MicroRNAs organize intrinsic variation into stem cell states

Meenakshi Chakraborty,a,b,c, Sofia Hu,a,b,d, Erica Visness,a, Marco Del Giudice,e,f,g, Andrea De Martino,a,h, Carla Bosia,i, and Phillip A. Sharp,b,1, and Salil Garga,d,j,1

*Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142; 1Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; 2Department of Genetics, Cambridge University, CB2 3EJ Cambridge, United Kingdom; 3Harvard–Massachusetts Institute of Technology Health Sciences and Technology, Cambridge, MA 02139; 4Italian Institute for Genomic Medicine, 10060 Candiolo, Italy; 5Department of Life Science and System Biology, Università degli Studi di Torino, 10124 Torino, Italy; 6Candiolo Cancer Institute, Istituto di Ricovero e Cura a Carattere Scientifico, 10060 Candiolo, Italy; 7Soft & Living Matter Lab, Istituto di Nanotecnologia, Consiglio Nazionale delle Ricerche, 00185 Rome, Italy; Department of Applied Science and Technology, Politecnico di Torino, 10129 Torino, Italy; and 8Department of Pathology, Massachusetts General Hospital, Boston, MA 02114

Contributed by Phillip A. Sharp, February 3, 2020 (sent for review November 25, 2019; reviewed by Richard W. Carthew and Carl D. Novina)

Pluripotent embryonic stem cells (ESCs) contain the potential to form a diverse array of cells with distinct gene expression states, namely the cells of the adult vertebrate. Classically, diversity has been attributed to cells sensing their position with respect to external morphogen gradients. However, an alternative is that diversity arises in part from cooption of fluctuations in the gene regulatory network. Here we find ESCs exhibit intrinsic heterogeneity in the absence of external gradients by forming interconverting cell states. States vary in developmental gene expression programs and display distinct activity of microRNAs (miRNAs). Notably, miRNAs act on neighborhoods of pluripotency genes to increase variation of target genes and cell states. Loss of miRNAs that vary across states reduces target variation and delays state transitions, suggesting variable miRNAs organize and propagate variation to promote state transitions. Together these findings provide insight into how a gene regulatory network can coopt variation intrinsic to cell systems to form robust gene expression states. Interactions between intrinsic heterogeneity and environmental signals may help achieve developmental outcomes.

A lan Turing first proposed the existence of chemical “morphogens” that could impart pattern formation during development (1). Later, Lewis Wolpert proposed that gradual differences in morphogen concentrations across cells’ external environment could suffice to define many different patterns (2–4). While organisms such as Drosophila and Xenopus develop by utilizing maternally deposited asymmetric morphogen gradients, mammals are unique because early stages of development take place without such obvious gradients (5). Investigations have aimed to recapitulate aspects of mammalian embryonic development through mixing cell types that signal to each other or supplying specified agonists in culture medium, often with external scaffolds to help organize cells (6–9). These studies have revealed “self-organized” body patterning axes, with formation of these axes relying to differing extents on the external signals provided by morphogens or scaffolds. Embryonic stem cells (ESCs) have proven to be particularly useful models for self-organizing processes that may shape the early embryo (6, 10, 11). Yet, the extent to which genetically identical cells can inherently generate diversity without external signals has remained less clear, as has a mechanism for such a phenomenon. This would require the gene regulatory network to give a variable output across cells that receive similar input. Prevailing views ascribe spontaneously arising cell-to-cell variation in gene expression to stochastic processes at gene loci, which include phenomena such as transcriptional bursting (12–19). However, an alternative hypothesis is that naturally arising variation within cell systems can be coordinated across multiple, specific loci by gene regulatory elements and coopted by the cell to enable diversification.

In ESCs, the core transcriptional regulatory network consists of pluripotency genes such as Pou5f1 (Oct4), Sox2, and Nanog (together “OSN”) (20–23). This regulatory network also contains microRNAs (miRNAs), small RNAs that bind and regulate genes in mammals through Argonaute (Ago) effector proteins (24), of which Ago2 is the principally active form in mammals. ESCs are known to express Nanog and Sox2 in a variable, heterogeneous fashion across cell populations (15, 25–30). The degree of cell-to-cell variation for Nanog and Sox2 is higher than that observed for Oct4 (15, 31, 32), though the molecular basis of this difference is unknown, as is the basis of Nanog and Sox2 heterogeneity. Additionally, the full extent to which the core ESC regulatory network can be subdivided based on factors driving cell-to-cell variation is unclear. In this study, we find ESCs exhibit intrinsic heterogeneity through formation of interconverting cell states in the absence of external gradients. Networks of genes and miRNAs that vary between states regulate each other, forming a circuit for variation that includes Nanog, Sox2, and Esrrb but not Pou5f1, Tcf3, or Smad1. These findings imply that the core transcriptional gene regulatory network of pluripotent embryonic stem cells self-organizes into distinct gene expression programs and displays distinct activity of microRNAs (miRNAs). Pluripotent embryonic stem cells (ESCs) contain the potential to form a diverse array of cells with distinct gene expression states, namely the cells of the adult vertebrate. Classically, diversity has been attributed to cells sensing their position with respect to external morphogen gradients. However, an alternative is that diversity arises in part from cooption of fluctuations in the gene regulatory network. Here we find ESCs exhibit intrinsic heterogeneity in the absence of external gradients by forming interconverting cell states. States vary in developmental gene expression programs and display distinct activity of microRNAs (miRNAs). Notably, miRNAs act on neighborhoods of pluripotency genes to increase variation of target genes and cell states. Loss of miRNAs that vary across states reduces target variation and delays state transitions, suggesting variable miRNAs organize and propagate variation to promote state transitions. Together these findings provide insight into how a gene regulatory network can coopt variation intrinsic to cell systems to form robust gene expression states. Interactions between intrinsic heterogeneity and environmental signals may help achieve developmental outcomes.

