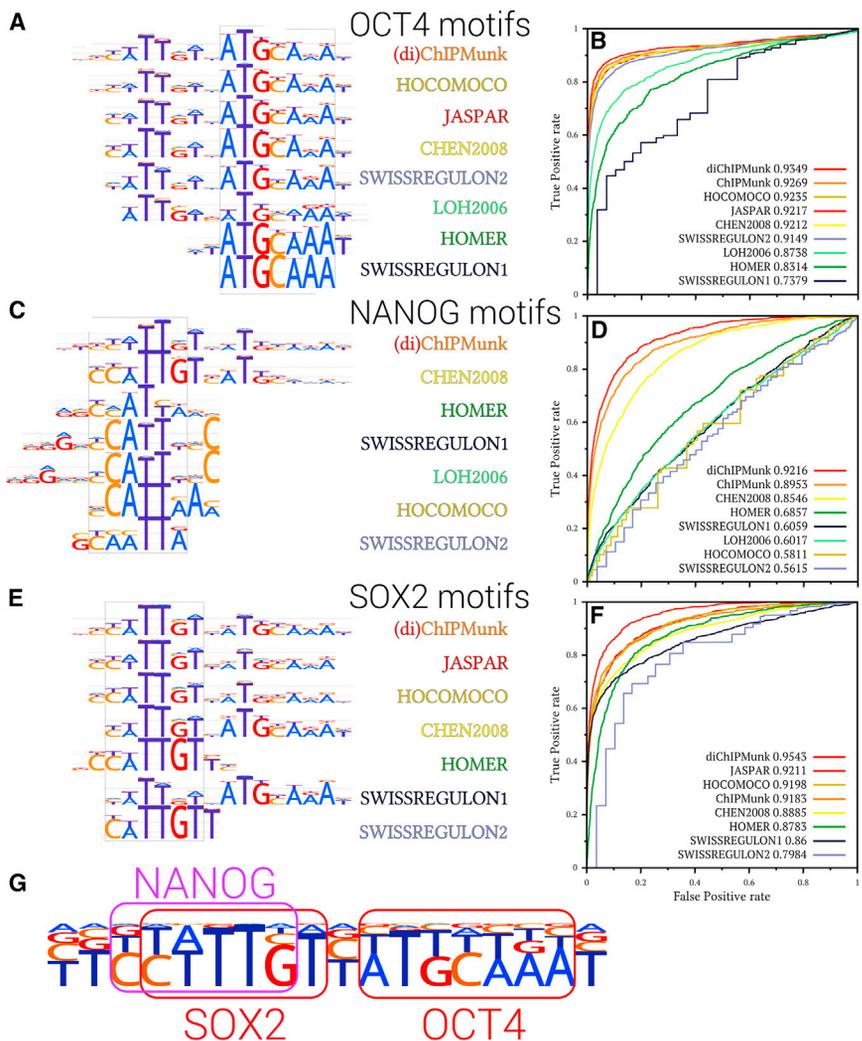
Figure 5. Composition of OCT4-SOX2 and OCT4-NANOG Elements

(A, C, and E) Survey of existing binding motif models for OCT4 (A), NANOG (C), and SOX2 (E) identified from ChIP data. In all cases, a flanking motif resembling the consensus binding sequence of another regulator is extracted along with the main motif. In the case of NANOG (C), and SOX2 (E) the flanking motif resembles an OCT4 motif, suggesting a tight link between the OCT4 and NANOG motifs.

(B, D, and F) The (ROC curves for OCT4 (B), NANOG (D), and SOX2 (F) comparing the performance of the corresponding published binding motif models for OCT4, SOX2, and NANOG shown in (A), (C), and (E). Strikingly, existing NANOG motifs without the flanking Oct4 sequences poorly recognize the original ChIP data. For every published binding motif version, it was required that the motif fit the available ChIP data from all sources.

(G) A proposed consensus for an OCT4-SOX2/NANOG composite element. The NANOG site overlaps with the SOX2 site, which may prevent synergistic activation of shared OCT4-SOX2 targets via competitive inhibition.

represented in gene clusters 1 or 2. A significantly larger number (p < 0.05, Fisher's exact test) of OCT4-SOX2/NANOG elements were found in cluster 2 genes (Table S6), supporting their alternative regulation by OCT4 and NANOG as established based on expression studies (Figures 4A and 4B).

