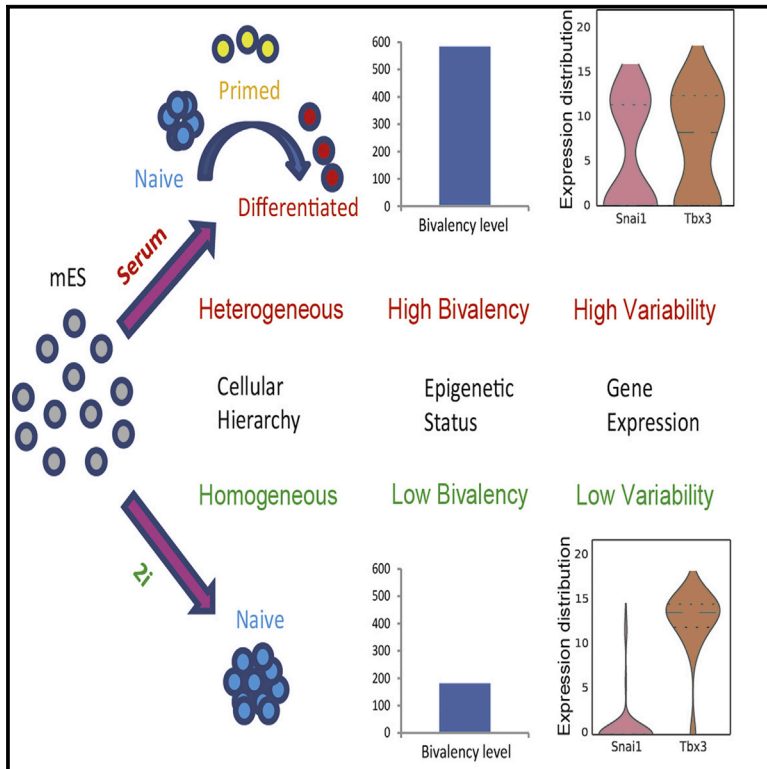


Cell Reports

Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis

Graphical Abstract



Authors

Guoji Guo, Luca Pinello, Xiaoping Han, ..., Keyong Zou, Guo-Cheng Yuan, Stuart H. Orkin

Correspondence

ggj@zju.edu.cn (G.G.),
gcyuan@jimmy.harvard.edu (G.-C.Y.),
stuart_orkin@dfci.harvard.edu (S.H.O.)

In Brief

Guo et al. investigate mechanisms of gene expression variation and cellular heterogeneity in mouse embryonic stem cells using single-cell mRNA-seq analysis. They find a differentiation-priming pathway in the cell culture system and show that serum provokes gene expression variability in mouse embryonic stem cells.

Highlights

- Single-cell mRNA-seq analysis reveals heterogeneity in ESCs cultured in serum
- Gene expression variability is associated with distinct chromatin characteristics
- Computational analysis identifies an ESC-priming pathway
- External culture system affects gene expression variability of ESCs

Accession Numbers

GSE75804



Guo et al., 2016, Cell Reports 14, 956–965
February 2, 2016 ©2016 The Authors
<http://dx.doi.org/10.1016/j.celrep.2015.12.089>

CellPress

Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis

Guoji Guo,^{1,3,8,*} Luca Pinello,^{2,8} Xiaoping Han,^{1,3,8} Shujing Lai,³ Li Shen,⁷ Ta-Wei Lin,⁵ Keyong Zou,⁶ Guo-Cheng Yuan,^{2,*} and Stuart H. Orkin^{1,4,*}

¹Division of Pediatric Hematology/Oncology, Dana Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

²Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

³Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou 310058, China

⁴Howard Hughes Medical Institute, Boston, MA 02115, USA

⁵Molecular Genetics Core Facility, Boston Children's Hospital, Boston, MA 02115, USA

⁶Boston Open Labs, Cambridge, MA 02138, USA

⁷Life Sciences Institute, Zhejiang University School of Medicine, Hangzhou 310058, China

⁸Co-first author

*Correspondence: ggj@zju.edu.cn (G.G.), gcyuan@jimmy.harvard.edu (G.-C.Y.), stuart_orkin@dfci.harvard.edu (S.H.O.)

<http://dx.doi.org/10.1016/j.celrep.2015.12.089>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

SUMMARY

Variation in gene expression is an important feature of mouse embryonic stem cells (ESCs). However, the mechanisms responsible for global gene expression variation in ESCs are not fully understood. We performed single-cell mRNA-seq analysis of mouse ESCs and uncovered significant heterogeneity in ESCs cultured in serum. We define highly variable gene clusters with distinct chromatin states and show that bivalent genes are prone to expression variation. At the same time, we identify an ESC-priming pathway that initiates the exit from the naive ESC state. Finally, we provide evidence that a large proportion of intracellular network variability is due to the extracellular culture environment. Serum-free culture reduces cellular heterogeneity and transcriptome variation in ESCs.

INTRODUCTION

Early mammalian development cells differentiate toward trophoblast (TE) and inner cell mass (ICM). The ICM goes on to form the epiblast (EPI) and the primitive endoderm (PE). Embryonic stem cells (ESCs) can be derived from the ICM in the presence of leukemia inhibitory factor (LIF) and fetal calf serum (FCS) (Evans and Kaufman, 1981). ESCs have two important characteristics: the capacity for differentiation into all somatic cell types and the property of unlimited self-renewal in vitro.

Previous studies suggest that ESCs in culture are not homogeneous. Transcription factors associated with ESC identity may be expressed in a heterogeneous manner. For example, Nanog and Dppa3 are expressed in only a fraction of cells (Chambers et al., 2007; Hayashi et al., 2008). Variation in expression of these individual genes has been implicated in controlling the differentiation potential of different subpopulations. However, traditional methods are limited to the analysis of small number of genes. The mechanisms underlying genome-scale ESC variability are not fully characterized.

