SI Appendix

Contents:

www.pnas.org/cgi/doi/10.1073/pnas.1920695117

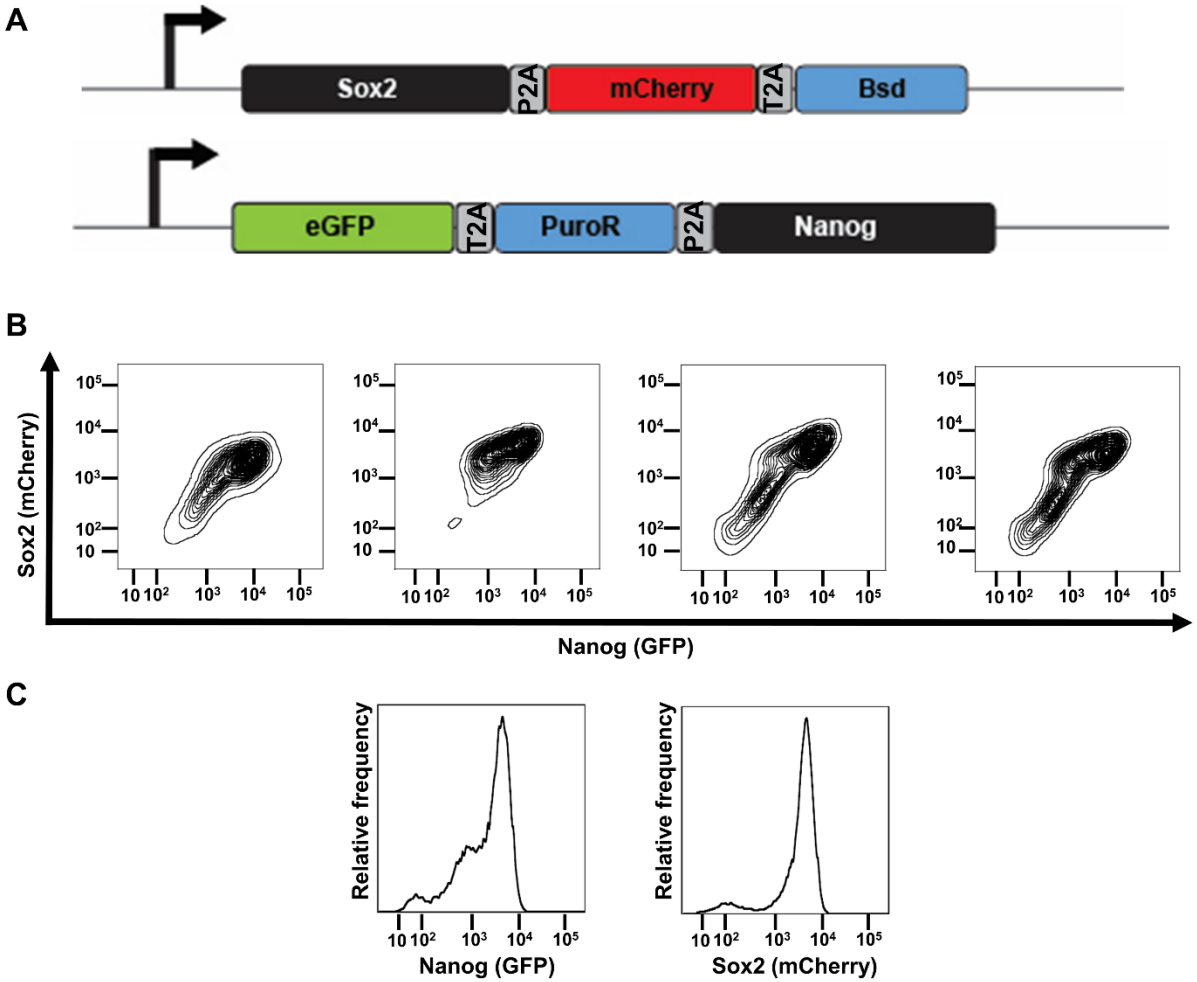

Figure S1: Embryonic stem cells show intrinsic heterogeneity in cell states. (A) Schematic depicting the modifications of *Nanog* and *Sox2* loci in embryonic stem cells used in this study. Note these modifications are used for all non-microscopy analyses. For microscopy, a nuclear localization sequence (NLS) was included prior to *GFP* and *mCherry*. See Methods for further details on the generation of cell lines using these reporters. **(B)** Three states (State 1 = Nanog-high, Sox2-high; State 2 = Nanog-low, Sox2-high; State 3 = Nanog-low, Sox2-low) across several distinct *Nanog-GFP/Sox2-mCherry* clones generated from the ESC line V6.5. **(C)** Frequency distributions of Nanog (GFP) and Sox2 (mCherry) in the *Nanog-GFP/Sox2-mCherry* clone shown in Fig. 1*B*.

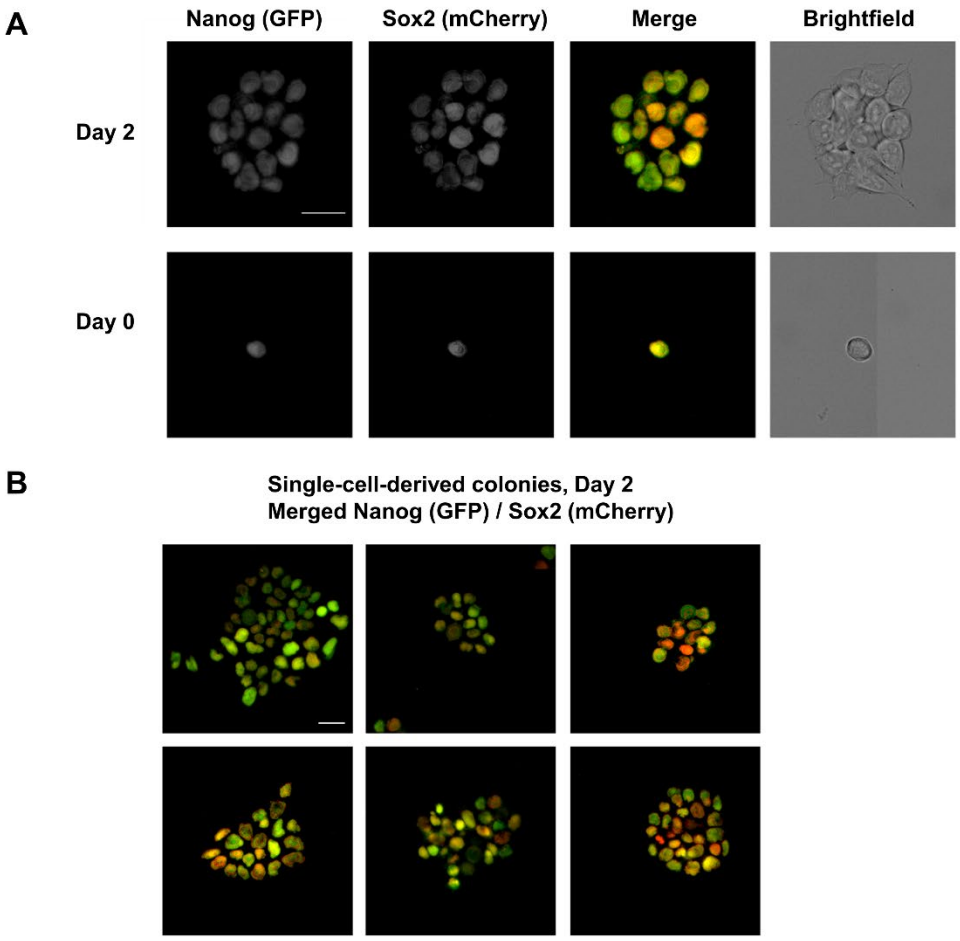

Figure S2: Cell states can arise from single ESCs. (A) A single-cell-derived ESC colony, shown imaged for Nanog (GFP), Sox2 (mCherry), merged Nanog/Sox2, and brightfield after 2 days of growth. **(B)** Single-cell-derived ESC colonies, shown imaged for merged Nanog (GFP) and Sox2 (mCherry) after 2 days of growth. Scale bar = 25 μm.