Identified Network Motifs Suggest a Mechanism for Stabilization of OCT4 Concentration

The identified iFFL feedback (iFFL-FB) motifs were explored using quantitative models based on transcriptional interactions (Papatsenko and Levine, 2011). iFFLs without feedback are essential for threshold responses of target genes (Goentoro et al., 2009 Figure 6A). The addition of a positive feedback from the target gene to the upstream regulator creates a dynamic system in which the prevalent non-zero solution is monostability (Figure 6B). Specific properties of the iFFL-FB motif include self-compensation, when proportional scaling of regulatory parameters (Figure 6; Supplemental Experimental Procedures) has a small effect on the steady-state concentration of the upstream regulator (Figure 6C). Indeed, proportional increases in both the activation and the repression strengths in iFFL-FB (orders of magnitude) compensate each other, and the steady-state concentration of OCT4 remains nearly the same (Figure 6C). Changing environmental and intrinsic conditions (temperature, stress, starvation, and epigenetic changes) may affect global system parameters such as the rate of protein synthesis. However, in the context of the iFFL-FB model, the concentration of Oct4 will tolerate changing global conditions.

PCA of OCT4 and SALL4 expression levels in single cells revealed a bimodal dependence between the factors, in which OCT4 and SALL4 levels are correlated positively over a broad range of OCT4 concentrations but correlated negatively at high OCT4 concentrations (Figures 6D and 6E). This observed bimodal dependence between OCT4 and SALL4 is consistent with the model predictions (compare Figures 6F and 6G).
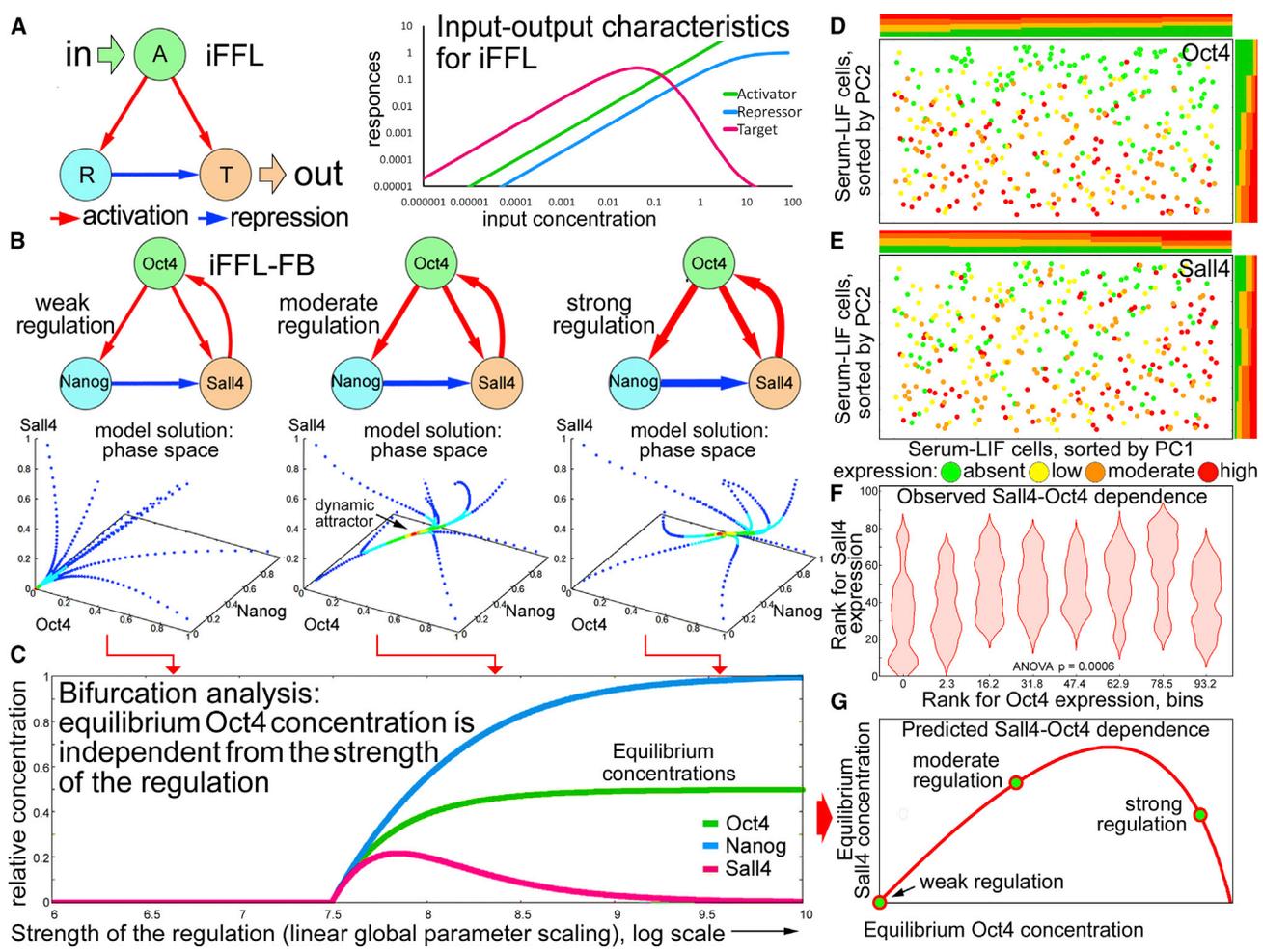
Figure 6. Properties of Incoherent Feedforward Loops with a Feedback Interaction