Single-cell gene expression analysis has been developed as a powerful tool for studying cellular heterogeneity and hierarchy. Several hallmark technical advances have been achieved. High-throughput single-cell qPCR is a dynamic approach for quantifying a set of target genes in systems of interest (Buganim et al., 2012; Dalerba et al., 2011; Guo et al., 2010, 2013; Moignard et al., 2013). Single-cell mass cytometry constitutes a complementary system for multiplexed gene expression analysis at the protein level (Bendall et al., 2011). Single-cell mRNA-sequencing strategies, which enable whole-transcriptome analysis from individual cells, have become increasingly mature and capable (Fan et al., 2015; Hashimshony et al., 2012; Islam et al., 2011; Jaitin et al., 2014; Klein et al., 2015; Macosko et al., 2015; Ramsköld et al., 2012; Sasagawa et al., 2013; Shalek et al., 2013; Tang et al., 2009, 2010; Treutlein et al., 2014; Xue et al., 2013; Yan et al., 2013).

Using single-cell technologies, several studies have reported transcriptomic analysis of mouse ESCs and uncovered signaling and microRNA pathways that influence heterogeneity of ESCs in culture (Grün et al., 2014; Kumar et al., 2014). More-recent studies have also examined transcriptional networks and cell-cycle regulators that contribute to transcriptional variation (Kolodziejczyk et al., 2015; Papatsenko et al., 2015). Epigenetic

regulation, which may also contribute to overall variability, has not been adequately explored. Moreover, the relevance of ESC culture heterogeneity to early embryonic development has yet to be analyzed.

In this study, we sought to combine the power of microfluidic-based single-cell mRNA-seq and single-cell qPCR to characterize in depth the molecular basis of heterogeneity among mouse ESCs in culture. We employ optimized computational strategies to reveal epigenetic mechanisms contributing to variation in gene expression and search for upstream pathways that induce network plasticity.

RESULTS

Single-Cell mRNA-Seq Analysis Reveals Heterogeneity among Mouse ESCs in Culture

We performed single-cell mRNA-seq analysis of undifferentiated ESCs in culture. Feeder-free J1 ESCs were grown in the presence of serum and LIF. Single ESCs were captured on a medium-sized (10–17 μ m cell diameter) microfluidic RNA-seq chip (Fluidigm) using the Fluidigm C1 system (Figure 1A). Whole-transcriptome-sequencing libraries were prepared using template switching-based amplification (Figure 1B). We compared the abundance of selected markers from single-cell cDNA amplified with the template switching (SMART) method, as well as the sequence-specific amplification (SSA) method. qPCR results from different amplification products revealed comparable expression patterns for wild-type ESCs, namely high-level detection of EPI markers *Pou5f1*, *Nanog*, and *Sox2*, as well as low-level detection of TE markers, *Cdx2* and *Gata3*. From amplified cDNA, we detected a bimodal distribution for *Id2* and sharp unimodal distribution for endogenous controls, *Actb* and *Gapdh* (Figure 1C).

Amplified single-cell libraries were barcoded, pooled, and sequenced to a depth of about 1.2 million reads per sample. For each gene in a sample, the median reads per kilobase of transcript per million reads mapped (RPKM) was approximately ten (Figures S1A and S1B). In order to filter out unreliable signals, we removed genes with low copy counts and ensured an average Pearson correlation of $R = 0.8$ between two sequencing duplicates for each single-cell library (Figures S1C–S1E; see Experimental Procedures for details). Using this strategy, we recover $\sim 9,000$ genes per cell (Figure 1D). The average of single-cell mRNA-seq profiles from ESCs showed high correlation with bulk mRNA-seq profiles from the same cell line (Figure 1E). However, we observed that a fraction of the samples had distinct global signatures from the others, suggesting strong heterogeneity under the tested culture conditions (Figure S1F). Although an endogenous control gene, *Actb*, and a pluripotency gene, *Pou5f1*, were homogeneously expressed among single cells, we observed strong variation of other markers, including *Lamb1*, *Clu*, and *Snai1* in both J1 and E14 cells (Figures S1G and S1H). By examining the expression correlation of key lineage regulators in the single-cell data, we defined different gene modules that correlate with this heterogeneity (Figure 1F). The tightly correlated pluripotency markers, *Pou5f1*, *Sox2*, *Nanog*, and *Fgf4*, define a module for maintaining the undifferentiated ESC state. A *Sox17*,

Gata6, and *Gata4* cluster reflects a PE module that is indicative of PE differentiation.

Gene Expression Variability Is Associated with Distinct Chromatin States

In order to study variability of gene expression within the single-cell transcriptome data, we first tested different ways to quantify the level of variability. As variability measurements are easily influenced by mean level and amplification bias, we sought to decouple gene expression variation from the mean expression. We fitted a Lowess curve to \log_2 of the mean expression versus the \log_2 of the SD and then calculated the distance from this curve for each gene (Figure 2A). Because the distribution of this distance is approximately normal, we further rescaled the values by converting to Z scores. The resulting value, which we term the Lowess coefficient of dispersion (LCOD), is used to quantify the variation of gene expression. We show that LCOD is the least correlated or anti-correlated with the mean expression level as compared to other measurements (Figure 2B). We then selected the most- (LCOD > 1.5) and the least-variable genes (LCOD < -1.5). Gene Ontology (GO) enrichment analysis indicated that the most-variable genes are related predominantly with developmental processes, whereas the least-variable genes are enriched for translation, mRNA processing, and splicing (Figure S2A).

To investigate the mechanism underlying variability at the single-cell level, we integrated our single-cell mRNA-seq data with the genome-wide transcription factor binding and chromatin state information obtained from publicly available bulk-level ChIP-seq data sets. We mapped both transcription factor occupancy and key chromatin marks in a 10-kb window at the transcriptional start site of the most- and least-variable genes. Of note, the master pluripotency regulators, including Oct4, Sox2, and Nanog, displayed similar binding patterns between the most- and least-variable genes. However, we observed a distinct chromatin state signature associated with the most-variable genes, including enrichment of the H3K27me3 mark and Ezh2 occupancy, as well as depletion of H3K36me3. Our analysis suggests that chromatin regulators may play an important role in mediating gene expression variability at the single-cell level (Figure 2C).

We then aimed to further discriminate the list of most-variable genes using the chromatin marks found to correlate for gene expression variation. As shown in Figure 2D, we observed three distinct patterns, suggesting multiple pathways leading to fluctuations in gene expression. Cluster 1 genes were strongly enriched for H3K27me3 and Ezh2 binding and moderately enriched for H3K4me3, suggesting a role of polycomb group proteins in mediating expression variability. Cluster 2 genes were moderately enriched for H3K27me3, H3K4me3, and H3K36me3. Cluster 3 genes were enriched for H3K4me3 and H3K36me3, indicating a possible role of Setd2.