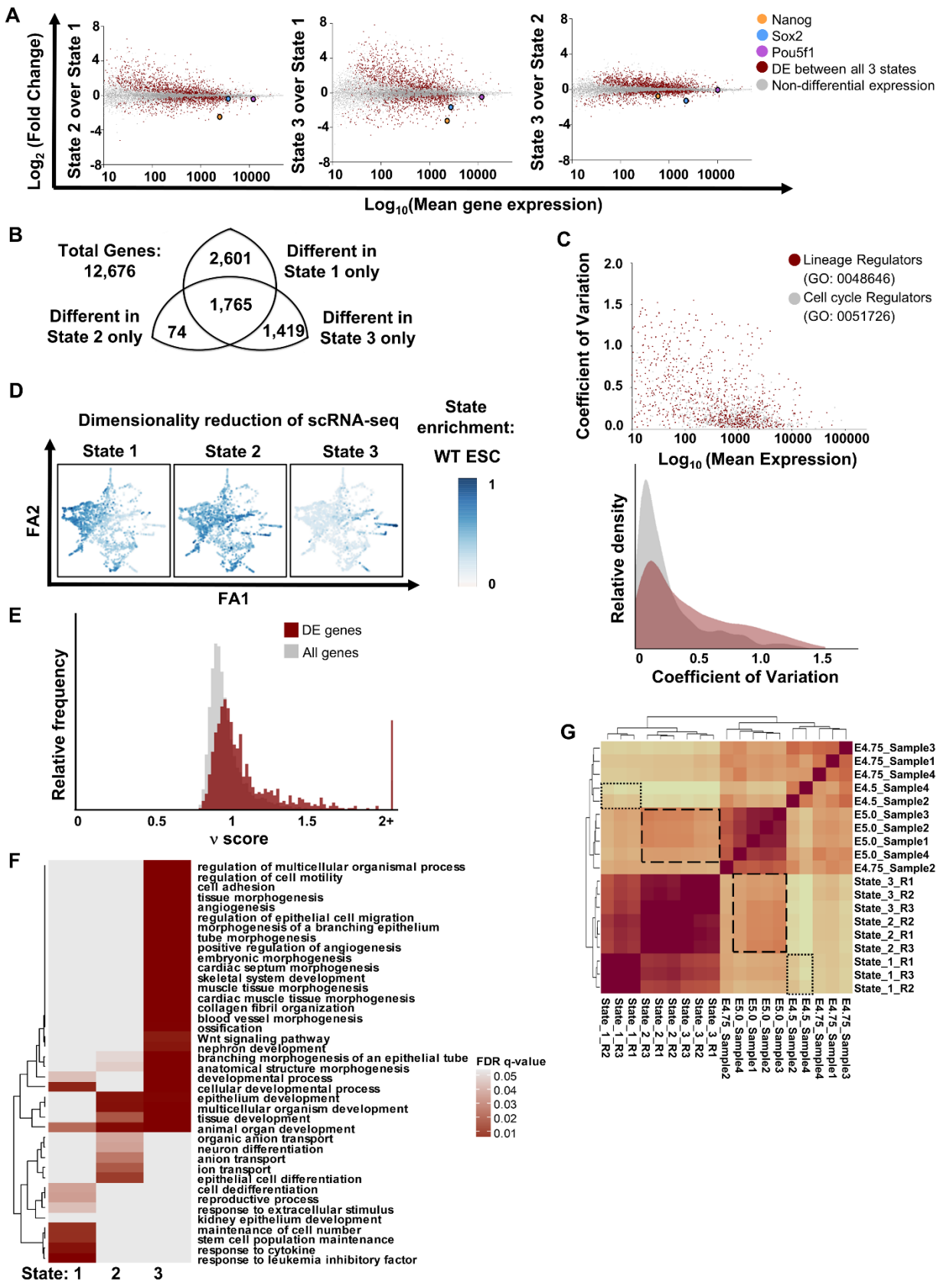
A
Log2 (Fold Change)
State 2 over State 1
State 3 over State 1
State 3 over State 2
Log10(Mean gene expression)
Nanog
Sox2
Pou5f1
DE between all 3 states
Non-differential expression
B
Total Genes: 12,676
2,601
Different in State 1 only
1,765
Different in State 2 only
74
1,419
Different in State 3 only
C
Coefficient of Variation
Log10 (Mean Expression)
Lineage Regulators (GO: 0048646)
Cell cycle Regulators (GO: 0051726)
Relative density
Coefficient of Variation
D
Dimensionality reduction of scRNA-seq
State 1
State 2
State 3
FA2
FA1
State enrichment: WT ESC
E
Relative frequency
DE genes
All genes
ν score
F
regulation of multicellular organismal process
regulation of cell motility
cell adhesion
tissue morphogenesis
angiogenesis
regulation of epithelial cell migration
morphogenesis of a branching epithelium
tube morphogenesis
positive regulation of angiogenesis
embryonic morphogenesis
cardiac septum morphogenesis
skeletal system development
muscle tissue morphogenesis
cardiac muscle tissue morphogenesis
collagen fibril organization
blood vessel morphogenesis
ossification
Wnt signaling pathway
nephron development
branching morphogenesis of an epithelial tube
anatomical structure morphogenesis
developmental process
cellular developmental process
epithelium development
multicellular organism development
tissue development
animal organ development
organic anion transport
neuron differentiation
anion transport
ion transport
epithelial cell differentiation
cell dedifferentiation
reproductive process
response to extracellular stimulus
kidney epithelium development
maintenance of cell number
stem cell population maintenance
response to cytokine
response to leukemia inhibitory factor
State: 1 2 3
FDR q-value
G
E4.75_Sample3
E4.75_Sample1
E4.75_Sample4
E4.5_Sample4
E4.5_Sample2
E5.0_Sample3
E5.0_Sample2
E5.0_Sample1
E5.0_Sample4
E4.75_Sample2
State_3_R1
State_3_R2
State_3_R3
State_2_R2
State_2_R1
State_2_R3
State_1_R1
State_1_R3
State_1_R2

