nature cell biology

SUPPLEMENTARY INFORMATION

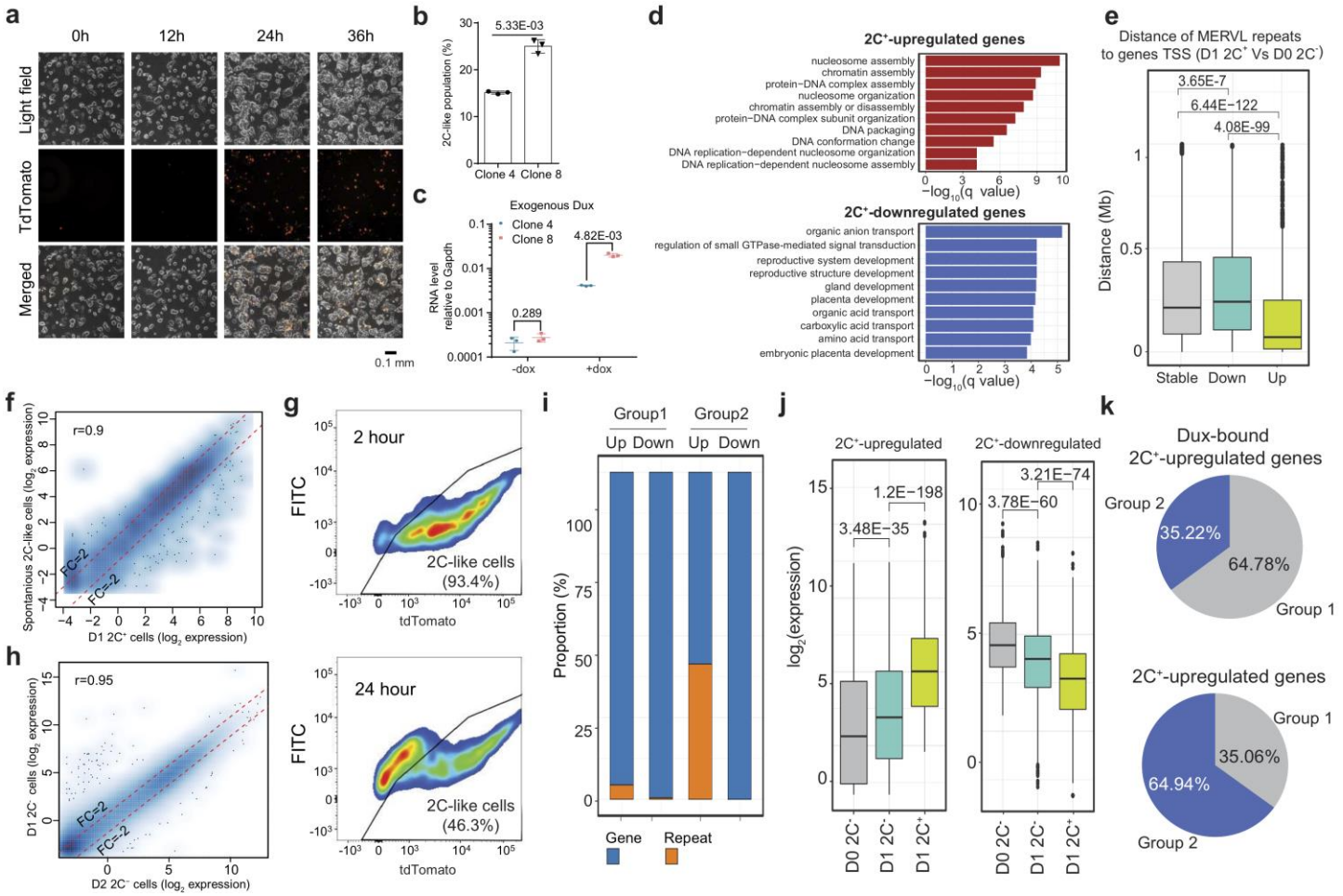
https://doi.org/10.1038/s41556-019-0343-0

In the format provided by the authors and unedited.

Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells

Xudong Fu[1,2,3,4,5], Xiaoji Wu[1,2,3,4,5], Mohamed Nadhir Djekidel[1,2,3,4,5] and Yi Zhang[1,2,3,4*]