Importantly, the most-variable genes were enriched for previously defined bivalent genes marked by both H3K27me3 and H3K4me3 in their promoters (22% of the most variable versus 6% for all the genes measured in our assay; Figures 2D and S2B–S2D; Bernstein et al., 2006). We also found that overall gene expression variability was significantly higher among

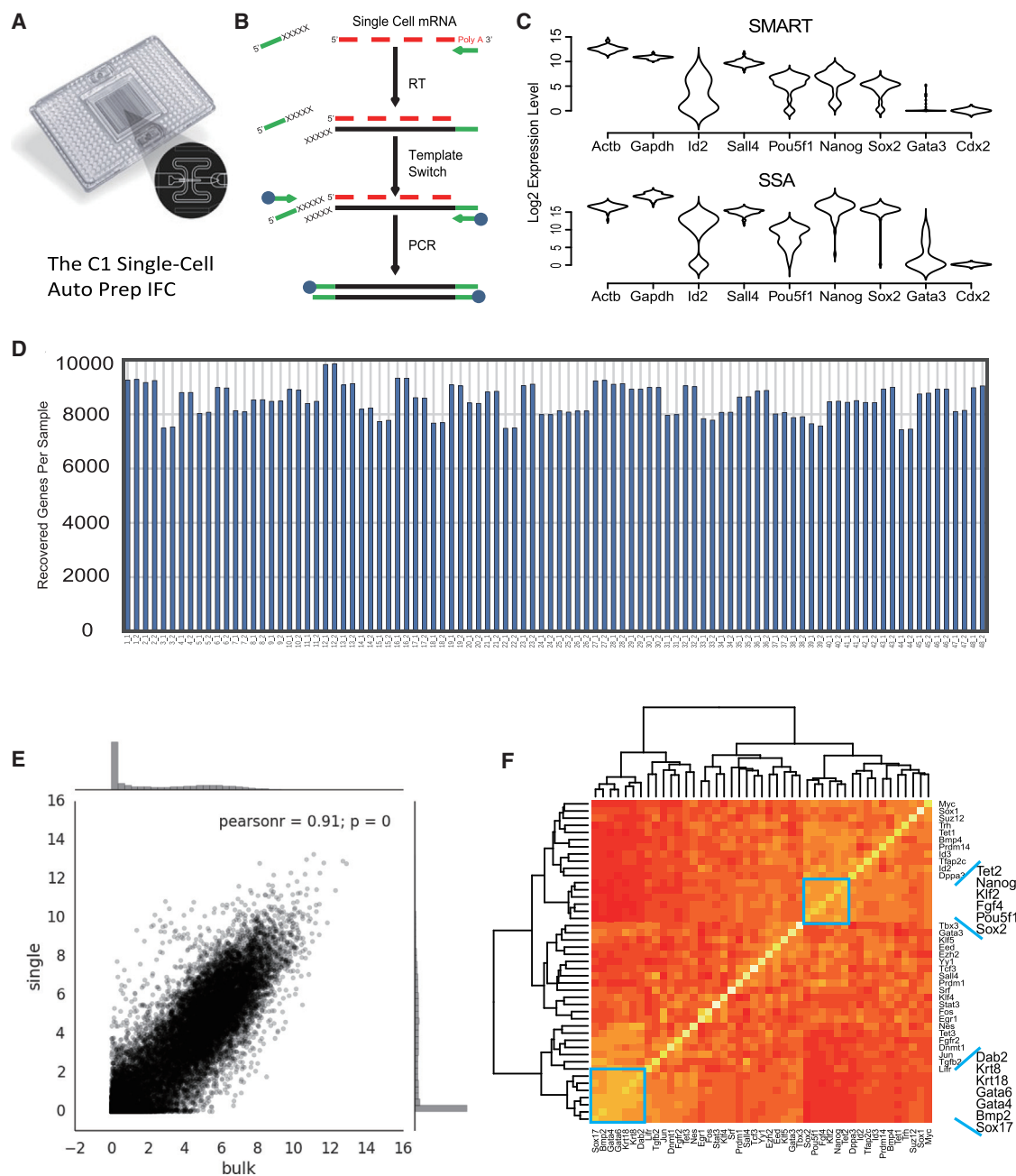


Figure 1. Single-Cell mRNA-Seq of Mouse Embryonic Stem Cells

(A) The C1 (Fluidigm) microfluidic system for single-cell capture and library generation.

(B) Protocol for the template-switch method (SMARTer Kit; Clontech) for global mRNA amplification from single cells.

(C) Comparison of results from template-switching amplification method (SMART) and sequence-specific amplification method (SSA) for single-cell mRNA quantification. Amplified single-cell cDNAs were tested by qPCR using selected gene primers. Expression level distributions are presented as violin plots.

(D) Bar chart depicts the number of expressed genes in each single-cell mRNA-sequencing samples.

(E) A scatterplot showing the correlation between J1 ESC single-cell mRNA-seq data and bulk-cell mRNA-seq data.

(F) A gene expression correlation heatmap from single-cell expression data reveals separation of different gene expression modules that reflect network heterogeneity in mouse ESC culture.

bivalent genes ($p = 1.0E-32$; KS test; Figure S2E). Whereas bivalent genes have been commonly considered to be silent in ESCs, previous studies have been limited to population level analysis.