Significance

Understanding how mammalian organisms achieve the full diversity of cell types in the adult organism is a central goal of developmental cell biology. Recent work has shown that some embryonic precursor cells can self-organize into developmental structures but the mechanisms of gene regulation that contribute to this process remain unknown. Here we show embryonic stem cells self-organize into distinct gene expression states that resemble developmental gene programs. We find that microRNAs, small noncoding regulators of gene expression, play a critical role in organizing fluctuations across gene networks to help achieve this organization into distinct expression states.

Author contributions: P.A.S. and S.G. designed research; M.C., S.H., E.V., M.D.G., A.D.M., C.B., and S.G. performed research; M.D.G., A.D.M., and C.B. contributed new reagents/analytic tools; M.C., S.H., E.V., and S.G. analyzed data; and S.G. wrote the paper.

Reviewers: R.W.C., Northwestern University; and C.D.N., Dana-Farber Cancer Institute. The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-NID).

Data deposition: RNA-sequencing and small-RNA sequencing data are available on the Gene Expression Omnibus (GEO) with accession number GSE132708. SI Appendix, Tables S2 and S3 show RNA and miRNA expression in states. Other scripts, data, and detailed descriptions of sequencing pipelines are available through Zenodo (DOI: 10.5281/zenodo.3694341).

1To whom correspondence may be addressed. Email: sharp@a0.it or salig@smit.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1920695117/-/DCSupplemental.

First published March 5, 2020.
network of ESCs contains a subcircuit that coherently amplifies
variation to achieve transition to new states.

**Results**

**ESCs Exhibit Intrinsic Variation between States Expressing Distinct Developmental Gene Expression Programs.** To explore ESC variation, we measured the coding transcriptomes of individual ESCs by single-cell RNA-sequencing (scRNA-seq) and identified highly variable genes by a statistic (s-score) that corrects the coefficient of variation for technical sampling noise (32). Consistent with previous reports (15, 25, 27–30, 32, 33), we found a remarkably high degree of variation for Nanog and Sox2 transcripts across single ESCs (Fig. 1A). To further investigate variation in these pluripotency factors, we generated cells with heterozygous insertions of fluorophore tags at the endogenous loci of Nanog and Sox2 joined by posttranslational cleavage sequences (GFP-P2A-Nanog and Sox2-P2A-mCherry, respectively; SI Appendix, Fig. S1A). Identically cultured ESCs showed remarkable heterogeneity in levels of Nanog and Sox2 (Fig. 1B and SI Appendix, Fig. S1B). We analyzed the frequency distributions of Nanog and Sox2 levels, noting a dominant peak of high expression with one (Sox2) or two (Nanog) minor peaks at lower expression (SI Appendix, Fig. S1C). Additionally, Sox2-high Nanog-low ESCs have been previously identified (29). Thus, to enable a coarse grain analysis of the continuum of ESC variation, we clustered cells into three predominant states of Nanog and Sox2 expression for subsequent analysis (Fig. 1B; state 1 = high Nanog, high Sox2; state 2 = low Nanog, high Sox2; and state 3 = low Nanog, low Sox2). When cells from these states were isolated by flow cytometric sorting and cultured identically, each state recapitulated the heterogeneity of the parental population (Fig. 1B, Bottom). To extend this analysis, we sorted single ESCs from each state and assessed their organization and evolving state distribution by fluorescence microscopy. Single ESCs from each state grew into colonies with mixed state morphology, with a high degree of intra- and intercolony variation in state distribution (Fig. 1C), consistent with a previous report tracking Nanog in single-cell–derived ESC colonies (29). We did not detect any reproducible orientation of states with respect to each other, with states 2 to 3 sometimes oriented toward the edges of colonies and sometimes located centrally with respect to state 1 (Fig. 1C and SI Appendix, Fig. S2A and B). To further delineate whether single ESCs displayed an inherent capacity to organize into states, we introduced a unique molecular barcode into each