Figure S3: ESC gene expression variation relates to developmental expression programs. **(A)** Gene expression changes between ESC states for protein coding genes, shown as average of expected counts versus fold change between states (MA plots). Highlighted in red are genes with significantly differential expression between all three states**.** *Nanog*, *Sox2*, and *Pou5f1* (*Oct4*) are indicated. **(B)** Diagram of the breakdown in expression patterns for the 5,859 genes with any significant expression differences between states. "Different in State X only" refers to genes that showed similar expression in two states but significantly different expression in State X. **(C)** Coefficient of variation (CV) across states (y-axis) versus mean expression across all three states (x-axis) for lineage regulators and cell cycle genes. All genes expressed in ESC (≥ 10 expected counts) from the listed GO terms were included. The CV data are also shown as an overlaid histogram to facilitate comparison. **(D)** Dimensionality reduction applied to scRNA-seq. Cells are plotted according to their low dimensional representations (force-directed, FA1 vs FA2). Each cell is colored for relative enrichment by gene expression signatures of States 1-3 (see Methods). Compare to Fig. 1*F*. **(E)** Comparison of variation across single cells (ν-score) for genes differentially expressed across States 1-3 (red) or non-differentially expressed (grey). Note these categories match those in main Figure 1. All expressed genes from Fig. 1 captured in scRNA-seq are included. **(F)** The top 300 genes uniquely highest expressed in each state were analyzed for ontology enrichment against the background of all expressed protein coding genes in ESC. Shown is a heatmap of FDR-q values for selected representative ontology terms. **(G)** Heatmap of gene expression distances between States 1-3 when compared to blastocyst expression profiles derived from days E4.5, E4.75, E5.0, and E5.5 (1). Short dashes highlight the comparison of State 1 with E4.5 samples and long dashes highlight the comparison of E5.0 with States 2-3. Comparison with all given sequencing replicates from the indicated reference are shown. See Methods for further details.

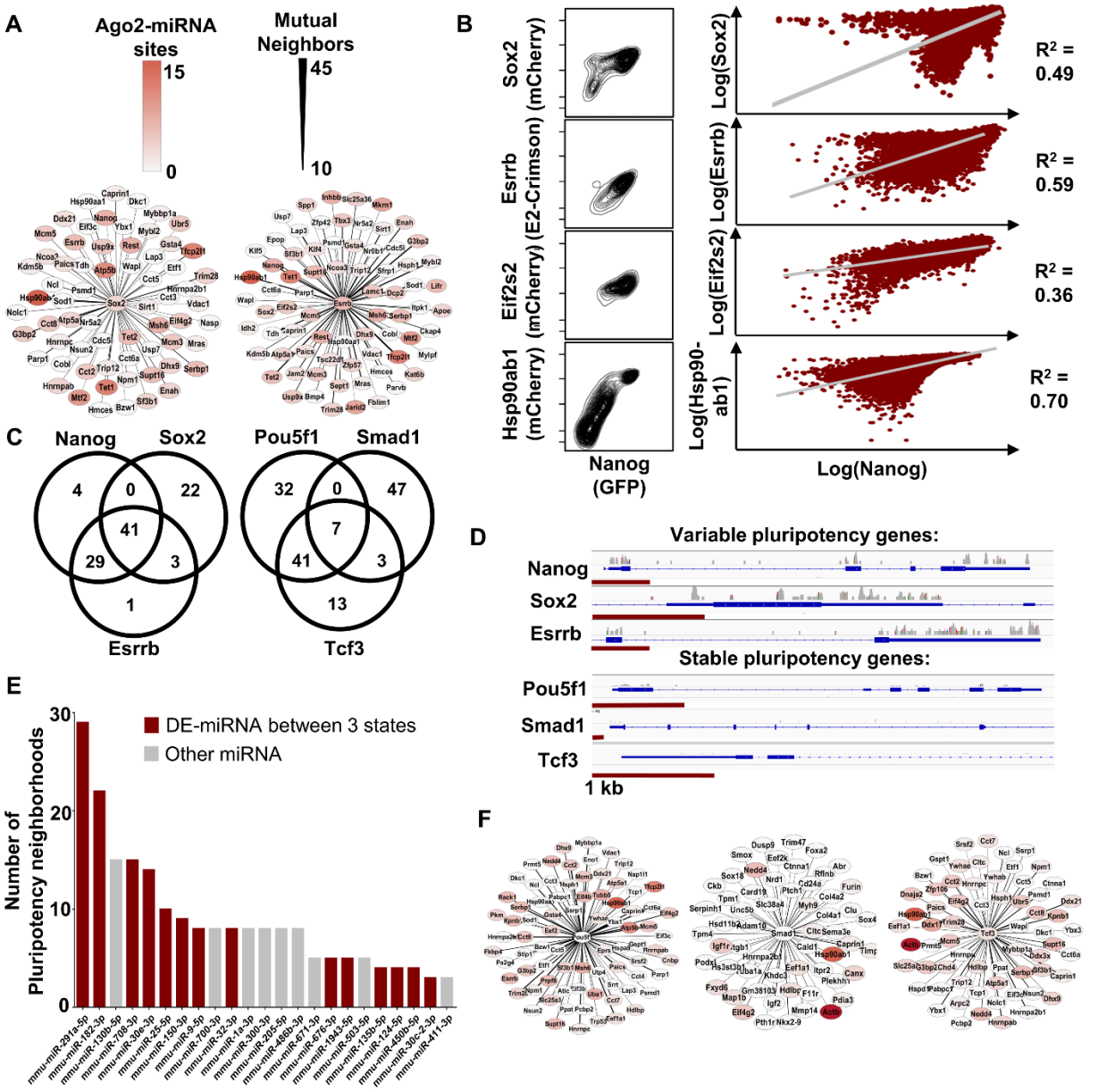

Figure S4: miRNA binding to gene neighborhoods. (A) *Sox2* and *Esrrb* neighborhoods and degree of binding by Ago2-miRNA; compare to Fig. 3*A*. **(B)** Covariation of proteins encoded by *Nanog* neighbors with Nanog. In each cell line, the neighbor of *Nanog (GFP*-P2A-*Nanog)* is then tagged at its endogenous locus with the indicated fluorophore (*Gene*-P2A-*fluorophore*). Cells are then analyzed by flow cytometry for fluorophore levels, which approximate the tagged genes. FACS distributions are shown at left; at right are the corresponding scatter plots, where each dot represents a single cell's fluorescence levels. Best fit lines and R^2 values are shown. **(C)** Venn diagram showing the overlap in neighbors for variable pluripotency genes and less variable pluripotency genes. Note that 41 neighbors occur in each of the *Nanog*, *Sox2*, and *Esrrb* neighborhoods, indicating these form a dense clique. In contrast, less variable pluripotency gene neighborhoods show fewer mutual neighbors, though *Pou5f1* and *Tcf3* do overlap. Only 20/41 genes overlapping in the *Nanog*/*Sox2*/*Esrrb* clique overlap with *Pou5f1* or *Tcf3*. **(D)** Ago2 CLIP coverage and gene structure are shown for variable pluripotency genes *Nanog, Sox2, Esrrb* and less variable pluripotency genes *Pou5f1, Smad1,* and *Tcf3*. Ago2 coverage is scaled identically across all six displayed loci (linear scale). For each gene, the 3'UTR is shown along with much of the gene body. Red scale bars indicate 1 kb for each gene. Binding data was mapped as previously described (2). **(E)** Number of pluripotency neighborhoods (out of 55) significantly enriched for binding by each miRNA by comparison to control neighborhoods (see Methods). **(F)** *Pou5f1*, *Smad1*, and *Tcf3* neighborhoods, degree of binding by Ago2-miRNA, and mutual neighbor strength; compare to Figs. 3*A* and S4*A*.