Here, using single-cell analysis, we observed that many bivalent genes are in fact actively transcribed in a subset of cells and that the overall distribution is bimodal, suggesting that the

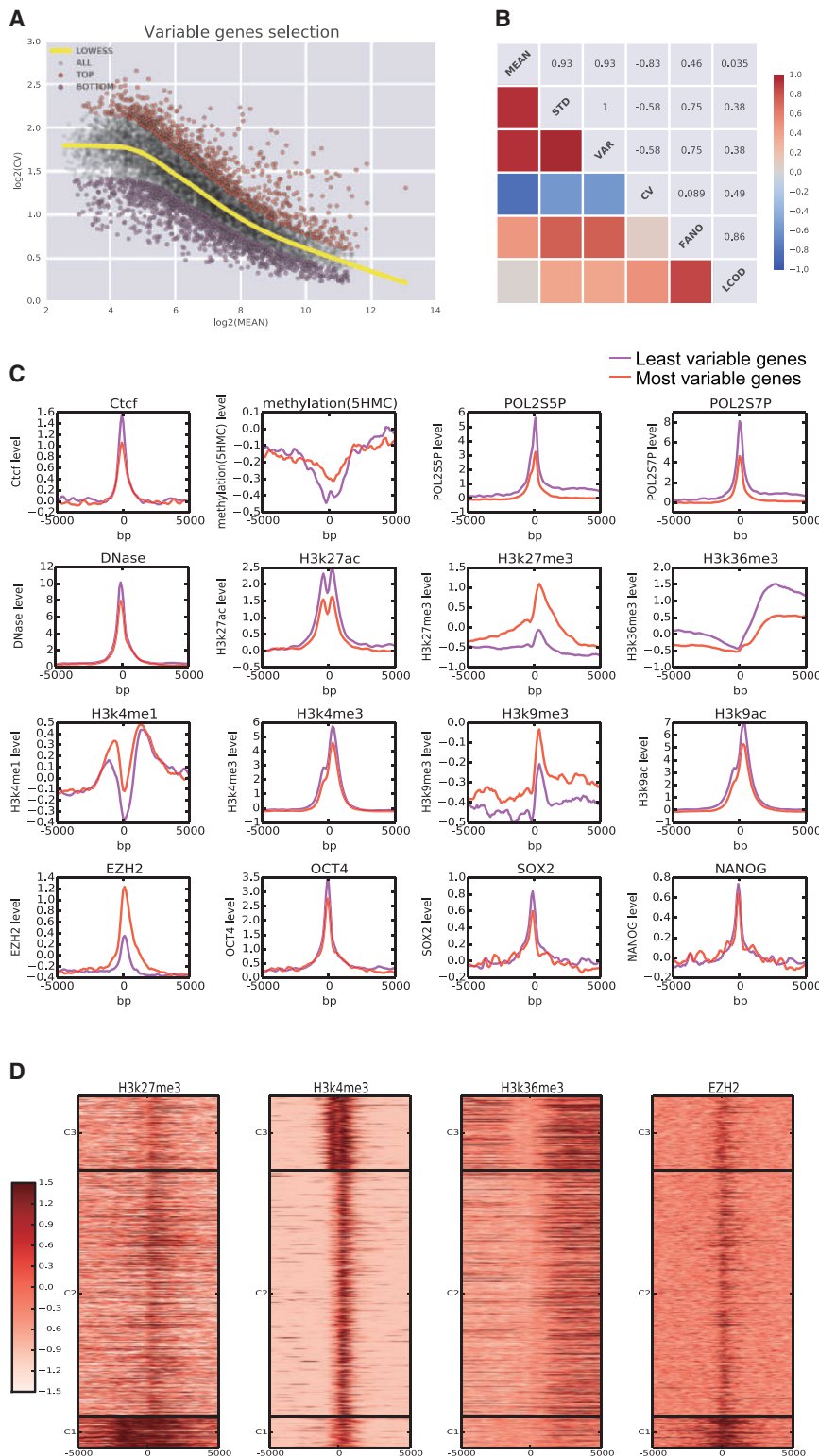


Figure 2. Distinct Chromatin States Mark Gene Expression Variability

(A) Selection of the most-variable genes (red) and the least-variable genes (purple) using Lowess coefficient of dispersion (LCOD) analysis. (B) The choice of LCOD as the criteria for measuring gene expression variability. (C) Comparison of chromatin states between the most- and least-variable genes in mouse ESC culture. (D) Analysis of selected chromatin marks on the most-variable genes reveals three clusters of genes with different characteristics.

Computational Analysis Reveals ESC Early Priming Pathway

To better understand the complex structure driven by heterogeneity in gene expression, we used locally linear embedding (LLE) dimensionality reduction analysis. LLE is an unbiased approach that computes a low dimensional representation of the data, preserving the original distances between neighborhoods points (Roweis and Saul, 2000). As seen from the LLE projection, a culture of morphologically “undifferentiated” ESCs was comprised of different subgroups (Figures S3A and S3B). The distribution of cell states suggests a defined pathway exiting pluripotency. In order to delineate this pathway more accurately, we applied a principal curve analysis and reconstructed a smooth path that passes through the cells at all stages (Figures 3A and S3C). By mapping the individual cells onto the principal curve, we identified three distinct cellular states (Figures 3A and S3C). On the left, the closely clustered population corresponds to the naive ESC state. The cells within this population express pluripotency markers (e.g., *Nanog*, *Sox2*, and *Klf2*) at high level (Figures 3A and S3C). In addition, expression of differentiation markers was not detected in this group. On the top of the curve, we defined a previously unrecognized population, consisting of “primed” cells. These ESCs simultaneously express pluripotent markers (e.g., *Sox2* and *Nanog*) and differentiation markers (e.g., *Gata4*, *Gata6*, and *Lamb1*). This population

transcriptional activities in an ESC may be highly dynamic and that the bivalent domains may play a role in modulating the frequency of gene activation.

appears to represent a transcriptionally primed cellular state in which cells are exiting the naive ESC state and under transition to a differentiated state. The third cluster of cells express