![Fig. 1. Single ESCs exhibit intrinsic variation between cell states. (A) Distribution of single-cell variation test statistic (s) scores for 7,259 genes across 2,299 well-sampled cells measured by scRNA-seq. Nanog, Sox2, and Esrrb are indicated, as are Pou5f1, Smad1, and Tcf3. (B) ESC labeled by heterozygous insertion of fluorophore tags at the endogenous loci for Nanog and Sox2 (GFP-P2A-Nanog, Sox2-P2A-mCherry) were separated into three distinct cell states by flow cytometric sorting and cultured identically. The population is shown over time. (C) ESCs were isolated from states 1 to 3 by flow cytometric sorting and plated at low density. Cells were analyzed by widefield fluorescence for Nanog (GFP) and Sox2 (mCherry) at the indicated timepoints. (D) A unique barcode was introduced into each ESC (SI Appendix, Supplementary Materials and Methods). Single ESCs from states 1 and 2, respectively, were isolated and cultured. State distribution and sequencing of the barcode region (red highlight) are shown. (E) Coefficient of variation (CV)-mean plot of protein coding gene expression across three states. Genes with differential expression between all three states (red) or between any two states (peach) are highlighted. Nanog, Sox2, and Pou5f1 (Oct4) are indicated. (F) Dimensionality reduction applied to scRNA-seq data. Cells are plotted according to their low dimensional representations by potential of heat-diffusion for affinity-based trajectory embedding (PHATE). Each cell is colored for relative enrichment by gene expression signatures of states 1 to 3 (SI Appendix, Supplementary Materials and Methods).](https://www.pnas.org/content/117/12/6943.full)
ESC. We sorted single cells from a given state and assessed their ability to repopulate other states. Over time, single-cell-derived ESCs with unique barcodes switched into other states (Fig. 1D). Together, these results establish that single ESCs contain an inherent ability to organize into a cell system containing a distribution of states. Individual cells switch between states to give rise to intrinsic heterogeneity at the population level.

To gain insight into the observed ESC states, we characterized their coding and noncoding transcriptomes by ribosomal RNA (rRNA)-depleted RNA-seq (GEO: GSE132708). Protein-coding genes differentially expressed between all three states were highly enriched for developmental regulators and certain pluripotency genes (including Nanog, Sox2, and Esrrb) and depleted for housekeeping, cell cycle, metabolic, and other pluripotency genes (including Pou5f1, Smad1, and Tcf3), suggesting variation across states was specific to particular developmental loci (Figs. 1E and 2A and B and SI Appendix, Fig. S3 A–C). Next, we sought to determine the extent to which the observed states reflected the scale of variation in ESCs by utilizing these state expression signatures. We represented ESC states in a gene-unbiased manner through dimensionality reduction of scRNA-seq data and colored cells by enrichment for states 1 to 3 expression programs. We chose methods of dimensionality reduction that emphasize progression along trajectories as opposed to clustering into groups (34, 35) to reflect the continuum of variation in ESCs. States 1 to 3, defined by Nanog and Sox2, captured a large portion of variation across single ESCs, as indicated by a gradual progression of cells enriched for each state’s expression across the major axes (Fig. 1F and SI Appendix, Fig. S3D). Further, variation across single cells was related to variation across states 1 to 3 for individual genes, as differentially expressed genes across states (including Nanog, Sox2, and Esrrb) showed relatively high t-scores (Fig. 1I and SI Appendix, Fig. S3E).

Next, we sought to test whether ESC states relate to specific developmental programs. Ontology analysis revealed that state 1 resembles a naïve, cytokine-responsive population of cells, whereas state 2 displays increased expression of preectodermal makers such as Sox18 and Neurod1, and state 3 displays increased expression of preendodermal and premesodermal markers such as Gata3 and Hoxa3 (Fig. 2A and C and SI Appendix, Fig. S3F). We compared states 1 to 3 to characterized gene expression profiles of the mouse blastocyst at developmental stages ranging from E4.5 to E5.5 in embryogenesis (9, 36). We calculated the distance of gene expression between conditions (SI Appendix, Supplementary Materials and Methods). While states 1 to 3 were most similar to each other, state 1 was closer in expression to E4.5 epiblasts than were states 2 and 3, whereas the latter were closer in expression to E5.0 or E5.5 (Fig. 2D and SI Appendix, Fig. S3G). Overall, we find that ESCs contain variation observable as cells transitioning between states that express distinct gene expression programs related to development.

Pluripotency Gene Neighborhoods Are Bound by miRNAs that Increase Target Variation. Diversification of single cells into new discrete states requires coordinated expression of gene programs, raising the question of how variation is related across different gene loci. To address this question, we constructed gene interaction neighborhoods using network inference methods (32, 37). The neighborhood of a chosen “node” gene represents the set of genes most closely correlated with it and with each other (SI Appendix, Supplementary Materials and Methods). Emphasis of this method on topology helps alleviate artifacts in correlation strength that may arise from the low technical sampling of transcripts in scRNA-seq. The neighborhoods of Nanog, Sox2, and Esrrb are shown (Fig. 3A and SI Appendix, Fig. S4A). We confirmed these in silico inferred neighborhoods were meaningful by testing the coexpression of Nanog with neighbors Eif2s2, Esrrb, and Hsp90ab1 by introducing a fluorophore tag at their endogenous locus in Nanog fluorophore-tagged cells. Eif2s2, Esrrb, and
Hsp90ab1 all showed covariation with Nanog by this method (SI Appendix, Fig. S4B). Further, the neighborhoods of Nanog, Sox2, and Esrrb all contain each other as members and have additional mutual neighbors (SI Appendix, Fig. S4C), supporting the idea that these genes interact and form an interconnected clique.