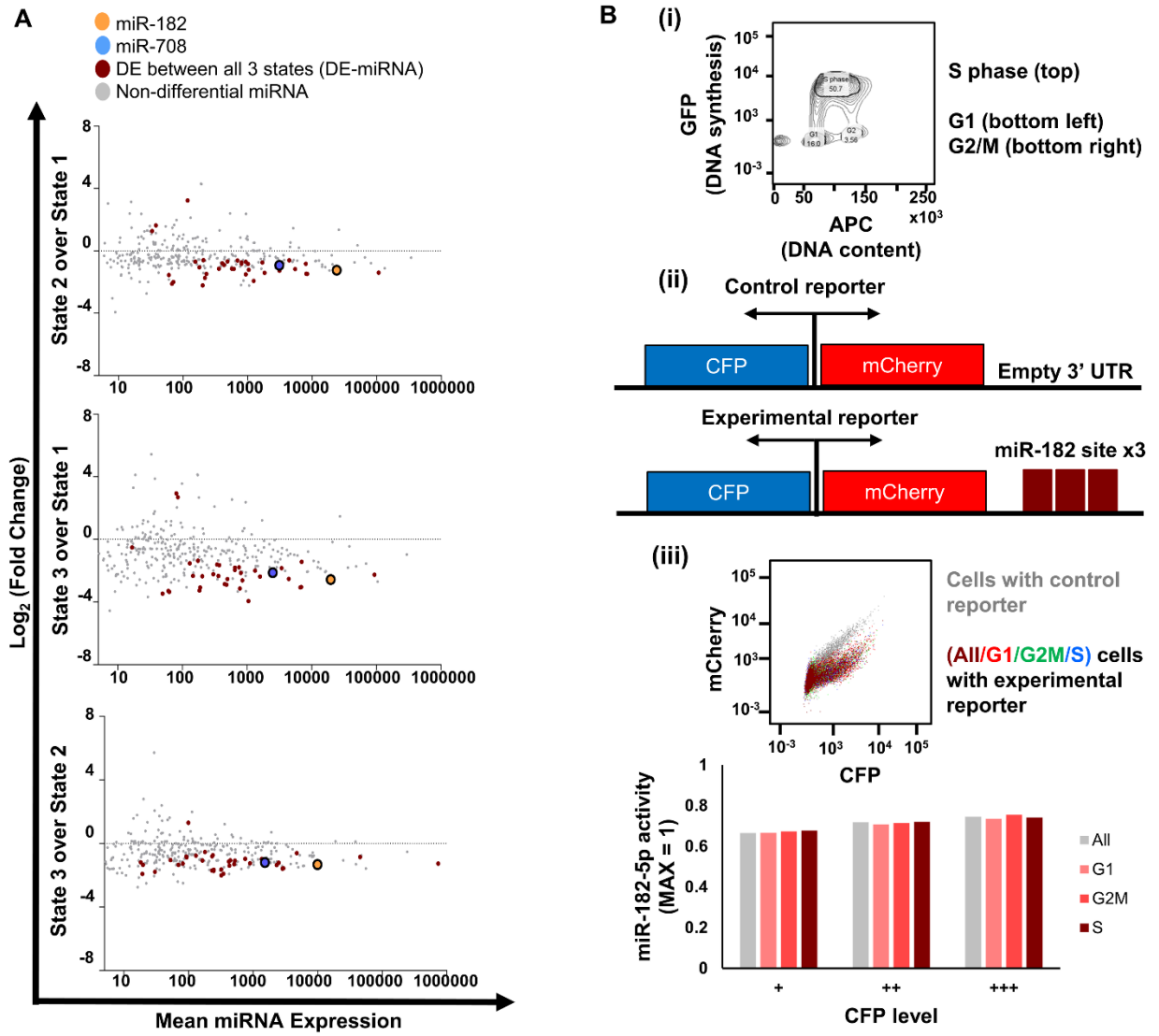

Figure S5: miRNAs differentially expressed across states, and their gene silencing activity across cell cycle stages. (A) Changes between ESC states for miRNAs, determined by small-RNA sequencing of sorted States 1-3. Expected counts versus fold change between states is plotted (MA plot). Highlighted in red are miRNAs with differential expression between all three states (DE-miRNA). MiR-182 and miR-708 are indicated**. (B)** (i) Flow cytometry-based method to distinguish cells by cell cycle stage. Prior to analysis, cells are fluorescently labeled for DNA content (FxCycle Far Red stain, ThermoFisher) and DNA synthesis (Click-It Plus EdU AlexaFluor 488 Flow Cytometry Assay Kit, ThermoFisher). (ii) Plasmid reporter assay for miR-182 activity. In brief, in cells transfected with plasmid, *mCherry* and *CFP* are simultaneously transcribed from a bidirectional promoter. Control reporter plasmids are not modified, whereas experimental reporters contain three miR-182 target sites in the 3' UTR of mCherry. Thus, CFP levels provide a measure of total transfection and induction efficiency, and across CFP levels, gene silencing activity by miR-182 is captured by the reduction in mCherry fluorescence, in cells transfected with the experimental reporter, compared to control. See (3, 4) for further details about this method. (iii) Comparison of miR-182 activity across cell cycle stages, in cells labeled by (i) and transfected with (ii).