(A) Input/output characteristics of iFFL without a feedback. Maximal expression of the target is achieved at an intermediate level of input.

(B) iFFL-FB network motifs and phase spaces demonstrating monostable solutions. Edge thickness corresponds to parameter values or regulatory strength.

(C) Bifurcation analyses of the solutions shown in (B). Proportional scaling of all regulatory parameters (binding constants) has little effect on the equilibrium OCT4 concentration for log(Ka) > 8.

(D and E) Interdependence between OCT4 (D) and SALL4 (E) expression levels in single ESCs detected by PCA. There is a negative correlation between OCT4 and SALL4 expression in single cells along PC1 (see the histograms at the top) and a positive correlation between OCT4 and SALL4 expression along PC2 (see the histograms on the sides).
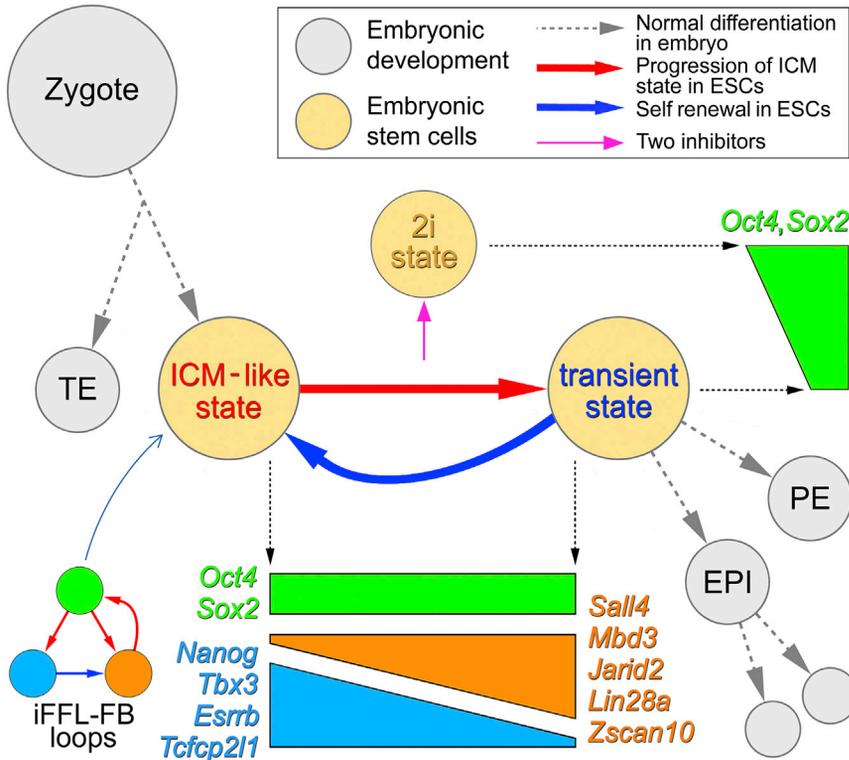
(F) Ranking of cells based on OCT4 and SALL4 concentrations and statistical analysis demonstrates relationships between OCT4 and SALL4 detected using PCA (D and E).

(G) Dependence between the equilibrium concentrations of OCT4 and SALL4 calculated from bifurcation data shown in (C). This dependence conforms to the bell-shaped curve as well as gene expression in single cells (compare F and G).