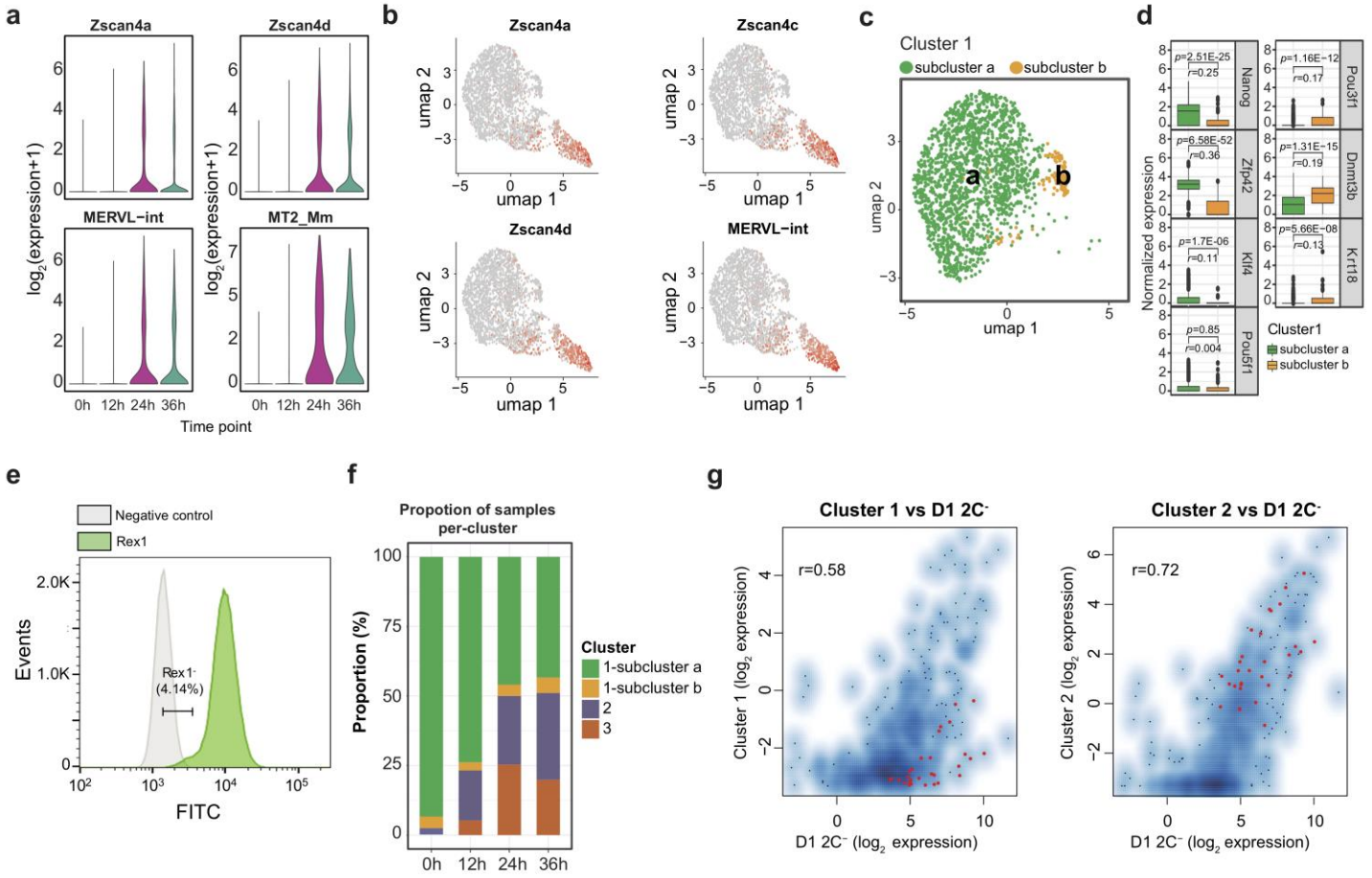
[1]Howard Hughes Medical Institute, Boston, MA, USA. [2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA. [3]Harvard Stem Cell Institute, Boston, MA, USA. [4]Department of Genetics, Harvard Medical School, Boston, MA, USA. [5]These authors contributed equally: Xudong Fu, Xiaoji Wu, Mohamed Nadhir Djekidel. *e-mail: yzhang@genetics.med.harvard.edu



Supplementary Figure 1

Transcriptomic anlysis indicated the existance of an intermediate cell state during ESC to 2C-like cell transition.