We analyzed these neighborhoods for molecular characteristics that could account for interactions between member genes giving rise to cell states. miRNAs are intriguing candidate cell state controllers because individual miRNAs can regulate hundreds of genes, which could allow cell-to-cell fluctuations in miRNA to generate relatively large effects on cell state (38). We mapped miRNA binding to target genes in ESCs using Ago cross-linking and immunoprecipitation (CLIP) data (SI Appendix, Supplementary Materials and Methods) (39). Next, we calculated whether neighborhoods were enriched for binding by particular miRNAs by comparing them to matched control neighborhoods constructed to contain the same number of genes of similar expression distribution and total binding of miRNAs (SI Appendix, Supplementary Materials and Methods). We found high miRNA binding of Nanog transcripts and of the entire variable pluripotency gene clique (Fig. 3A and B and SI Appendix, Figs. S4A and D). Pluripotency gene neighborhoods were enriched for binding by particular miRNAs, such as mir-182 and mir-708 (SI Appendix, Fig. S4E). Notably, transcripts and neighborhoods of less variable pluripotency genes Pou5f1, Smad1, and Tcf3, showed lesser binding by miRNAs and contained fewer mutual neighbors than variable pluripotency genes (Fig. 3B and SI Appendix, Figs. S4 C and D and S4F). The consistent enrichment of particular miRNAs within neighborhoods suggested a role for miRNA in regulating these neighborhoods.

To test if this was the case, we first determined variably expressed miRNAs (differentially expressed [DE]-miRNA) across states 1 to 3 (Fig. 3C and SI Appendix, Fig. S5A). DE-miRNA included many miRNAs enriched for binding pluripotency gene neighborhoods, including mir-182 and mir-708. Notably, although DE-miRNAs varied across states, their activity did not vary by cell cycle phase (SI Appendix, Fig. S5B, shown for mir-182). Classically, miRNAs are thought to regulate genes by repressing targets through mRNA destabilization or translational inhibition (40). We found increased mean miRNA levels for targets of DE-miRNAs in ESCs deficient for miRNA activity (Fig. 3D and SI Appendix, Fig. S6B) (41–43). Next, we tested the effect of a single DE-miRNA by generating ESCs deficient in mir-182 using CRISPR-Cas9 targeting of the miRNA hairpin loop (Mir182mut), a strategy previously reported to

Fig. 3. Single ESC variation is organized into neighborhoods bound by miRNAs. (A) Nanog interaction neighborhood, constructed by analyzing covariation across single cells. Thickness of Nanog connection represents number of mutual neighbors between Nanog and that gene, red shading indicates degree of Argonaute binding (miRNA activity). (B) Histogram for average miRNA binding per gene (miRNA binding events/number of neighbors) for the 6,577 non-empty neighborhoods produced for all of the genes present in WT scRNA-seq data. Values for Nanog, Sox2, and Esrrb neighborhoods are indicated in red; those for Pou5f1, Smad1, and Tcf3 neighborhoods are shown in black. (C) CV-mean plot for miRNA expression across states. DE-miRNAs are marked in red. Mir-182 and mir-708 are indicated. (D) CDF for the ratio of gene expression in Ago2-inducible ESCs with (WT) or without (KO) Ago2 induction. In these cells, Argonaute 2 is expressed from a doxycycline-inducible transgene in an endogenous Ago1

0.0003

1.0

0.0003

200

400

600

800

1000

0.0

0.2

0.4

0.6

0.8

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0

2.5

3.0

3.5

4.0

0.0

0.5

1.0

1.5

2.0
generate functional miRNA knockouts (KOs) (44). We validated that Mir182indel cells had little to no detectable miR-182 activity using reporters (45) (SI Appendix, Fig. S6A). We did not detect changes for miR-182 targets on average across Mir182indel cells (SI Appendix, Fig. S6C), leading us to consider whether DE-miRNA might be acting cooperatively to regulate targets. Consistent with this idea, binding-site analysis indicated that DE-miRNA targets are more highly bound by Ago2-miRNA complexes than non-variable miRNA targets (Fig. 3E). Interestingly, DE-miRNA binding was skewed toward lower affinity “6-mer” miRNA site type matches and therefore away from higher affinity “7/8-mer” sites (hypergeometric P value = 0.0003) when considering the degree of complementarity between the miRNA and its messenger RNA (mRNA) target (Fig. 3F).

Weak, cooperative interactions are a hallmark of molecular events prone to variation. Strikingly, targets of DE-miRNA showed significantly increased variation across single cells (ν-score) compared to all genes, with particularly high variation for miR-182 targets (Fig. 4A). We analyzed Mir182indel cells by scRNA-seq in parallel to WT cells and noted the increase in variation for miR-182 targets in wild-type (WT) ESCs was lost in Mir182indel ESCs (Fig. 4A). Therefore, while miR-182 did not have a detectable effect on average target expression across all cells, it appeared to have a significant effect increasing cell-to-cell variation of its targets (Fig. 4A).