Modifications to the original iFFL-FB model have been considered, including high cooperativity models (Figure S7A), NANOG self-repression models (Figure S7B), and OCT4 dual regulation models (Figure S7C). The OCT4 dual regulation models (Marikawa et al., 2011; Papatsenko and Levine, 2008) produce very robust bistable solutions, sustaining 4-fold proportional scaling of regulatory parameters (Figures S7C–S7L, see also Movie S1).

DISCUSSION

Major Cell States Defining ESCs and Self-Renewal

Analyses of single-cell gene expression data in mESC populations have demonstrated the presence of two major subpopulations or states in cells grown in the presence of serum + LIF (LIF) and a single subpopulation in cells grown under serum-free 2i + LIF conditions (2i) (Figure 1). Cell

Relative positions of the detected ICM and transient pluripotent states are shown in the context of mouse embryo development. The color bars below display the relative expression levels of uniformly (OCT4 and SOX2) and differentially (TBX3 and NANOG) expressed genes between the two pluripotent states. The color bar on the right demonstrates higher absolute expression levels of core pluripotency factors in the 2i state. Self-compensatory network motifs, such as the one at the bottom left (iFFL-FB), may be responsible for the stabilization of ICM or other pluripotent states. The transient state is more mature than the ICM state, and it is more similar to epiblasts. ESC self-renewal may correspond to a dedifferentiation (blue arrow), which occurs when differentiation is stalled and the cells begin to roll back from the transient state to the ICM state. Such a dynamic exchange between the two states may ensure maintenance of pluripotent mESCs.

identities in the two LIF subpopulations were established using published data (Hailesellasse Sene et al., 2007; Lu et al., 2009; Tang et al., 2010). Cells in one subpopulation matched, by their expression profiles, cells from the E2.5 morula/E3.5 ICM, and cells in the other subpopulation matched epiblast cells (E5.5; Figure 2).

Earlier studies of NANOG-GFP-expressing mESC lines have demonstrated the restoration of a subpopulation structure from isolated GFP high and low fractions, suggesting a dynamic equilibrium between the two cell states (Chambers et al., 2007). A possible reverse exchange between alternative ESC states has been also suggested based on the expression of differentiation markers (Canham et al., 2010; Morgani et al., 2013). Taken together, these observations suggest a view of LIF-supported self-renewal as "stalled" differentiation (Figure 7). If the progress of normal differentiation (as observed in the embryo) is blocked or stalled at the onset of differentiation, then the cell identities "revert" back to an ICM-like state.

Several genes were found outside of the identified clusters (states). For example, OCT4 and SOX2 have similar expression levels in both states, potentially suggesting their general maintenance functions in self-renewal. The expression of differentiation markers was also similar (Figure S3D), suggesting that both LIF states are pluripotent. Conversely, STAT3 and TCF3 were highly expressed in cells occupying an intermediate or transitional position be-

tween the two major LIF states (Figure 2). Both of these transcription factors mediate signaling pathways, emphasizing the role of signaling in cell fate transitions between the two LIF subpopulations.

Molecular Signatures of ICM-like and Transient Pluripotent States

The presence of an ICM-like state in ESCs grown under LIF conditions is expected given the origin of ESCs. Our current studies show that the major determinants of the ICM state are TBX3, KLF4, NANOG, ESRRB, and other factors associated with pluripotency (Figure 1; Figure S3). TBX3 shows the greatest differences, with much higher levels in the ICM-like state. Both Tbx3 and KLF4 are activated by LIF, suggesting that LIF may be the reason for the formation of the two ES subpopulations. Interestingly, pulsed expression of KLF2 and NANOG promotes conversion of epiblast-like hESCs into a more naive ground state, which may be analogous to the ICM-like mESC state (Takashima et al., 2014). Epigenetic modifiers promoting pluripotency, such as WDR5 (Ang et al., 2011) and others, were also highly expressed in the ICM-like state (Table S3).

The properties of the transient LIF state (E5.5) may provide a clue regarding the nature of self-renewal. Addition of LIF to ESCs promotes JAK-STAT3 and other pathways, which, in turn, activate transcription factors such as TBX3, KLF4, TCFCP2l1, and others (Chen et al., 2008).