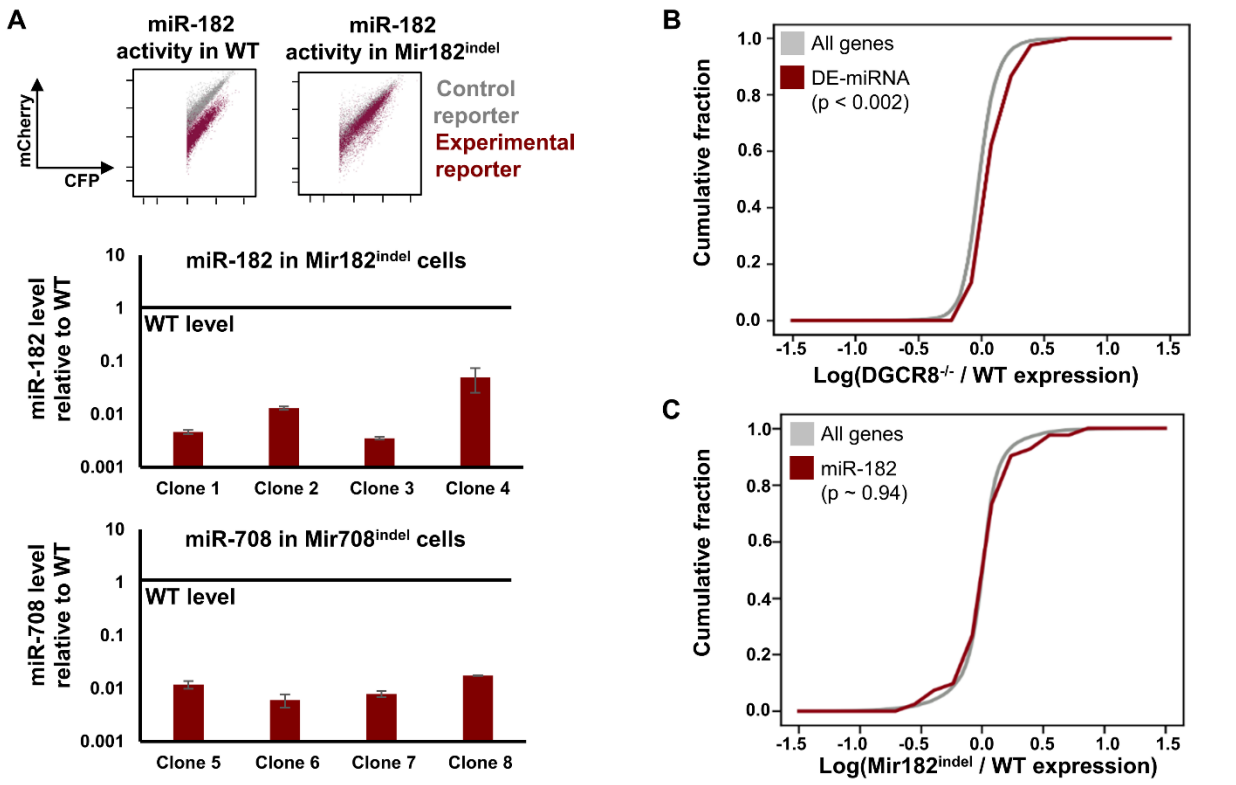

Figure S6: Gene silencing by DE-miRNAs in WT vs. miRNA-deficient cells. (A) Top: Comparing results of miR-182 activity assay (described in S5B(ii)) in WT vs. Mir182indel cells. **Bottom**: miR-182 and miR-708 levels in *Mir182*^{indel} cells and *Mir708*^{indel} cells by quantitative PCR (using the ddCt method relative to WT ESC) respectively. Error bars represent n=3 replicates. Insertion/deletions were verified by sequencing across the hairpin region of the miRNA gene from amplified genomic DNA. Note that clones 3-4 & 7-8 were labeled with fluorophores at *Nanog* and *Sox2* loci, and clones 3 & 7 are those used in Fig. 5 and S8. Clones 1-2 and 5-6 were used in biological replicates for bulk RNA-sequencing of miR-182 and miR-708 targets in *Mir*^{indel} cells. Clone 2 is used for reporter activity assay shown in this figure. **(B)** CDF for the ratio of gene expression in DGCR8$^{-/-}$ ESC vs WT ESC for all genes or genes targeted by ≥ 2 DE-miRNAs. **(C)** Expression of all genes (gray) or genes targeted by miR-182 in WT vs. *Mir182*^{indel} ESC displayed as cumulative distribution function (CDF) over the logarithm of expression ratio between cell types. Three biological replicates for each condition were used in bulk RNA-sequencing (Methods).

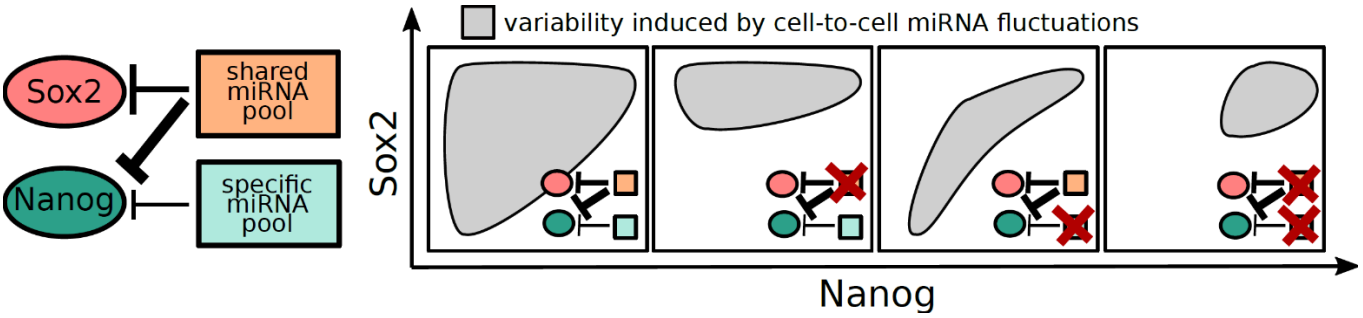

Figure S7: Qualitative model for miRNA impact on state dynamics. Cell state distributions expected for a minimal qualitative model involving Nanog, Sox2, and miRNAs in which cell-to-cell variation is added only in miRNA levels. "Shared" miRNAs regulate both Nanog & Sox2 (Sox2 less strongly than Nanog) while "specific" miRNAs regulate only Nanog. Cell-to-cell variation in miRNA levels in this network induces highly variable expression of Sox2 and Nanog in WT ESC. When the shared pool is knocked out, Sox2 is derepressed and variation mainly affects Nanog expression. Conversely, when the Nanog-specific pool is knocked out, the shared pool tends to correlate its targets focusing variation close to the diagonal axis in the Nanog-Sox2 plane. Finally, knocking all miRNAs out will concentrate cells in State 1 (high Sox2, high Nanog). Hence loss of variable miRNAs generically reduces diversity. Experiments suggest that, while weak in absolute terms (so that most cells reside in State 1), miRNA repression is highly variable across cells due to variation in miRNA expression. In such conditions, the minimal model just discussed will robustly reproduce the experimentally observed states so long as the interaction hierarchy shown above is preserved.