Coordination and Propagation of Variation across Neighborhoods. Coordinated regulation of gene neighborhoods by miRNAs could provide a mechanism for individual miRNAs to impact the variation of many genes by binding and regulating their interacting neighbors. First, we asked what effect loss of miR-182 activity has on the DE-miRNA–bound variable pluripotency gene clique. We compared the variation across single cells (ν-score) for Nanog, Sox2, and Esrrb in WT vs. Mir182indel ESCs. These genes had lower variation in Mir182indel than in WT ESCs, even if they were not direct targets of miR-182 (Fig. 4B). By contrast, variation for Pou5f1, Smad1, and Tcf3, three pluripotency genes not bound by DE-miRNAs, was unchanged (Fig. 4B). This suggested that miR-182 propagates variability through the neighborhoods it binds in addition to promoting variability of its direct targets. To further assess whether variation can be propagated across neighborhoods, we plotted variation of all genes and their neighbors (Fig. 4C). Remarkably, highly variable genes across single cells showed a strong tendency to group into the same interaction neighborhoods, showing synchronous co-variation when measured across a population of uniformly cultured cells (Fig. 4C, note that red shading indicating higher ν-score is clustered in a subset of neighborhoods rather than being evenly distributed across all neighborhoods). This indicates that individual genes do not vary stochastically with respect to each other. Rather, variation is organized at the level of neighborhoods. The degree and concentration of variation is significantly decreased in Mir182indel ESCs (Fig. 4C and D), consistent with the idea that miR-182 loss reduces cell-to-cell variation of bound targets, which in turn leads to less variation across Mir182indel ESC neighborhoods compared to WT ESC neighborhoods. This synchronized co-variation of a group of genes within a cell could give that cell an inherent ability to diversify into a new state.

**Fig. 4.** ESC variation is coordinated and propagated across neighborhoods. (A) CDF plot of variation score across single cells (ν-score) for all genes (gray), genes targeted by two or more DE-miRNAs (peach), genes targeted by miR-182 in WT ESC (red), and genes targeted by miR-182 in Mir182indel ESC (violet). Kruskal–Wallis P values are shown. (B) Variation score across single cells (ν) for variable pluripotency genes Nanog, Sox2, and Esrrb and less variable (stable) pluripotency genes Pou5f1, Smad1, and Tcf3 in WT vs. Mir182indel ESCs. (C) Variation scores (ν) for all nonempty neighborhoods in WT and Mir182indel ESCs. The variation score of the node gene is shown in the Left column. Next, in each row we plot the score of each node gene’s neighbors (arranged from highest variation score at Left to lowest at Right). Neighborhoods are plotted Top to Bottom by decreasing average variation of all neighbors for WT and Mir182indel ESCs separately. (D) Overlay histogram of variation (ν-scores) in WT cells and Mir182indel cells for 6,107 genes well-sampled in both WT and Mir182indel cells.
Fig. 5. Variation in microRNA can drive variation in ESC states. (A) Cell state distributions in Ago2-inducible ESCs (Fig. 3D) labeled at Nanog and Sox2 loci by fluorophores (SI Appendix, Fig. S1). Cells were cultured in the indicated concentrations of doxycycline (Ago2-miRNA) for 48 h. Shown are variation in Nanog (*P < 0.0001, Levene’s test) and the fraction of cells outside state 1 (*P < 0.0001, binomial test). Error is 95% CI from bootstrapping. See also SI Appendix, Fig. S8A. (B) Cell state distributions in WT, Mir182indel, and Mir708indel ESCs labeled at Nanog and Sox2 loci by fluorophores (SI Appendix, Fig. S1). Shown are variation of Nanog and the fraction of cells outside state 1. Statistics were calculated as in A. See also SI Appendix, Fig. S8B. (C) Reintroduction of miR-182 into Mir182indel cells. Mir182indel cells were transfected with a bidirectional expression plasmid expressing pri-miR-182 tightly coupled to cerulean fluorescent protein (CFP), or with a CFP-only plasmid (“empty” control). Cells were isolated by flow cytometric sorting at increasing levels of CFP. miR-182 and miR-293 levels are quantified and plotted relative to WT ESCs. Error is SD for n = 3 replicates. (D) Nanog (GFP) and Sox2 (mCherry) levels in transfected Mir182indel cells, plotted by CFP (miR-182) levels as in C; *P < 0.0001, Levene’s test, pooled maroon vs. gray. Also shown is fraction of cells outside of state 1 (*P < 0.0001, binomial test). See also SI Appendix, Fig. S8C. (E) State 1 WT and Mir182indel ESCs were isolated by flow cytometric sorting and cultured. The fraction of cells outside state 1 is shown (*P < 0.0001, binomial test). See also SI Appendix, Fig. S8D.
Variation in miRNA Can Drive Variation in ESC States. Cell-to-cell variation in DE-miRNA expression and binding of these miRNAs within key neighborhoods could drive cell diversification into new states. This raises the possibility of miRNA regulation of Nanog and Sox2 contributing to the observed distribution of cell states through variation of miRNA levels. We explored qualitative models that could recapitulate the observed distribution of states. Previous work has established that bimodal distributions of target expression can emerge from the interplay between cell-to-cell miRNA variation and the threshold-like response of their targets (45, 46). Thus, we explored minimal qualitative models that could yield the observed distribution of states. We found that a model in which two distinct miRNA pools regulate Nanog and Sox2, one targeting both genes while the other controls Nanog specifically, recapitulated three cell states (SI Appendix, Fig. S7). In this configuration, cell-to-cell variation in miRNA levels results in highly variable expression of Nanog and Sox2. Loss of the shared miRNA pool, Nanog-specific pool, or of all miRNAs results in changes in the distribution of cell states (SI Appendix, Fig. S7).