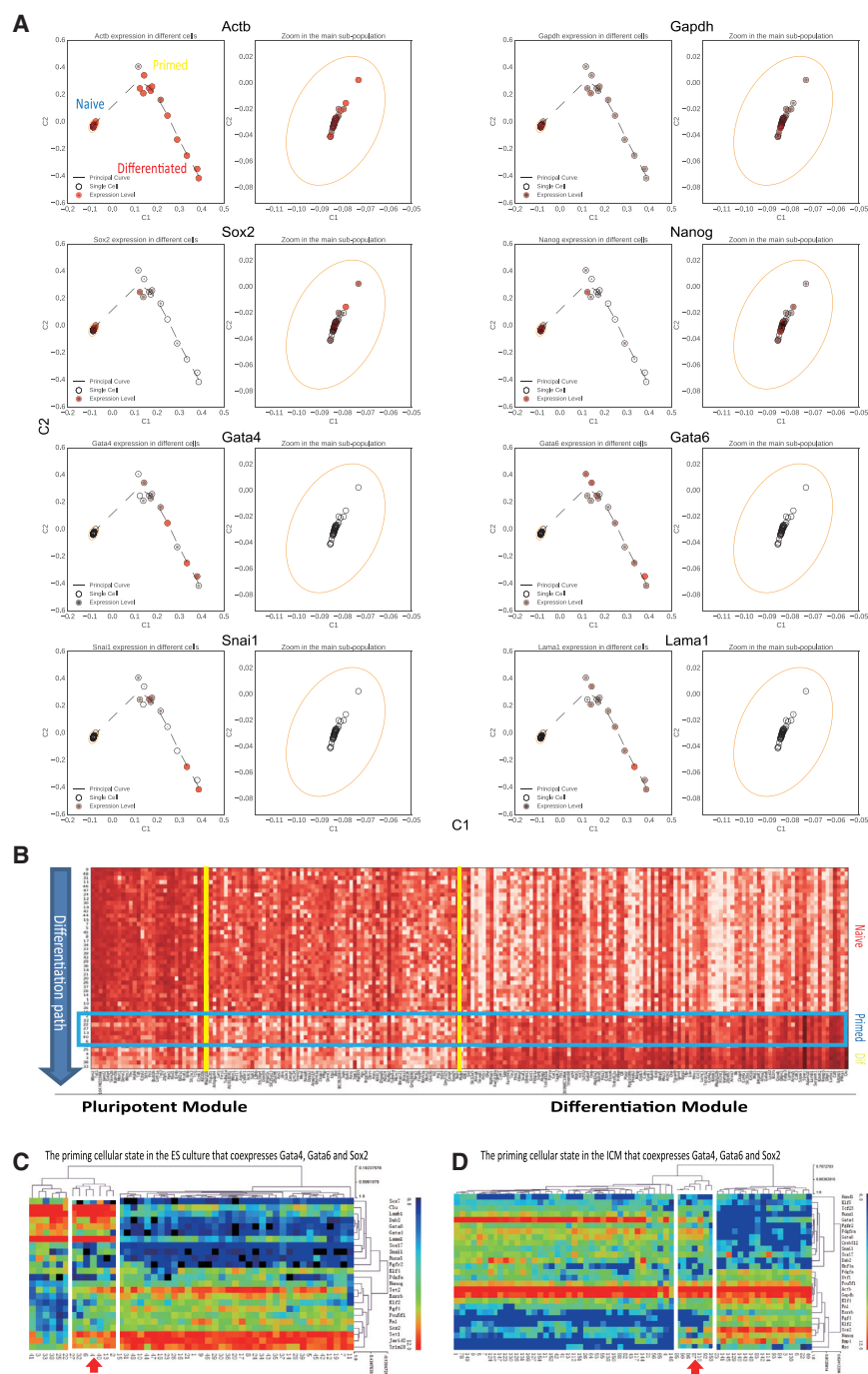


Figure 3. Computational Analysis Reveals ESC-Priming Pathway

(A) Local linear embedding plus the principal curve analysis reveals early priming pathway in the mouse ESCs in culture.

(B) Expression pattern of most-variable genes through the ESC-priming pathway indicates a transitional state that co-expresses pluripotent markers and differentiation markers.

(C) Hierarchical clustering of single-cell gene expression data reveals the primed pluripotent cells in ESC culture.

(D) Hierarchical clustering of single-cell gene expression data reveals the primed pluripotent cells in the blastocyst stage ICM. Note that the primed EPI cells co-express pluripotent markers and PE markers.

To examine in a systematic fashion the contribution of each gene to the pluripotency exit pathway, we calculated the Pearson correlation between its expression level in a cell and the mapped position on the principal curve. For the most-variable genes, we identified a subset whose expression levels were highly correlated with the differentiation path, including *Lama1* and *Lamb1* (Figure S3D). These genes are likely to play an important role in initiating cell differentiation. Similarly, we also identified another subset whose expression levels are anti-correlated with the differentiation path, such as *Tet1* and *Tet2*. These genes are likely to play an important role in maintenance of pluripotency. Figure 3B depicts an expression heatmap of highlighted genes along the early ESC differentiation path. In the heatmap, we show that primed cells co-express pluripotent and differentiation modules.

The primed ESC state maintains a distinct gene expression signature (Figure S3E). Hierarchical clustering of single-cell data also distinguishes this state as a unique cell-type cluster that co-expresses pluripotency markers and differentiation markers (Figure 3C). In order to link the state with in vivo developmental

Gata6 and *Gata4* at high level and pluripotent markers at low level. Cells of this population are predominantly representative of PE lineage cells, which are considered to be the default differentiation state for wild-type ESCs in culture. Of note, expression of *Tet1*, *Ezh2*, and *Suz12* was high in the naive state, reduced in the primed state, and then repressed in differentiated cells (Figure S3C), whereas the endogenous control markers, *Actb* and *Gapdh*, were robustly expressed in all cells.

processes, we reanalyzed previously published single-cell data from blastocyst stage ICM cells (Guo et al., 2010). We found a corresponding primed cell-type cluster that is distinct from known PE and EPI cell clusters in the blastocyst ICM (Figure 3D). The special cluster of blastocyst cells also co-expresses *Sox2*, *Gata4*, and *Gata6*. The identification of a primed state adds to the complexity of seemingly homogenous pluripotent cells and suggests stepwise exit from the naive pluripotent state both in vitro and in vivo.

External Culture System Affects Network Variability

A central question regarding cellular heterogeneity is whether variability in gene expression is derived from internal transcriptional “noise” or results from fluctuation in response to external signals. To address this question, we searched for upstream regulators of the variably expressed genes defined by our analysis. We used the Haystack pipeline (Pinello et al., 2014) to identify enriched transcription factor motifs upstream of different groups of highly variable genes (Figure 4A). For cluster 1 variable genes, we observed enrichment for a motif recognized by Zbtb33. For cluster 2 genes, the motif for TCF factors was enriched. For cluster 3 genes, the most-enriched motif corresponded to that for serum response factor, SRF. TCF factors lie downstream of glycogen synthase kinase 3 (GSK3) pathways in ESCs (Martello et al., 2012). SRF is a critical transcription factor that binds to the *c-fos* serum response element (Norman et al., 1988) that lies downstream of serum response and the MAPK pathways (Hill et al., 1993). These clues suggest that the serum-based culture conditions generally employed for ESCs might be a major contributor to variable gene expression observed in single-cell analysis. Indeed, downstream effectors for these signaling pathways are highly variable in cultured ESCs (Figures S4A and S4B).