This process may slow down the natural progression of the ICM-like state toward differentiation and promote the formation of the transient or stalled pluripotent state. The transient state is characterized by increased expression of many epigenetic factors, such as MBD3, JARID2, or CTCF. Increased levels of these molecules suggest that the transient state is a turning point toward differentiation in which the cells are inherently ready for silencing their genomes and losing pluripotency. However, LIF or 2i conditions force the expression of the core pluripotency factors. Therefore, overriding the natural progression and "rolling back" the cell fates to the ICM-like state. Genome-wide predictions (Figure 3) produced a longer list of genes highly expressed in the transient pluripotent state. Some of these, such as the TCEA3/LEFTY1/NODAL group, are involved in blocking differentiation rather than promoting pluripotency.

The above cases suggest the existence of molecular mechanisms or checkpoints demarcating a transition toward differentiation. The transient state, with its high expression levels of differentiation-blocking genes, may be one such checkpoint. What controls "crossing the border" events, and how do the switches work? Reconstruction of gene networks has pointed to feedback control mechanisms involving downstream OCT4 targets such as *Sall4* or *Zscan10*, which are also highly expressed in the transient pluripotent state.

SALL4 May Be among the Key Self-Renewal Factors

Our network reconstructions (Figure 4) revealed an interesting connectivity pattern in which OCT4 and NANOG alternatively regulate genes highly expressed in the transient state (cluster 2). The literature suggests that at least two genes from this group, *Sall4* and *Zscan10*, are not only regulated by the core factors but also provide feedback control to the system by regulating *Oct4* (Yang et al., 2010; Yu et al., 2009; Figure S6). Feedback control in dynamic systems is required to achieve equilibria or steady states under changing (initial) conditions.

Several recent studies suggest that the OCT4-NANOG-SALL4 network regulates pluripotency via a feedback control mechanism (Tan et al., 2013; Yang et al., 2010). Systematic knockdown studies of 100 transcription factors (KD-100) in mESCs place SALL4 in fourth position (after ESRRB, OCT4, and SOX2) based on the number of downstream targets (Nishiyama et al., 2013). Known SALL4 functions include blocking trophectodermal commitment (Oron and Ivanova, 2012), regulating extraembryonic endodermal genes (Lim et al., 2008), stimulating human hematopoietic stem cells (Yang et al., 2011), and many others.

Our current analyses of in vivo binding data for core pluripotency factors suggest the presence of a composite DNA element combining OCT4/SOX2 and NANOG binding sites (Figure 5). Both *Sall4* and *Zscan10* loci contain this element, suggesting the direct alternative regulation of these two genes by Oct4 and Nanog, which is also supported by expression data. The composite element includes the known OCT4/SOX2 element, in which the part of the SOX2 site overlaps with the Nanog site. Binding of NANOG to the composite element may interfere with binding of SOX2, therefore disrupting OCT4-SOX2 synergy, which may appear as direct repression by Nanog. Quantitative modeling of the network motif (iFFL-FB) loop combining *Oct4*, *Sall4*, and *Nanog* demonstrates how the interplay of these factors may stabilize OCT4 concentration and, therefore, safeguard pluripotency.

iFFL-FB: Self-Compensatory Network Motifs that Stabilize Pluripotency

The identified iFFL-FB motifs and their variants were explored using quantitative models (Figure 6). One of the most frequent solutions for such motifs is monostability. For certain modifications of iFFL-FB motifs, bistable solutions are frequent but not robust to parameter variations. Instead, monostable solutions display a remarkable robustness: the equilibrium concentration of OCT4 in these solutions sustained proportional scaling of regulatory parameters over orders of magnitude. Indeed, proportionally increasing both activation and repression rates in the context of the iFFL-FB compensates for the outcome and maintains a stable equilibrium concentration of OCT4.

In terms of biology, the identified self-compensatory feature of the iFFL-FB motif suggests that the equilibrium concentration of OCT4 may be resistant to global changes in cell metabolic and biochemical reaction rates caused by changing temperature, stress, starvation, as well as global chromatin configuration. The models described in this study (Figures 6B and 6C) predict that, under such changing conditions, the concentration of OCT4 would remain constant, therefore safeguarding or buffering pluripotency. Interestingly, the major impact of global parametric changes appears to be "absorbed" by SALL4. Therefore, SALL4 acts as a buffering component of the system, whereas OCT4 is the buffered component. NANOG, in this configuration, mediates the negative feedback in the context of OCT4-NANOG-SALL4 iFFL-FB. The measured distribution of OCT4 and SALL4 expression levels in single cells is consistent with the proposed model, assuming changing metabolic rates during mESC transitions from one state to another (Figures 6D–6G).