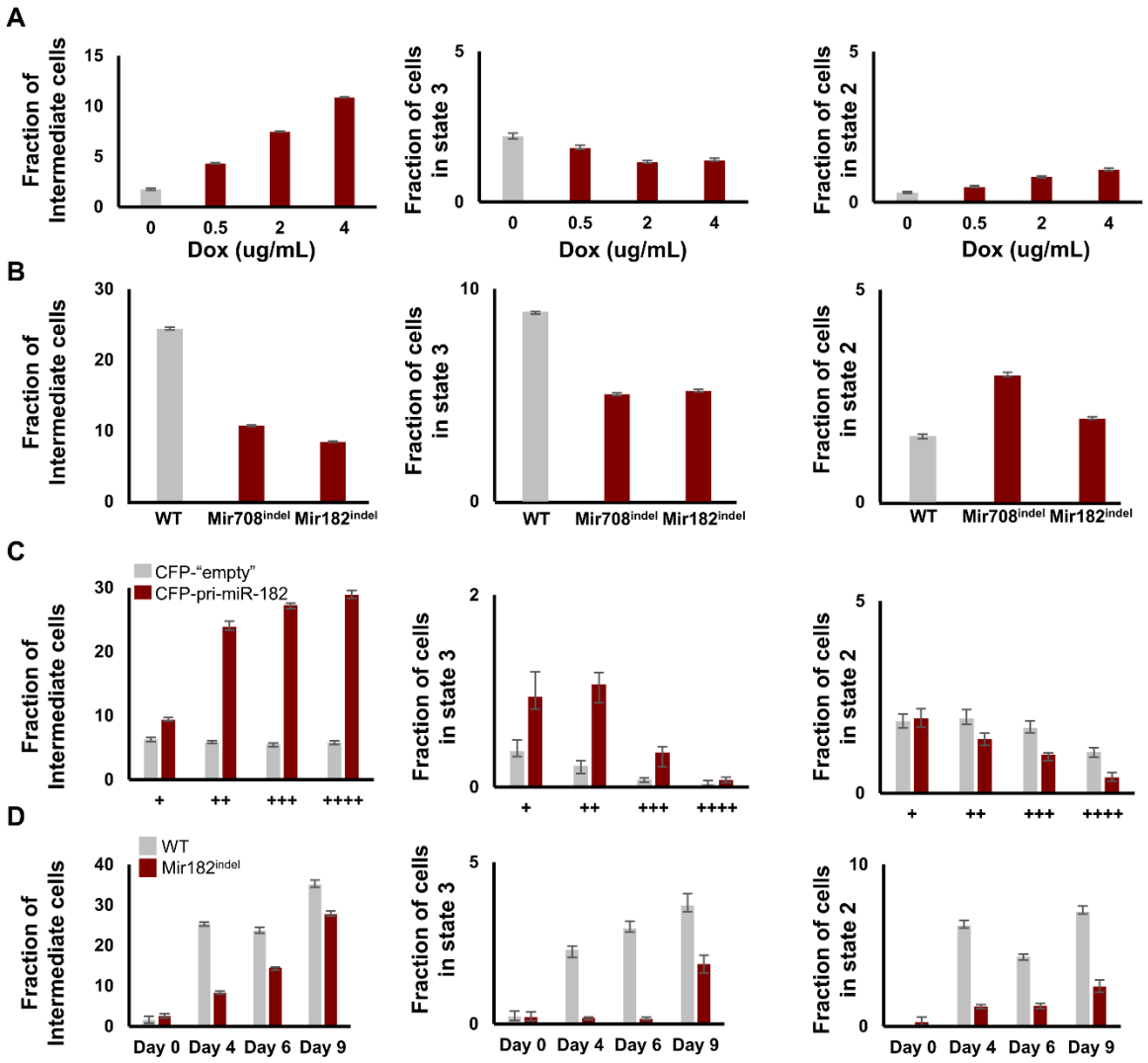

Figure S8: Distributions amongst cell states for variable miRNA reconstituted cells. (A) Fraction of cells in State 2, State 3, and intermediate between States 1-3 (outside of state gating) for Ago2-inducible ESC. All cells were cultured at 1 ug/Ml doxycycline continuously and then switched into the indicated amounts of doxycycline for 48 hours before measuring cell state distributions by flow cytometry. Compare to Fig. 5*A*. Error is 95% CI from bootstrapping (see Methods). **(B)** Fraction of cells in State 2, State 3, and intermediate between States 1-3 for *Mir*^{indel} ESC compared to WT ESC. Compare to Fig. 5*B*. Error is 95% CI from bootstrapping. **(C)** Fraction of cells in State 2, State 3, and intermediate between States 1-3 for *Mir182*^{indel} cells transfected with pri-miR-182 expressing plasmid or transfected with empty control plasmid. Compare to Fig. 5*D*. **(D)** Fraction of cells in State 2, State 3, and intermediate between States 1-3 for WT and *Mir182*^{indel} ESC sorted for State 1 and cultured for the indicated number of days. Compare to Fig. 5*E*.

Supplementary Materials and Methods

Cell line maintenance

Cells were maintained in culture medium on 10cm tissue culture plates pre-coated with 0.2% gelatin in phosphate-buffered saline (PBS). Plates were maintained in a humidified 5% CO_2 incubator at 37 deg. C. Culture medium consisted of: 415 Ml DMEM, 5 Ml 1M HEPES, 5 Ml 0.1 Mm non-essential amino acids, 5 Ml 0.1 Mm P/S antibiotics, 5 Ml 0.1 Mm L-glutamine, 4 Ml 14.3 M beta-mercaptoethanol, 82.5 Ml HyClone fetal bovine serum (FBS), and 55 Ml 1000U/Ml leukemia inhibitory factor (LIF). All components were passed through a sterile filter prior to addition of FBS + LIF. Cells were passaged at a minimum of every two days using the protocol in Supplementary using the following protocol: 1) Wash with 0.5 Ml HEPES-buffered saline (HBS) per Ml culture medium; 2) Detach from plate with 0.05-1 Ml 0.25% trypsin per Ml medium; 3) Inactivate trypsin with culture medium after 1-2 min; 4) Spin detached cells at 233 rcf for 5 min; 5) Resuspend cell pellet in culture medium; 6) Count cells using Trypan Blue staining with a hemocytometer; 7) Re-plate at a density of 25,000-50,000 cells / Ml medium or analyze as described below.

Fluorophore tagging of pluripotency genes

As indicated above, for 9/10 cell lines, fluorophore tags were inserted at the endogenous loci of pluripotency genes. CRISPR-Cas9 induced homology directed repair was used, with a guide RNA targeting immediately upstream of the start codon (*Nanog*) or downstream of the stop codon (*Sox2, Esrrb, Eif2s2, Hsp90ab1*). For each targeted gene, the guide RNA sequence (Supplementary Table 1) was BbsI-cloned into PX330 (Addgene #42230); the modified plasmid was then introduced into cells by cationic lipid transfection (Lipofectamine 2000, Invitrogen) along with a homology-directed repair construct containing the relevant fluorophore, T2A/P2A, and drug resistance (see Fig. S1) following the method of(5). (Note: for the “NLS” cell line, a nuclear localization sequence was added directly upstream of *GFP* for *Nanog* tagging and directly downstream of *P2A* for *Sox2* tagging.) Transfected cells

were selected by drug resistance; PCR was used to confirm insertion of the repair construct at the endogenous locus (primers: Supplementary Table 1).

Generation of Mir^{indel} ESCs

As indicated above, 2/10 cell lines were made miRNA-deficient (termed Mir^{indel}) after fluorophore tagging. CRISPR-Cas9 was used to introduce an indel near the hairpin stem-loop in the relevant miRNA gene. PX330 plasmid containing the relevant guide (Supplementary Table 1) was transiently transfected into cells and single-cell transfectants were isolated by Fluorescence-activated cell sorting (FACS) into individual wells of a 96-well flat-bottom plate. PCR was used to confirm the presence of an indel (primers: Supplementary Table 1). Reduction in mature miRNA production was confirmed using RNA extraction by TriZOL® and RT-Qpcr using the Mir-X miRNA First-Strand Synthesis Kit (Takara) for miR-182 and the miScript System (Qiagen 218161 and 218073) for miR-708.

Reintroduction of miRNAs into deficient ESCs

MiRNA re-expression in *Mir^{indel}* cells was performed as follows. First, bidirectional reporter constructs were generated (see above) in which *mCherry* was replaced with the *pri-miRNA* sequence. Primers used to amplify *pri-miRNA* for cloning into Ptre-BI-Tight were: AATCGGATCCTCACTGCCTAATGCCCCTAC and GCAAAAGCTTAGCCATCTGTCTCTCCCTCA (for miR-182, using HindIII and KpnI restriction sites to replace *mCherry*) and AATCGGATCCTGAATAGCCAATGAAAATGACTTG and ATTGAAGCTTCAAGCCCAGGAGTTGAAGAG (for miR-708 using HindIII and BamHI restriction sites, also replacing *mCherry*). The use of the bi-directional expression plasmid allowed detection of pri-miRNA expression by CFP fluorophore expression. These vectors were then separately transfected into *Mir^{indel}* cells along with rtTA plasmid and the distribution of cells into States was determined by flow cytometry as described above. Restoration of the relevant miRNA to at or above WT levels was confirmed using RNA extraction by TriZOL® and RT-Qpcr using the Mir-X miRNA First-Strand Synthesis Kit (Takara) for miR-182 and the miScript System (Qiagen 218161 and 218073) for miR-708.