Besides the classical serum-based culture conditions, a serum-free 2i culture system targeting both the MAPK and GSK3 pathways has been found to maintain mouse ESC pluripotency (Ying et al., 2008). To ascertain the contribution of serum-based culture conditions on expression variability, we assessed expression in J1 ESCs at the single-cell level in medium containing normal serum, knockout serum replacement, or 2i chemicals (PD184352 and CHIR99021). We analyzed these three ESC cultures using a more-cost-effective single-cell qPCR protocol that we previously described (Guo et al., 2013). We selected 96 genes for analysis, including known pluripotency regulators and differentiation markers, as well genes that displayed strong variability under standard culture conditions.

On examination of the single-cell data from cells under the three culture conditions, we found that ESCs cultured in serum expressed more markers of differentiation (e.g., *Id2*, *Lamb1*, and *Snai1*) than ESCs in 2i medium (Figure S4C). On LLE projection of single-cell data, global expression of 2i ESCs was more tightly distributed than that of ESCs cells cultured in serum or serum replacement (Figure 4B). Specifically, by focusing on distribution over the first principal component and distribution of expression SD, we confirmed that ESCs cultured in 2i medium exhibit the least variation. Medium with serum led to the greatest heterogeneity in expression (Figures 4C, 4D, and S4D). As revealed by violin plots in Figure 4D, the expression distribution of key regulators suggested a more-homogenous transcriptional network in 2i ESCs. For example, *Tbx3*, which is a highly variable pluripotency marker, showed clear bimodal distribution in serum-cultured ESCs. However, under 2i conditions, the percentage of *Tbx3*-expressing cells was significantly increased, whereas the differentiation priming marker *Snai1* was repressed. These findings were confirmed in E14 ESCs, as well as an independent clone of J1 ESCs (Figure S4E).

We next asked whether the reduced heterogeneity under 2i conditions was accompanied with altered epigenetic status. We searched for an effect of 2i culture on the bivalent marks of highly variable genes using available epigenomic data (Marks et al., 2012). The overall number of bivalent genes was reduced dramatically in 2i-cultured ESCs, as compared with ESCs cultured in serum-containing medium (Figure S4F). We also found that, among the most-variable genes defined in serum-cultured ESCs, two-thirds of bivalent markers lost their bivalency in 2i conditions (Figure 4E).

We then utilized published single-cell DNA methylation data (Smallwood et al., 2014) to interrogate the link between gene expression variation and DNA methylation variation. For each gene, we considered the region [TSS-2kb, TES+2kb] and calculated the difference in methylation variance between serum and 2i conditions. Interestingly, we found a moderate correlation ($r = 0.33$; p value = 0.0016) between the difference in methylation variance and the difference in gene expression variance between serum and 2i conditions (Figure 4F). When cells were cultured with serum, *Tbx3* and *Snai1*, two variable markers in ESCs, showed strong variation of DNA methylation level in gene bodies. However, when cultured with 2i medium, such epigenetic variation was significantly reduced (Figure S4G), suggesting that reduction of DNA methylation variability may in part contribute to reduction of gene expression variability.

In summary, the nature of the culture conditions represents an important contributor to bivalency, gene expression variation, and DNA methylation variation in mouse ESCs. With replacement of serum and proper targeting of the related signaling pathways, variability among ESCs is largely controllable without hampering pluripotency and self-renewal.

DISCUSSION

Cellular heterogeneity has been accepted as a hallmark of both embryonic and adult stem cells (Graf and Stadtfeld, 2008; Chambers et al., 2007). It has been proposed that variation in gene expression arises from transcriptional noise and network fluctuation and that associated heterogeneity accounts for stochasticity of cell fate decisions in stem and progenitor cells (Chang et al., 2008). Using mouse ESCs as a model, we have investigated global gene expression variability at single-cell resolution.

In agreement with recently published single-cell analyses of mouse ESCs (Grün et al., 2014; Kolodziejczyk et al., 2015; Kumar et al., 2014; Papatsenko et al., 2015), we observed significant heterogeneity in gene expression in the serum-cultured mouse ESCs. Using LLE analysis, we showed that heterogeneity does not appear to be stochastic but rather follows a defined differentiation pathway toward PE-like cells. Importantly, we defined a primed ESC state that reflects transition from a naive to differentiated state. ESCs in the primed state co-express pluripotency and differentiation modules. We have also provided evidence that the primed state is developmentally relevant, as the same signature is found in the developing mouse blastocyst during PE and EPI lineage specification.