Although iFFL-FB models predict monostable solutions, cells grown under LIF conditions form at least two subpopulations. Is there a contradiction between the data and the model? Apparently, LIF media may be considered an "enforced" condition where the stability of the pluripotent

states and self-renewal are maintained artificially. Instead, the 2i condition, producing a single subpopulation of cells, may represent a more natural, self-sustainable PGRN steady state in which most of the signaling is bypassed by the chemical inhibitors. In this case, a single predicted state is in agreement with the single detected cell subpopulation. Under enforced LIF conditions (two subpopulations), two iFFL-FB loops may be involved in stabilizing either the ICM-like or transient state or, possibly, both. Moreover, switching between states may be explained not only by bistable solutions (Figure S7C), but may be the consequence of more sophisticated processes in which the regulatory parameters undergo changes and generate new dynamic solutions (bifurcation).

In summary, we identified and characterized two cell subpopulations present in mESC, proposed a molecular mechanism explaining the antagonistic actions of Oct4 and Nanog in the pluripotency gene regulatory network, developed quantitative models explaining the stabilization of Oct4 concentration as well as pluripotency states, and proposed a "state exchange model" to explain self-renewal.

EXPERIMENTAL PROCEDURES

mESC Culture and Data Collection

CCE mESCs were cultured on gelatin-coated plates in high-glucose DMEM supplemented with 15% fetal bovine serum, 100 mM non-essential amino acids, 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, 100 U/ml penicillin/streptomycin, 1 mM sodium pyruvate, and 1000 U/ml LIF. When transferring to 2i serum-free conditions, cells were seeded on gelatin-coated plates in 2i medium composed of a 50/50 mixture of Neurobasal and DMEM/F12 media, N2 and B27 supplements, 0.05% BSA, 100 U/ml penicillin/streptomycin, 1 mM L-glutamine, 1,000 U/ml LIF, 3 μM CHIR99021, and 1 μM PD0325901. For single-cell gene expression analyses, 4′,6-diamidino-2-phenylindole (DAPI)-stained live cells were sorted. Cells were sorted into 96-well PCR plates containing pre-amplification mix, Taqman probes, Superscript III RT/Platinum Taq mix, and CellsDirect reaction mix (Invitrogen). Reverse transcription, pre-amplification, and amplification were performed as follows: 50°C for 15 min and 95°C for 2 min, followed by 21 cycles of 95°C for 15 s and 60°C for 4 min. A well containing no sorted cells served as a negative control, and a well with 100 cells served as a positive control for the population state. RT and pre-amplification products were diluted and run on a Biomark Fluidigm machine following the manufacturer's instructions using 96.96 dynamic arrays. Each probe set was run in duplicate.

Single-Cell Data Filtering and Normalization

Single-cell gene expression data were first filtered to remove cells with multiple failed probes. The threshold (<60% of failed measurements) was established based on the dependence of marker concentration on the fraction of failed cells (Figure S2E). This procedure returned a total of 552 cells for seven Fluidigm chips, including 375 cells for LIF (5 × 96 cells initially) and 177 cells for 2i (2 × 96 cells initially) conditions. The single-cell data were processed using relative normalization. First, data from each individual cell were normalized using the linear zero-mean method, and technical replicates were averaged. Next the data for each gene from each chip were normalized in the same way, therefore eliminating chip-to-chip variations (Figure S2). Finally the normalized data were integrated for multiple chips (either for five LIF chips or for two 2i chips) and ranked. The ranks were converted to a 0–100 range, with failed readings assigned a value of 0. All published single-cell data used in the comparisons were rank-normalized using the same procedure to match the data format. Comparisons between the raw cycle threshold (CT) values and the normalized data are given in Figures S2F and S2G. Raw and normalized single-cell data produced in this study are available in Table S1.

Clustering Cell Gene Expression Profiles

Clustering of data with respect to cell expression profiles was performed using three different methods: hierarchical clustering (de Hoon et al., 2004), co-expression clustering (Stuart et al., 2003), and PCA (de Hoon et al., 2004). In the case of co-expression clustering, failed measurements were ignored, and the corresponding probabilistic similarity matrix was constructed by calculating correlation p values given the number of non-failed readings in all pairwise comparisons. The p values were corrected using the Benjamini-Hochberg technique to establish an appropriate similarity cutoff (p < 0.01 was adopted). Visualization of the resulting co-expression clusters was performed with force-directed layout using the Cytoscape program (Figures 1A–1D; Smoot et al., 2011).