Image Acquisition for Fluorescence Microscopy

V6.5 ESC (*NLS-GFP-P2A-Nanog*, *NLS-Sox2-T2A-mCherry)* cells were sorted into States 1-3 using BD FACSAria and seeded in gelatinized µ-Slide 8 Well (Ibidi catalogue #80826) at 2000 cells/well in ESC media for State 1, State 2, and Unsorted conditions and 2400cells/well for State. Cells were transferred to Dpbs (Corning, 21-031-CV) when imaging. Cell were imaged 3, 24, 48, and 72 hours after plating on a Axio Observer 7 (Zeiss) with Axiocam 702 Mono camera (Zeiss), Solid-State Light Source Colibri 7, Type R[G/Y]CBV-UV (Zeiss 423052-9741-000) on a 20x/.8 M27, (.29µm/px) objective using Zen blue software. Images were stitched into an 8x10 tiled plane using Zen Blue during acquisition. Acquisition settings were maintained constant across all samples.

Image Processing for Fluorescence Microscopy

Image processing was done in Fiji (http://fiji.sc). Each fluorescent channel was rolling ball background subtracted (50px, disable smoothing), despeckled, masked using size exclusion on a lower size threshold. The upper threshold for brightness and contrast were auto set for each image except for the Day0 (3 hr) images in Fig. 1*C*, which were set to the same upper bound as defined by Day0 State1 to facilitate comparison. Lower signal bounds were set manually for each image to remove background signal.

Molecular Barcoding of ESC

ESC were barcoded using the ClonTracer library (Addgene #67267) described previously(6). Lentivirally-encoded barcodes were transduced into *Nanog-GFP/Sox2-mCerulean3* ESC using spinoculation and selected by FACS sorting for RFP expression. Transduced library cells were cultured together in 10 cm plates or separately as single cell clones in 96-well plates (Fig. 1*E*) for the indicated time periods. Pooled genomic DNA was extracted using Sigma GenElute Mammalian Genomic DNA Prep Kit (Catalog #G1N70) and PCR amplified according to the protocol provided by(6) on the Addgene website. Amplicons were sequenced by Sanger sequencing for single cell clones.

RNA-sequencing analysis pipeline and data plotting

For complete details, including commands used, please see the "RNA-seq analysis pipeline and data plotting" document at https://bit.ly/2qfUSLQ. In brief, RNA-sequencing reads were trimmed before alignment to the the mm10 genome, obtainment of RSEM counts, and differential expression analysis using EBSeq. The 'data plotting' section explains the generation of RNA-seq-based MA plots, CV-mean plots, scatter plots, and heatmaps.

Gene Ontology Analysis

Lists of genes differentially expressed between 3 states were inputted into GOrilla (http://cbl-gorilla.cs.technion.ac.il/) on 1-11-18 to determine ontologies highest represented in this group. Display in Fig. 2*B* corresponds to reduction of redundant terms. Fig. S3*C* displays CV-mean plot for selected gene ontologies (GO: 0048646 and 0051726) chosen to represent lineage regulators and cell cycle genes respectively of similar set size and expression distribution. To determine ontology representations of each state, we took the top 300 expressed protein coding genes in each state and analyzed by GOrilla. Fig. S3*F* represents z-score of FDR q-value for representative ontology terms in each state. Selected terms from amongst the strongest processes for each state are shown in Fig. 2*C*.

Gene neighborhood construction after single cell RNA-sequencing

To ensure robust inferred neighborhoods based on correlation, cells with fewer than 5,000 uniquely captured transcripts were discarded from analysis, yielding 2,299 cells remaining. We then restricted our analysis to the best sampled genes, keeping those with a total of >500 counts across all remaining cells OR with >4 counts in any one cell to account for rare cells with high expression of an individual transcript. Additionally, analysis comparing WT ESC and *Mir182*^{indel} ESC are restricted to 6,107 genes captured by these criteria in both cell types. Data tables were then processed by total count normalization(7) prior to neighborhood analysis. Raw fastq files, count tables and merged count tables, and normalized tables are provided at https://bit.ly/2qfUSLQ.

To construct interaction neighborhoods, we utilized a topology-based method originally based on(8) and similar to(7). We reasoned similarly to the latter authors that such a method would minimize artifacts from technical sampling noise in scRNA-seq. In brief, a node gene of interest is selected (G_0 gene). The 50 most correlated (Pearson r correlation coefficient) genes with this node are then selected (G_1 set), and then the 50 most correlated with each of these G1 genes is selected (50 G_2 sets for every G_1 set). Thus an initial network is created with 2550 directed edges (G_0->G_1 or G_1- >G_2). We then trim the network by iteratively removing any genes with < 10 incoming edges (from G_0 or G_1 genes) until we are left with a final network that must include G_0 and in which all genes have ≥ 10 incoming edges. We found relatively little difference in final networks when modifying the size of G_1/G_2 or the trimming parameters.

Single Cell Variation (ν) Score Calculation

We calculated a "variability" score to describe the variation in the expression of a gene across a population of cells, based on a test statistic described in (7). For any given mean expression level, this test statistic weights genes whose coefficient of variation is significantly larger than a Poisson random variable with the same mean. The test statistic, ν, is:

$$v = \frac{CV_{\hat{m}}^2}{\frac{(1 + CV_M^2)\left(1 + CV_{\frac{1}{N}}^2\right)}{E_{[\hat{m}]}} + CV_{\frac{1}{N}}^2}$$

where $\hat{m}$ is normalized read counts (total count normalization), M is total number of reads and N is total Mrna content. The additive constant noise $CV_{\frac{1}{N}}^2$ weights against genes whose variation in expression is largely due to variation in differences in cell size. In this study, we use $CV_{\frac{1}{N}}^2$ of 0.25.

Single Cell State Analysis (PHATE mapping, PCA, Force Directed Graphs)

Single Cell state representations were made primarily using the PHATE(9) and ScanPy(10) packages (Python 3.7). A full script with detailed commands used to produce plots in Fig. 1*F* is given at: https://bit.ly/2qfUSLQ in /scripts/Seurat_Python2.py. In brief, the combined (WT and *Mir182*^{indel}), mapped single cell expression tables were normalized and then analyzed in PHATE using default parameters to power the PHATE operator. Cells were pruned by removing those with the top 5% of mitochondrial reads and transformed by square root. The total number of counts and percent of mitochondrial reads was regressed out using ScanPy prior to applying the PHATE operator. PCA and force directed graphs were generated using built-in ScanPy functions using default parameters in analogy to the methods shown in (11). Cells were colored according to their Z-score for State 1-3 genes in the following way. First, the top300 genes uniquely most highly expressed in each state (see GO analysis or Heatmap analysis) are defined. Then, for each gene, a Z-score for its expression across cells is calculated for each cell. Next, the Z-scores for all 300 State 1 genes are summed together for each cell, as are the Z-scores for all 300 State 2 genes and State 3 genes. These summed Z-scores are then re-scaled across all cells to the range [0,1] for each state. Thus, the cell with the 'most State 1 character' is given a State1Z-score of 1.0 and that with the least such character is given a score of 0. These scores were then used for blue coloration in Fig. 1*F* and Fig. S3*D*.

miRNA target analysis

Genomic targets of miRNAs were determined primarily based on crosslinking and immunoprecipitation (Iclip) experiments previously described (12). This analysis identified 6,816 high confidence "clusters" of Argonaute 2 binding. The list of the Iclip clusters with genomic locations and nucleotide sequences is given at https://bit.ly/2qfUSLQ ("miRNA targets" → "Ago2clusters.txt"). "Stacked 5P sequencing ends" indicate the site of crosslinking and putative exact location of miRNA binding within each cluster (12). In "Ago2clusters.txt," the notation in the "Stacked 5P end" column is as follows: -_1 indicates no observed stacked 5P end; otherwise, position of the stacked 5P end, with respect to the cluster "start" location, is

given, followed by an underscore and the number of reads at that position. Multiple putative stacked 5P ends are separated by semicolons prior to the underscore.