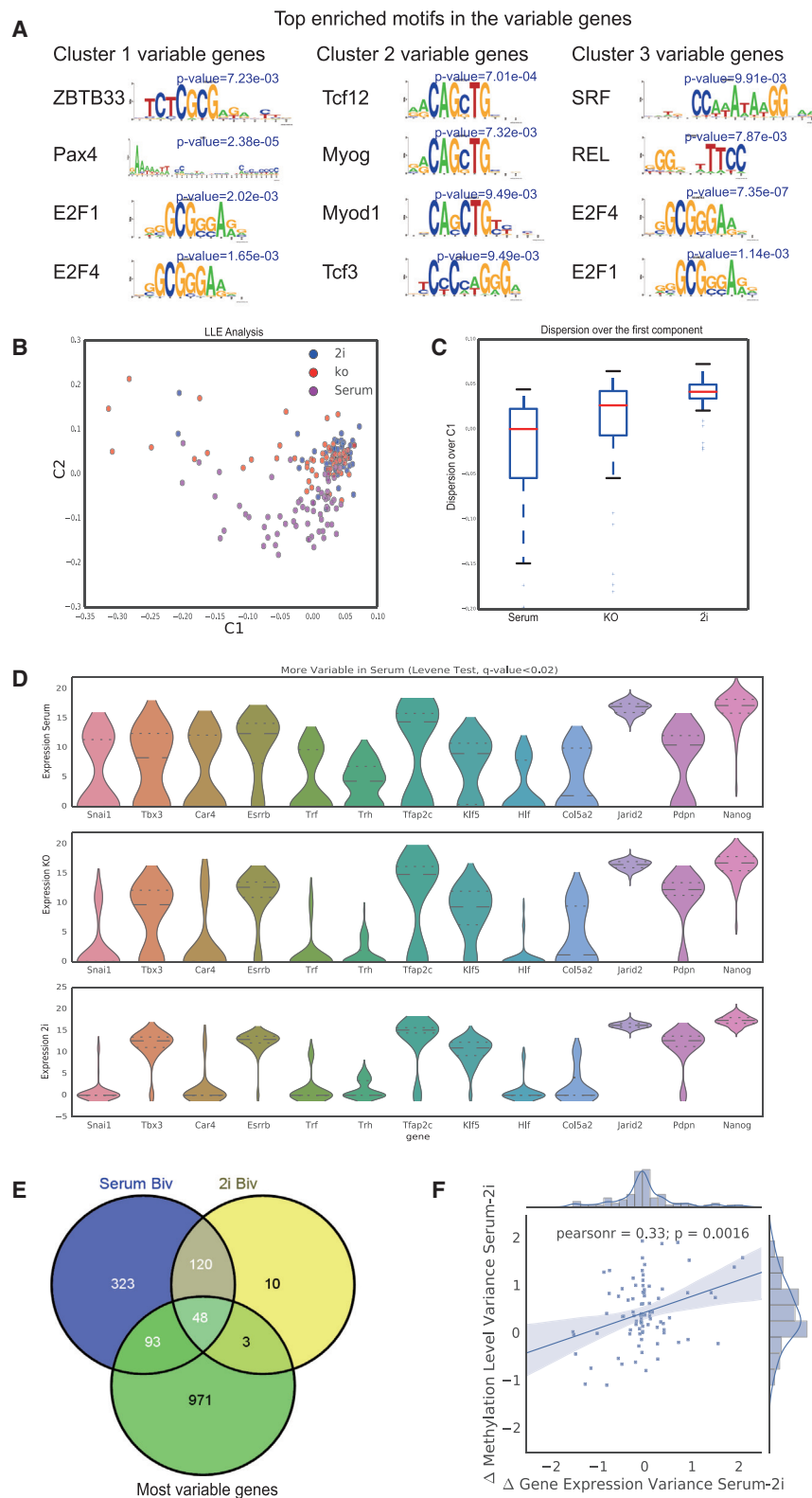


Figure 4. External Culture System Affects ESC Network Stability

(A) Motif analysis of most-variable genes predicts the roles of several important signaling pathways in regulating gene expression variability.

(B) LLE projection of single-cell analysis data from ESCs cultured with serum, knockout serum replacement, or 2i medium.

(C) Box plots for the expression distribution over the first PC reveal reduced gene variability in the 2i-medium-cultured ESCs.

(D) Violin plots showing expression level distribution of selected genes in ESCs cultured with serum, knockout serum replacement, or 2i medium.

(E) 2i medium reduces bivalency in the list of most-variable genes defined with serum-cultured ESCs.

(F) Correlation of single-cell-level gene expression variability and single-cell-level DNA methylation variability in 2i and serum-cultured ESCs.

Prior studies emphasized transcriptional networks and microRNA pathways that lead to gene expression variation (Klodziejczyk et al., 2015; Kumar et al., 2014). In the current work, we associate gene expression variation with epigenetic characteristics. We used LCOD analysis to extract true variability from mean expression level and describe the unique epigenetic status that distinguishes the highly variable genes. We propose that a proportion of previously defined bivalent marked genes are actually highly variable in their expression in cultured ESCs, suggesting a possible role for bivalent domains in modulating the frequency of transcription activation. One caveat is that ChIP-seq data are obtained from population-level studies. As a result, it remains unclear whether bivalent domains are established in all cells or only a fraction of cells. Future developments of methods for the mapping of epigenetic marks at single-cell resolution are needed to resolve these issues.

Importantly, we demonstrated that the culture environment contributes strongly to observed gene expression variability. Upon replacement of serum and targeting the MAPK and GSK3 pathways by 2i conditions, ESCs in culture exhibit greater homogeneity in gene expression. Our results confirm findings from other recent studies (Grün et al., 2014; Kumar et al., 2014). The FGF and MAPK pathway are closely related with EPI cell differentiation (Ying et al., 2008; Guo et al., 2010). WNT and GSK3 signaling has been implicated in control of gene expression noise during development (Arias and Hayward, 2006). Using motif analysis, we have connected gene expression variation with these important signaling pathways. Interestingly, replacing serum with knockout serum replacement alone also reduces gene expression variation, suggesting that other serum-responsive pathways contribute to ESC culture heterogeneity. Moreover, we show that 2i-cultured ESCs exhibit reduced bivalency and altered single-cell-level gene expression variation correlates with single-cell-level methylation status. We provide an example that gene expression variation is controllable through proper perturbation of key signaling pathways.

The plasticity of mammalian cellular states complicates an understanding of cell fate decision mechanisms. Comprehensive characterization of dynamic stem cell differentiation pathways requires single-cell gene expression analysis. Acquisition of similar analyses from different cellular systems should eventually allow for the mapping of the cell fate decision landscape and the modeling of dynamic network configurations during mammalian development.

EXPERIMENTAL PROCEDURES

Cell Culture

Mouse ESCs were cultured in feeder-free conditions. Cells cultured in serum were grown in DMEM + Glutamax (GIBCO) + 15% fetal bovine serum (FBS) (Gemini) + penicillin/streptomycin (GIBCO) + non-essential amino acids (NEAAs) (GIBCO) + b-mercaptoethanol (EmbryoMax ES cell-qualified; Millipore) + 1,000 U/ml LIF (Millipore). Cells cultured in knockout serum replacement were grown in DMEM + Glutamax (GIBCO) + 20% knockout serum replacement (GIBCO) + penicillin/streptomycin (GIBCO) + NEAAs (GIBCO) + b-mercaptoethanol (EmbryoMax ES cell-qualified; Millipore) + 1,000 U/ml LIF (Millipore). Cells cultured in 2i medium were grown in DMEM + Glutamax (GIBCO) + 20% knockout serum replacement (GIBCO) + penicillin/streptomycin (GIBCO) + NEAAs (GIBCO) + b-mercaptoethanol (EmbryoMax ES cell-qualified; Millipore) + 1,000 U/ml LIF (Millipore) + 1 mM MEK inhibitor

PD0325901 (StemGent) + 3 mM GSK3 inhibitor CHIR99021 (StemGent). Cells from the same line, same stock, and same passage were used for the culture system comparison experiment from Figures 4B–4D.