Gene Network Reconstruction

Gene networks were reconstructed using the following methods: co-expression clustering (Stuart et al., 2003) of single-cell data using an approach described for cells but applied to genes (clustering columns versus rows), Bayesian network inference using Banjo software (Hartemink, 2005), mutual information analysis to detect non-linear interactions between genes, and reconstruction based on genome-wide knockdown studies compiled from multiple sources (Martello et al., 2012) (Supplemental Experimental Procedures; Table S4).

Analysis of Published ChIP Datasets

Available ChIP data and existing binding motif models (positional weight matrices [PWMs]) for OCT4, SOX2, and NANOG were collected from available sources (Chen et al., 2008; Heinz et al., 2010; Kulakovskiy et al., 2013b; Loh et al., 2006; Portales-Casamar et al., 2010; Figure S5; Table S5). The ChIPMunk for classic PWMs (Weirauch et al., 2013) and diChIPMunk (Kulakovskiy et al., 2013a; Levitsky et al., 2014) programs were used for de novo motif discovery. Raw published ChIP data were processed, and receiver operating characteristic (ROC) curves were constructed as described earlier (Kulakovskiy et al., 2013a). Sequences for motif discovery were extracted with an offset of ±150 bp with respect to the peak summits. Transcription factor binding site (TFBS) predictions were produced using diChIPMunk dinucleotide PWMs for the OCT4/SOX2/NANOG subtypes of the SOX2/OCT4 composite element.

Quantitative Modeling

Quantitative models for network motifs were constructed using ordinary and stochastic differential equations describing interactions between genes in transcriptional networks (Supplemental Experimental Procedures). Model solutions were analyzed using the NetExplore web server (http://line.bioinfolab.net/nex/netexplore.htm) (Papatsenko and Lemischka, 2015). At least 10^4 solutions were analyzed. The explored additive and cooperative models have four open edge-specific parameters (each specific to one of the four regulatory links) and three global parameters (common to all nodes). For all node/edge absolute concentrations, cooperativity levels and synthesis rates were identical. The NANOG self-repression model has five open parameters, and the OCT4 dual regulation model has six edge-specific open parameters plus three node-specific parameters for cooperative interactions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, six tables and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.07.004.

REFERENCES

Abdelalim, E.M., and Tooyama, I. (2014). Knockdown of p53 suppresses Nanog expression in embryonic stem cells. Biochem. Biophys. Res. Commun. 443, 652–657.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell 145, 183–197.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. 16, 516–528.

Canham, M.A., Sharov, A.A., Ko, M.S., and Brickman, J.M. (2010). Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol. 8, e1000379.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230–1234.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106–1117.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics 20, 1453–1454.

Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J., et al. (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat. Cell Biol. 11, 197–203.

Festuccia, N., Osorno, R., Wilson, V., and Chambers, I. (2013). The role of pluripotency gene regulatory network components in mediating transitions between pluripotent cell states. Curr. Opin. Genet. Dev. 23, 504–511.

Goentoro, L., Shoval, O., Kirschner, M.W., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol. Cell 36, 894–899.

Hailesellasse Sene, K., Porter, C.J., Palidwor, G., Perez-Iratxeta, C., Muro, E.M., Campbell, P.A., Rudnicki, M.A., and Andrade-Navarro, M.A. (2007). Gene function in early mouse embryonic stem cell differentiation. BMC Genomics 8, 85.

Hartemink, A.J. (2005). Reverse engineering gene regulatory networks. Nat. Biotechnol. 23, 554–555.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533–538.

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273–1282.

Kulakovskiy, I., Levitsky, V., Oshchepkov, D., Bryzgalov, L., Vorontsov, I., and Makeev, V. (2013a). From binding motifs in ChIP-Seq data to improved models of transcription factor binding sites. J. Bioinform. Comput. Biol. 11, 1340004.

Kulakovskiy, I.V., Medvedeva, Y.A., Schaefer, U., Kasianov, A.S., Vorontsov, I.E., Bajic, V.B., and Makeev, V.J. (2013b). HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Res. 41, D195–D202.

Kumar, R.M., Cahan, P., Shalek, A.K., Satija, R., DaleyKeyser, A.J., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J.J., et al.

(2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature *516*, 56–61.

Levitsky, V.G., Kulakovskiy, I.V., Ershov, N.I., Oshchepkov, D.Y., Makeev, V.J., Hodgman, T.C., and Merkulova, T.I. (2014). Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data. BMC Genomics *15*, 80.