To experimentally test this idea, we measured variation in ESC states in cells with inducible Ago2 expression (and therefore inducible miRNA activity) in an endogenous Ago1−/−−/−/3′′/4′′′ background (43). We observed a titratable increase in Nanog expression and cells exiting state 1 with increasing Ago2 miRNA activity (Fig. S4 and SI Appendix, Fig. S8A). A similar effect was observed upon loss of individual DE-miRNA, as Mir182indel and Mir708indel ESCs showed reduced state diversity compared to WT, with reduction in variation of Nanog and fewer cells exiting state 1 (Fig. S5B and SI Appendix, Fig. S6B). These results were consistent with the qualitative predictions of our model, and together they established that loss of miRNA could reduce cell state variation in ESCs.

To determine if reintroduction of miRNA could restore diversity, we constructed an inducible bidirectional plasmid expressing miR-182 tightly coupled to a fluorophore (CFP) and introduced it into Mir182indel ESCs. This allows single-cell measurement of miR-182, Nanog, and Sox2 levels through their respective fluorophores (CFP, GFP, and mCherry, respectively). CFP levels tracked miR-182 expression and the latter was restored to wild-type levels (∗∗∗+) or overexpressed (∗∗∗+, Fig. 5C). As a control, we transfected Mir182indel cells with a CFP-only “empty” plasmid in parallel to reexpression of miR-182. Mir182indel cells with miR-182 reexpressed showed an increase in cell state variation, with an increased fraction of cells outside state 1 and increased variation in Nanog and Sox2 levels (Fig. 5D and SI Appendix, Fig. S8C, note that levels of Nanog and Sox2 change across red “miR-182” bins but stay similar across gray empty bins). This indicates that adding exogenous variation in cell-to-cell miRNA levels by taking advantage of natural variation in plasmid transfection and expression efficiency increases variation in cell states across the population. Finally, we isolated state 1 ESCs from WT and Mir182indel ESCs by flow cytometry sorting and assessed their ability to diversify into other states over time. Mir182indel ESCs were delayed in diversifying out of state 1 compared to WT ESCs (Fig. 5E and SI Appendix, Fig. S8D). This supports the idea that miR-182 fluctuations impact transition of cells out of state 1, which is expected when miRNAs act as feedback repressors of genes subject to transient fluctuations (47). We conclude that cell-to-cell variation in DE-miRNA levels can drive variation in bound pluripotency gene neighborhoods in a subset of cells, enabling their diversification into new states.

Discussion

We find naturally arising variation in ESC gene expression can be described by three cell states with distinct expression programs related to embryonic development. Genes that vary across these three states encode specific developmental factors such as Nanog, Sox2, and Esrrb, but not others such as Pou5f1, Tcf3, or Smad1 that are equally well expressed. When isolated, single ESCs form a cell system that recapitulates the heterogeneity of the parental distribution despite the absence of an externally supplied morphogen gradient or scaffold. Variation within the cell system is concentrated at specific genes and miRNAs and excluded from others, forming an integrated genetic subcircuit that organizes variation into three cell states. Deletion of variable miRNAs reduces variation of target genes and reduces propagation of variation across the gene network, resulting in delayed ability for states to repopulate one another (Fig. 6). Reexpression of variable miRNA restores variation. These results define variation within cell systems as a fundamentally regulated process subject to modulation by noncoding elements such as miRNAs.

These results are consistent with previous reports describing heterogeneity in the expression of pluripotency factors in ESCs (15, 25–30). The molecular basis of this heterogeneity is still under debate, and the results here suggest miRNAs are important to ensure ESC heterogeneity by accelerating the dynamics of interconversion between states. A previous report utilized continuous long-term single-cell tracking to examine Nanog expression dynamics (29), similarly identifying that Nanog-low/ negative ESCs can revert to high Nanog expression under standard culture conditions. Interestingly, this study found that the context of cells surrounding Nanog-low cells influenced their propensity to differentiate into precursor cells expressing Foxa2 and Sox1. Future work aimed at understanding how the intrinsic state within a cell interacts with the milieu of states surrounding it to achieve appropriate diversity will be of great interest.

Our results show variable miRNAs such as miR-182 act on a clique of variable transcription factors (Nanog, Sox2, and Esrrb). More broadly, we show that variation across ESCs is organized into a specific subset of interacting gene neighborhoods. This contrasts with stochastic models in which variation at different loci is unrelated to one another or all loci are equally variable. Further, we find genes that vary between states encode developmental factors such as Nanog and Sox2. Since Nanog and Sox2 are themselves involved in promoting expression of many miRNAs (24), this provides opportunity for higher-order organization.
of intrinsic fluctuations within cells. Even within a genetically identical population of cells cultured under uniform conditions, fluctuations will inevitably arise in a subset of cells. If these fluctuations are always focused to a particular set of interacting genes in a highly ordered way, as identified here, and those genes encode developmental factors, variation of these factors within this subset of cells will inevitably result in their transition to another developmental cell state. This causes intrinsic heterogeneity or the ability of cells to repopulate specific cell states when separated.