Single-Cell mRNA-Seq

Feeder-free J1 ESCs were grown in the presence of serum and LIF. ESCs were dispersed via trypsin-EDTA treatment. Single-cell whole-transcriptome amplification was performed using the Fluidigm C1 Single-Cell Auto Prep System (C1 System) as per the manufacturer's recommendations (full details available at <http://www.fluidigm.com>). Amplified cDNAs were diluted with C1 DNA dilution reagent, quantified using Quant-it HS system, and validated by qPCR with selected primers. Successfully amplified single-cell cDNA samples were selected and diluted to the same concentration. Single-cell libraries were constructed using the Nextera XT DNA Sample Preparation kit (Illumina), pooled using Nextera XT DNA Sample Preparation Index Kit (Illumina), and then sequenced using HiSeq 2500 (Illumina). Each library was sequenced twice on two lanes.

Single-Cell qPCR

Individual primer sets (total of 96) were pooled to a final concentration of 0.1 μ M for each primer. Individual cells were sorted directly into 96-well PCR plates loaded with 5 μ l RT-PCR master mix (2.5 μ l CellsDirect reaction mix, Invitrogen; 0.5 μ l primer pool; 0.1 μ l RT/Taq enzyme, Invitrogen; 1.9 μ l nuclease-free water) in each well. Sorted plates were immediately frozen on dry ice. After brief centrifugation at 4°C, the plates were placed immediately on PCR machine. Cell lyses and sequence-specific reverse transcription were performed at 50°C for 60 min. Then, reverse transcriptase inactivation and Taq polymerase activation was achieved by heating to 95°C for 3 min. Subsequently, in the same tube, cDNA was subjected to 20 cycles of sequence-specific amplification by denaturing at 95°C for 15 s, annealing, and elongation at 60°C for 15 min. After preamplification, PCR plates were stored at –80°C to avoid evaporation. Pre-amplified products were diluted 5-fold prior to analysis. Amplified single-cell samples were analyzed with Universal PCR Master Mix (Applied Biosystems), EvaGreen Binding Dye (Biotium), and individual qPCR primers using 96.96 Dynamic Arrays on a BioMark system (Fluidigm). Ct values were calculated using the BioMark Real-Time PCR Analysis software (Fluidigm).

ACCESSION NUMBERS

The accession number for all RNA-seq data reported in this paper is GEO: GSE75804.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.12.089>.

AUTHOR CONTRIBUTIONS

Conceptualization, G.G., L.P., G.-C.Y., and S.H.O.; Methodology, G.G., L.P., and X.H.; Investigation, G.G., L.P., X.H., S.L., L.S., and T.-W.L.; Writing – Original Draft, G.G. and L.P.; Writing – Review & Editing, K.Z., G.-C.Y., and S.H.O.; Supervision, S.H.O.; Funding Acquisition, S.H.O.

ACKNOWLEDGMENTS

We thank Y. Fujiwara, M. Nguyen, H.E. Benjamin, A. Saadatpour, E. Marco, and Z. Shao for help on experiments and Y. Zhang, Z. Li, H. Ouyang, Y. Zhou, and J. Ji for insightful discussions on the project. This work was supported by funding from NIH and the Harvard Stem Cell Institute (S.H.O.). S.H.O. is an investigator of the Howard Hughes Medical Institute (HHMI). G.-C.Y. is supported by NIH grant R01HG05085. G.G. is supported by the 1000 Talent Plan from China and the Fundamental Research Funds for the Central Universities. L.P. is supported by NHGRI Career Development Award K99HG008399.

Received: February 15, 2015
Revised: October 14, 2015
Accepted: December 18, 2015
Published: January 21, 2016

REFERENCES

- Arias, A.M., and Hayward, P. (2006). Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.* 7, 34–44.
- Bendall, S.C., Simonds, E.F., Qiu, P., Amir, el-A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332, 687–696.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326.
- Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209–1222.
- 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.
- Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544–547.
- Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* 29, 1120–1127.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotent cells from mouse embryos. *Nature* 292, 154–156.
- Fan, H.C., Fu, G.K., and Fodor, S.P. (2015). Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* 347, 1258367.
- Graf, T., and Stadtfeld, M. (2008). Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* 3, 480–483.
- Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11, 637–640.
- Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685.
- Guo, G., Luc, S., Marco, E., Lin, T.W., Peng, C., Kerenyi, M.A., Beyaz, S., Kim, W., Xu, J., Das, P.P., et al. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* 13, 492–505.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.
- Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 3, 391–401.
- Hill, C.S., Marais, R., John, S., Wynne, J., Dalton, S., and Treisman, R. (1993). Functional analysis of a growth factor-responsive transcription factor complex. *Cell* 73, 395–406.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 21, 1160–1167.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elieff, N., Paul, F., Zaretzky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776–779.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201.
- Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Illicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühler, M., Liu, P., et al. (2015). Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* 17, 471–485.
- 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.
- Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214.
- Marks, H., Kalkan, T., Menafr, R., Denissov, S., Jones, K., Hofmeister, 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.
- Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kingston, S., Joshi, A., Hannah, R., Theis, F.J., et al. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* 15, 363–372.
- Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55, 989–1003.
- Papatsenko, D., Darr, H., Kulakovskiy, I.V., Waghray, A., Makeev, V.J., MacArthur, B.D., and Lemischka, I.R. (2015). Single-cell analyses of ESCs reveal alternative pluripotent cell states and molecular mechanisms that control self-renewal. *Stem Cell Reports* 5, 207–220.
- Pinello, L., Xu, J., Orkin, S.H., and Yuan, G.C. (2014). Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. *Proc. Natl. Acad. Sci. USA* 111, E344–E353.
- Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30, 777–782.
- Roweis, S.T., and Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326.
- Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K.D., Imai, T., and Ueda, H.R. (2013). Quartz-seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* 14, R31.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaubomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240.
- Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382.
- Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. *Cell Stem Cell* 6, 468–478.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523.