Lim, C.Y., Tam, W.L., Zhang, J., Ang, H.S., Jia, H., Lipovich, L., Ng, H.H., Wei, C.L., Sung, W.K., Robson, P., et al. (2008). Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. Cell Stem Cell *3*, 543–554.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Lu, R., Markowetz, F., Unwin, R.D., Leek, J.T., Airoldi, E.M., MacArthur, B.D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L.A., et al. (2009). Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature *462*, 358–362.

Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. Nat. Struct. Mol. Biol. *18*, 120–127.

Marikawa, Y., Tamashiro, D.A., Fujita, T.C., and Alarcon, V.B. (2011). Dual roles of Oct4 in the maintenance of mouse P19 embryonal carcinoma cells: as negative regulator of Wnt/β-catenin signaling and competence provider for Brachyury induction. Stem Cells Dev. *20*, 621–633.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491–504.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824–827.

Morgani, S.M., Canham, M.A., Nichols, J., Sharov, A.A., Migueles, R.P., Ko, M.S., and Brickman, J.M. (2013). Totipotent embryonic stem cells arise in ground-state culture conditions. Cell Rep. *3*, 1945–1957.

Nishiyama, A., Sharov, A.A., Piao, Y., Amano, M., Amano, T., Hoang, H.G., Binder, B.Y., Tapnio, R., Bassey, U., Malinou, J.N., et al. (2013). Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. Sci. Rep. *3*, 1390.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

Oron, E., and Ivanova, N. (2012). Cell fate regulation in early mammalian development. Phys. Biol. *9*, 045002.

Papatsenko, D., and Levine, M.S. (2008). Dual regulation by the Hunchback gradient in the Drosophila embryo. Proc. Natl. Acad. Sci. USA *105*, 2901–2906.

Papatsenko, D., and Levine, M. (2011). The Drosophila gap gene network is composed of two parallel toggle switches. PloS One *6*, e21145.

Papatsenko, D., and Lemischka, I.R. (2015). NetExplore: a web server for modeling small network motifs. Bioinformatics *31*, 1860–1862.

Parisi, S., and Russo, T. (2011). Regulatory role of Klf5 in early mouse development and in embryonic stem cells. Vitam. Horm. *87*, 381–397.

Park, K.S., Cha, Y., Kim, C.H., Ahn, H.J., Kim, D., Ko, S., Kim, K.H., Chang, M.Y., Ko, J.H., Noh, Y.S., et al. (2013). Transcription elongation factor Tcea3 regulates the pluripotent differentiation potential of mouse embryonic stem cells via the Lefty1-Nodal-Smad2 pathway. Stem Cells *31*, 282–292.

Plusa, B., and Hadjantonakis, A.K. (2014). Embryonic stem cell identity grounded in the embryo. Nat. Cell Biol. *16*, 502–504.

Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. *38*, D105–D110.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431–432.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249–255.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

Tan, M.H., Au, K.F., Leong, D.E., Foygel, K., Wong, W.H., and Yao, M.W. (2013). An Oct4-Sall4-Nanog network controls developmental progression in the pre-implantation mouse embryo. Mol. Syst. Biol. *9*, 632.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat. Protoc. *5*, 516–535.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909–918.

Weirauch, M.T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T.R., Saez-Rodriguez, J., Cokelaer, T., Vedenko, A., Talukder, S., et al.; DREAM5 Consortium (2013). Evaluation of methods for modeling transcription factor sequence specificity. Nat. Biotechnol. *31*, 126–134.

White, A.K., VanInsberghe, M., Petriv, O.I., Hamidi, M., Sikorski, D., Marra, M.A., Piret, J., Aparicio, S., and Hansen, C.L. (2011). High-throughput microfluidic single-cell RT-qPCR. Proc. Natl. Acad. Sci. USA *108*, 13999–14004.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

Yang, J., Gao, C., Chai, L., and Ma, Y. (2010). A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. PLoS ONE *5*, e10766.

Yang, J., Aguila, J.R., Alipio, Z., Lai, R., Fink, L.M., and Ma, Y. (2011). Enhanced self-renewal of hematopoietic stem/progenitor cells mediated by the stem cell gene Sall4. J. Hematol. Oncol. *4*, 38.

Yu, H.B., Kunarso, G., Hong, F.H., and Stanton, L.W. (2009). Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. J. Biol. Chem. *284*, 31327–31335.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. *9*, 2635–2645.