Previous descriptions of cell-to-cell variation in gene expression have attributed it to noise or stochastic processes (12–19). Yet we find manipulating the level of miR-182 in ESCs can directly change cell-to-cell variation in Nanog levels. The presence of a genetic element (in this case, miRNA) whose manipulation explicitly impacts the distribution of a gene’s expression across cells indicates distribution of expression across cells for a gene is in fact highly regulated.

We identified a role for miRNA to increase variation, in contrast to their previously emphasized role reducing cell-to-cell variation in gene expression (48). Many additional gene regulatory components beyond miRNAs, such as enhancers, RNA binding proteins, or splicing factors may also contribute to cell state variation by virtue of their ability to regulate many genes. Although challenging, it will be important to further define the structural and regulatory features of molecular classes that vary in activity across cell to cell.

No matter how cell-to-cell variation originates, gene regulatory networks are poised to organize and direct it to achieve diversification of states in a robust manner without requiring varied input to each cell from an external gradient. Thus, organization of variation intrinsic to cell systems is a sufficient principle to allow for single cells to diversify into distinct states, providing a mechanism for cell-intrinsic symmetry breaking. This feature of gene regulatory networks may in part explain how self-organization occurs, and its description here builds upon previously observed links between gene expression variation and differentiation (49–51). We show that single ESCs can both self-organize into three states and that these states can repopulate one another. In adult tissues such as intestine and lung, stem-like cell types have shown the ability to repopulate each other (52–54). Development of tissues is also known to be intimately tied to miRNA function (47). Whole-body organoids and tissues receive external signals in addition to experiencing intrinsic fluctuations, these may cooperate in homeostatic ratios of cell states. Future studies will be necessary to determine the extent to which phenomena in mammalian development and tissue homeostasis are related to the intrinsic heterogeneity observed in ESCs. In addition to development, other biological systems demonstrate spontaneous organization of seemingly equivalent cells into states, such as bacteria in forming persistor states or cancer cells in forming therapy resistant subpopulations (55–59). Determining the configurations of gene regulatory networks within such systems will be of great interest. Nevertheless, the results presented here suggest that naturally arising cell-to-cell variation, sometimes described as stochastic fluctuation, is in fact coherently organized biology.

Materials and Methods

For additional information, please see SI Appendix, Supplementary Materials and Methods.

Cell Lines. Ten mouse ESC lines were used in this study: 1) DGC87−/− ESC (42); 2) Ago2-inducible ESCs (43), Nanog-GFP[Sox2-mCherry]; 3) V6.5 ESC (Jaenisch Laboratory, Whitehead Institute, Massachusetts Institute of Technology [MIT]), Nanog-GFP[Sox2-mCherry]; 4) V6.5 ESC, Nanog-GFP[Sox2-mCherry] (nuclear localization sequence [NLS]-tagged fluorophores); 5) V6.5 ESC, Nanog-GFP[Sox2-mCherry]; 6) V6.5 ESC, Mir107[nt], Nanog-GFP[Sox2-mCherry]; 7) V6.5 ESC, Nanog-GFP[Sox2-mCerulean3]; 8) V6.5 ESC, Nanog-GFP[Sox2-mCherry]; and 10) V6.5 ESC, Nanog-GFP[Hsp90a1b-mCherry]. All fluorophore tags, NLSs, and microRNA gene indels were added by us (see relevant sections in SI Appendix).

Flow Cytometry and Fluorescence Activated Cell Sorting (FACS). A total of 50 to 90% confluent ESCs were analyzed on BD LSR or LSRFortessa with FACSComp v8.0 acquisition software. Flow cytometry standard files were analyzed with FlowJo V9.9. Samples were gated first for live cells based on forward scatter area vs. side scatter area and then for single cells (forward scatter width vs. forward scatter height). Cells singly transfected with transient fluorophore expression constructs or simply tagged with either GFP or mCherry were used as fluorescence compensation controls. Channel gains were adjusted based on native V6.5 ESC.

BD FACsARIA was used to sort Nanog-GFP[Sox2-mCherry] ESCs into states (as defined by GFP[mCherry protein levels]), and to sort single cells. States were sorted into fresh culture medium in 5-ml collection tubes, then immediately spun at 1,000 rpm for 5 min and resuspended for either plating or RNA isolation. Single cells were sorted into 200 ml fresh culture medium in one well of a 96-well flat bottom plate (VWR catalog no. 29442-054).

RNA-Sequencing Sample Preparation. Standard TRizol protocol was used to isolate total RNA from sorted states 1/2/3 or the total unseparated ESC population passed through the sorter. Following DNase I (NEB M0303) treatment and ethanol precipitation, samples were analyzed by Agilent BioAnalyzer and accepted for sample RNA integrity number > 7.0. RNA-seq libraries represent those GeneChip replicates isolated by flow sorting on 3 separate days. rRNA-depleted RNA-sequencing libraries (~100 ng RNA/sample) were prepared using Kapa RNA HyperPrep Kit with RiboErase (HMR) KKB8561 using 11 rounds of PCR amplification after addition of External RNA Controls Consortium spike-in controls at the recommended concentration. The final libraries were quality control (QC) checked by fragment electrophoresis and qPCR for colony-forming units prior to pooling and loading on an Illumina FlowCell (NextSeq 500; 150 cycles for paired reads). Each sample library was sequenced to 30 to 45-M read depth.