We assumed that each distinct Ago2 cluster resulted from the binding of one particular miRNA. To determine which miRNA gave rise to each cluster, we first made two assumptions: 1) if the cluster had stacked 5P end reads, it most likely arose from the activity of a miRNA with the highest affinity target site (seed match) immediately upstream of the stacked 5P end location (12) and 2) if the cluster did not have a stacked 5P end, it arose from the activity of the highest expressed miRNA with a predicted seed match to a target site anywhere in the cluster. Higher affinity '8-mer' matches between Mrna target and miRNA were prioritized over lower affinity '6/7-mer' matches, and higher expression miRNAs were prioritized over lower expression miRNAs when no evidence for strong affinity-based interaction was present. MiRNAs with identical '7-mer' seed region sequence were grouped for all analyses. Commented code for generating miRNA-Mrna assignments is available at https://bit.ly/2qfUSLQ ("Code for making biochemical miRNA-target map.py") as is the resulting miRNA-target map ("Biochemical miRNA-target map.txt"). Note that the map only predicts targets for the first 250 miRNAs in "Information about miRNA families.txt." These are the top 250 expressed miRNAs in the *Nanog-GFP/Sox2-mCherry* ESCs used in this study.

MiRNA enrichment in neighborhoods

To determine miRNA enrichment within a given gene network neighborhood (Network N_0) (Fig. S4*E*), we first listed the total number of binding events for each of the top 250 miRNA seed families within all neighbors. Next, we constructed 10,000 control neighborhoods (N_1 set), each of which had an equal number of genes to N_0. Each N_1 control neighborhood was also chosen to contain: roughly the same average gene expression of the same expression distribution as N_0, the same number of highly Ago2 bound genes as N_0 (genes with >5 clusters of bound Ago2-miRNA), and roughly the same number of total miRNA binding events as N_0. These simulated, control neighborhoods (N_1) thus represented a

closely matched group to the original gene network neighborhood under interrogation. For each of the top 250 miRNA seed families, we then empirically determined the total number of binding events in each N_1 control neighborhood. A miRNA seed family was enriched in a N_0 gene network neighborhood if 500 out of 10,000 N_1 control neighborhoods or fewer (empirical p-value 0.05) contained a total number of miRNA binding sites ≥ the total number observed in the N_0 gene network neighborhood. Given the relative rarity of genes bound by > 5 miRNAs and the stringency of these parameters, many N_1 control neighborhoods shared considerable overlap with the N_0 gene network neighborhood under interrogation, giving us a conservative list of enriched miRNAs for each N_0. Next, we defined groups of genes such as pluripotency (taken from reference (7), 57 total expressed in our scRNA-seq data with non-zero Ago2 binding in their network), cytoskeleton (GO: 0007010, 192 expressed), or highly variable genes (defined by □ test-statistic > 3.0 in our scRNA-seq data, 109 genes). For each of the top 250 miRNA seed families, we totaled the number of times it was enriched in these sets of N_0 neighborhoods and plotted (Fig. S4*E*) or analyzed for overlaps in enrichment between sets of N_0 neighborhoods.

Flow cytometry-based method for distinguishing between cell cycle stages

Prior to flow cytometric analysis, cells were fluorescently labeled for DNA content, using FxCycle Far Red stain (ThermoFisher). They were also labeled for DNA synthesis, using the Click-It Plus EdU AlexaFluor 488 Flow Cytometry Assay Kit (ThermoFisher). Cell cycle stages were distinguished as follows (see Fig. S5*B*): G1 phase = low DNA content, low DNA synthesis; S phase = high DNA content, high DNA synthesis; G2 phase = high DNA content, low DNA synthesis.

Exact cell numbers, Figs. 5 and S8

Fig. 5*A* / S8*A* (Dox 0, 0.5, 2, and 4 ng/mL): 257254, 278817, 306193, 488957

Fig. 5*B* / S8*B* (WT, *Mir708*^{indel}, *Mir182*^{indel}): 257229, 299486, 383833

Fig. 5*D* / S8*C* ('+' through '++++'; empty/pri-miR-182): 9793/9268, 12087/11390, 15345/21585, 14735/10794

Fig. 5*E* / S8*D* (Days 0, 4, 6, and 9; WT/*Mir182*^{indel}): 1752/1905, 20219/20334, 21604/20798, 17022/13556

Bidirectional reporters for miRNA activity

Activity of individual miRNAs in ESCs was determined using a bidirectional reporter system similar to previously reported (3) (Figs. S5*B*, S6*A*). In brief, reporters were derived from the Takara Ptre-Tight-BI plasmid in which two fluorophores are expressed under tight co-transcriptional control under regulation by TET-on promoter system. The reporters used in this study are derived from (12) and contain mCherry fluorescence as a readout of miRNA activity, but are modified to contain *mCerulean3 (CFP)* substituted for *YFP* due to an observed "tighter" correlation between mCerulean3 and mCherry than between YFP or ZsGreen and mCherry. Additionally, the *mCherry* UTR contains a small insertion to introduce a SpeI restriction site for cloning. MiRNA activity is detected via insertion of target sites for the relevant miRNA into the 3' UTR of *mCherry* using mCerulean3 as a normalization control. We chose to insert three perfect matches to each miRNA, separated by spacer sequences (cloned using ClaI and SpeI). Spacer sequences: Upstream of each miRNA site – CTGGGCACCAACTCAACTTC, Downstream of each miRNA site – ACAACTTGGTGTGTTAGTGT. Activity of a given miRNA is determined by comparing mCherry expression in cells transiently transfected (Lipofectamine, see above) with a control plasmid (i.e., one with no miRNA sites in the *mCherry* 3' UTR) to mCherry expression in cells transiently transfected with a plasmid with miRNA 3X sites in the 3' UTR (additionally, cells are cotransfected with rtTA plasmid to engage the Tet system). Reduced mCherry expression compared to control indicates miRNA activity.

Supplementary References:

1. M. N. Shahbazi *et al.*, Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* **552**, 239-243 (2017).
2. J. R. Zamudio, T. J. Kelly, P. A. Sharp, Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. *Cell* **156**, 920-934 (2014).
3. S. Mukherji *et al.*, MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854-859 (2011).
4. J. M. Schmiedel *et al.*, Gene expression. MicroRNA control of protein expression noise. *Science* **348**, 128-132 (2015).
5. J. Stewart-Ornstein, G. Lahav, Dynamics of CDKN1A in Single Cells Defined by an Endogenous Fluorescent Tagging Toolkit. *Cell Rep* **14**, 1800-1811 (2016).
6. H. E. Bhang *et al.*, Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21**, 440-448 (2015).
7. A. M. Klein *et al.*, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201 (2015).
8. A. Li, S. Horvath, Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics* **23**, 222-231 (2007).
9. K. R. Moon *et al.*, Visualizing Structure and Transitions for Biological Data Exploration. *bioRxiv* (2019).
10. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
11. F. Paul *et al.*, Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663-1677 (2015).
12. A. D. Bosson, J. R. Zamudio, P. A. Sharp, Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell* **56**, 347-359 (2014).