Single-Cell RNA-Seq. scRNA-seq was performed by the Koch Institute Nanowell Cytometry Core using SeqWell (60) technology. In brief, single-cell suspensions of fluorophore-tagged V6.5 ESCs were made by trypsinization followed by serial passage through 50-μm cell strainer meshes. Approximately 10,000 cells were loaded onto a SeqWell array (54) and prepared as single-cell cDNA libraries (60). Libraries were sequenced using a NextSeq 500 and aligned to the mm10 genome using the SeqWell analysis pipeline (Love Lab, MIT). The entire process was repeated on a different day to generate two data tables representing reads per cell across mm10 annotated transcripts (GenCODE M15). These two tables were merged together and analyzed, prior to gene neighborhood construction (SI Appendix, Supplementary Methods and Materials).

miRNA-Seq and Data Analysis Pipeline. Total RNA samples were prepared from ESC states identically to RNA-seq above for two biological replicates isolated by flow cytometric sorting. Small RNA libraries were then prepared using the NEB small RNA-sequencing kit (E7300S) according to the manufacturer’s instructions using 13 cycles of PCR amplification. QC assessment was done by electrophoresis and colony forming units prior to pooling and loading on an Illumina FlowCell (NextSeq 500; 150 cycles for paired reads). Each sample library was sequenced to 30 to 45-M read depth.

Statistical Analyses. For all analyses with P values, significance was determined at P < 0.05. P values are shown in the figure, or in the figure legend, wherever they are used. The Kolmogorov–Smirnov (K–S) test was used to assess whether paired cumulative distribution functions (CDFs) were significantly different (Fig. 3 D and E and SI Appendix, Fig. S6 B and C). The Kruskal–Wallis one-way analysis of variance test was used to assess differences between 3+ CDFs (Fig. 4A). Both the K–S and Kruskal–Wallis test statistics were calculated in GraphPad Prism. In all cases, a CDF based on the population was compared to CDF(s) based on sample(s). The cumulative hypergeometric statistical test for enrichment was used in Fig. 3F. This test detects enrichment for a property in a population sample, compared to what would have been expected based on the prevalence of that property in the whole population. Four numbers are required: population size (N), number of population successes (n), sample size (k), and number of sample successes (k). The test statistic was calculated using Python v3.6, using scipy.stats.hypergeom.pmf and summing from k to min(n,K) to calculate the cumulative value. Box-whisker plots in Fig. 5 A and B were generated using matplotlib (Python v3.6), and associated P values reflect the results of Levene’s
28. H. Ochiai, T. Sugawara, T. Sakuma, T. Yamamoto, Stochastic promoter activation affects gene expression in mouse and human embryos. Nature 561, 186–190 (2018).

29. A. Filipczyk, W. A. Whyte, A. Tripicchio, S. Mirin, A. Kless, E. Kondos, M. D. Carver, J. M. Dowen, C. Merlo, A. Beal, et al., Network effects control lineage choice in mammalian progenitor cells. Cell 173, 214–228 (2018).

30. V. Karwacki-Neisius et al., Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell 12, 531–545 (2013).

31. A. M. Klein et al., Droplet barcoding for single-cell transcriptomics applied to cancer. Nature 515, 471–476 (2014).

32. M. A. Mayer et al., Her2 expression identifies dynamic functional states within circulating breast cancer cells. Nature 537, 428–433 (2016).

33. A. S. Tsoi, D. P. Bartel, The role of extrinsic noise in microRNA regulation and its consequences. Nat. Rev. Mol. Cell Biol. 15, 512–525 (2014).

34. D. P. Bartel, MicroRNAs: Target recognition and regulatory functions. Cell 136, 215–233 (2009).

35. J. Li, S. Horvath, Network neighborhood analysis with the multi-node topological data analysis. PLoS Comput. Biol. 9, e1003258 (2013).

36. S. Chen et al., Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell 160, 1264–1276 (2015).

37. S. Mukherji et al., MicroRNAs can generate thresholds in target gene expression. Nat. Genet. 43, 854–859 (2011).

38. J. R. Mckinney et al., Specific metabolic pathways in melanoma. Nat. Genet. 44, 152–157 (2012).

39. D. P. Bartel, MicroRNAs: Target recognition and regulatory functions. Cell 136, 215–233 (2009).

40. J. Beumer, H. Clevers, Regulation and plasticity of intestinal stem cells during homeostatic and regenerative states. Gastroenterology 147, 859–872 (2014).

41. J. S. Kim et al., Network-level control of cell identity and disease. Cell 160, 1264–1276 (2015).

42. V. Sharma et al., A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141, 69–80 (2010).

43. S. M. Shaffer et al., Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature 546, 431–437 (2017).

44. N. V. Jordan et al., HER2 expression identifies dynamic functional states within circulating breast cancer cells. Nature 537, 107–106 (2016).

45. T. Tammela et al., A wealth of transcriptional flexibility drives proliferative potential and progression in lung adenocarcinoma. Nature 545, 355–359 (2017).

46. T. M. Gierahn et al., Seq-Well: Portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods 14, 395–398 (2017).