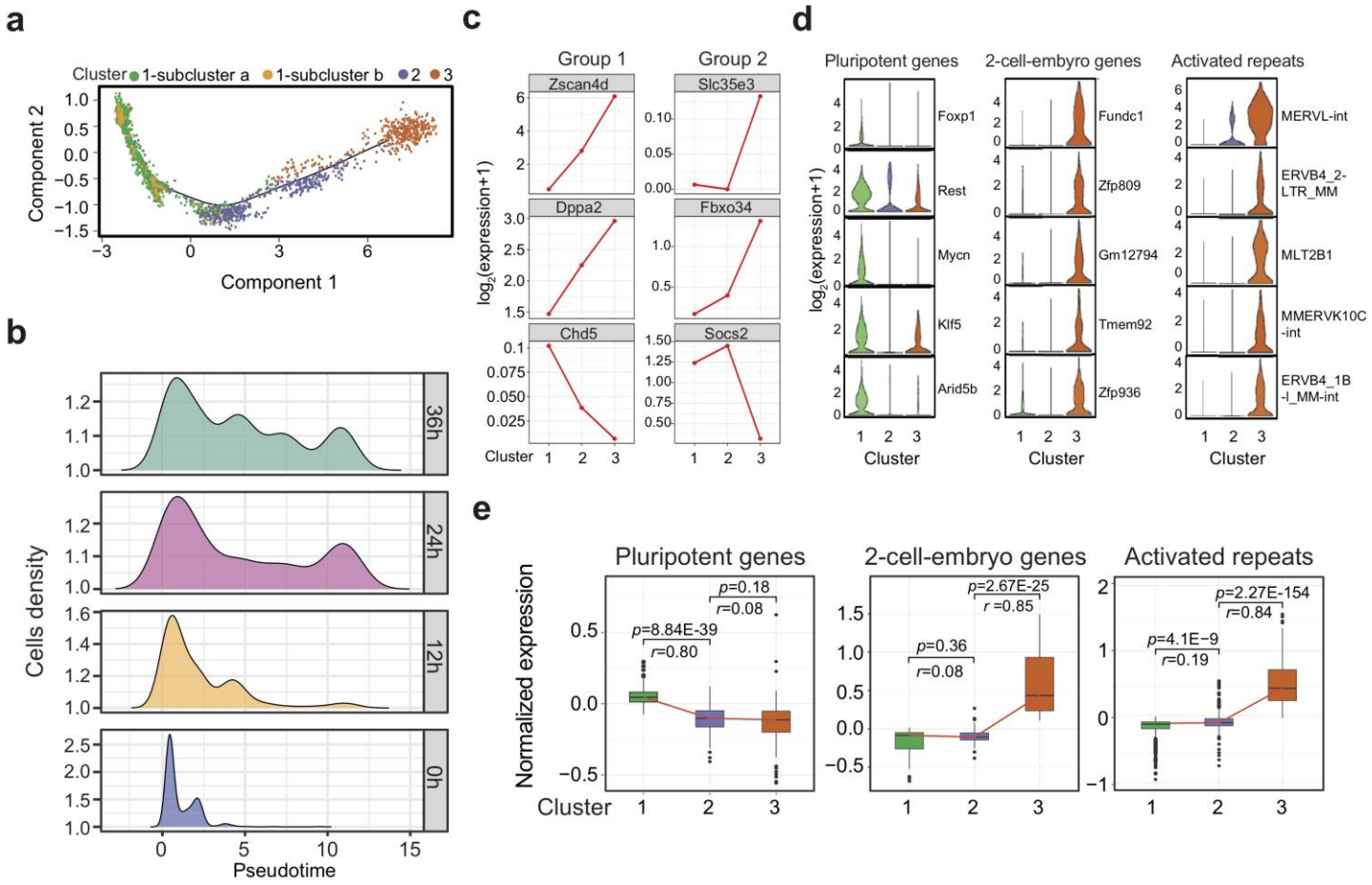
a, Images showing *Dux* induction can gradually induce 2C-like cells generation in cultured ESCs. Scale bar: 0.1 mm. Experiment was repeated independently twice with similar results. **b,** The 2C-like cell percentage of two independent ESC clones after one day of *Dux* induction. **c,** *Dux* transgene mRNA level relative to *Gapdh* in two independent ESC clones. **b-c,** Shown are mean ± SD, n=3 biologically independent cell cultures. *P* values (indicated as numbers in the graphs) are calculated by unpaired t-test, two-tailed, two-sample unequal variance. Experiments were repeated independently twice with similar results. **d,** Bar plots showing the -$\log_{10}$(q value) (right-tailed Fisher-exact test) of the GO terms of upregulated and downregulated genes in 2C-like cells. **e,** Box plot showing the distance of the nearest gene TSS (n=Stable: 24,739, Down: 2,466, Up: 2,146 genes) to MERVL repeats (median: Stable: ~213kb, Down: ~242.5kb, Up: ~93kb). **f,** Scatter plot comparing 2-cell-embryo-specific genes expression profile between spontaneous 2C-like cells and Dux-induced 2C-like cells (Pearson correlation, r=0.9 , n=1,400 genes). **g,** The percentage of 2C-like cells after 2-hour and 24-hour culture of D1 2C$^+$ cells. **h,** Scatter plot comparing gene expression profile between cells which exit from Dux-induced 2C-like state (D2 2C$^-$ cells) and D1 2C$^-$ cells, (Pearson correlation, r=0.95, n=27,221 genes/repeats), supporting that Dux-induced 2C-like cells can exit from 2C-like state spontanesouly. **i,** Relative percentage of genes and repeats in each category. **j,** Box plot showing the $\log_2$ expression level of 2C$^+$-upregulated genes/repeats (n=2,976 genes/repeats) and 2C$^+$-downregulated genes/repeats (n=2,726 genes/repeats) in each of the D0 2C$^-$, D1 2C$^-$, and D1 2C$^+$ cell populations. **e, j,** *P* values were calculated by two-tailed mann-whitney *U*-test. The black central line is the median, boxes limits indicate the upper and lower quartiles, whiskers indicate the 1.5 interquartile range, dots represent outliers. **k,** Pie chart reporting the percentage of Group1 and Group2 genes in 2C$^+$-upregulated genes and Dux bound 2C$^+$-upregulated genes. **d-k** were based on two independent biological replicates of RNA-seq experiments. The effect-size of each statistical analysis and statistical source data can be found in Supplementary Table 10.



Supplementary Figure 2

Single-cell transcriptomic analyses revealed an intermediate state from ESC to 2C-like cell transition.

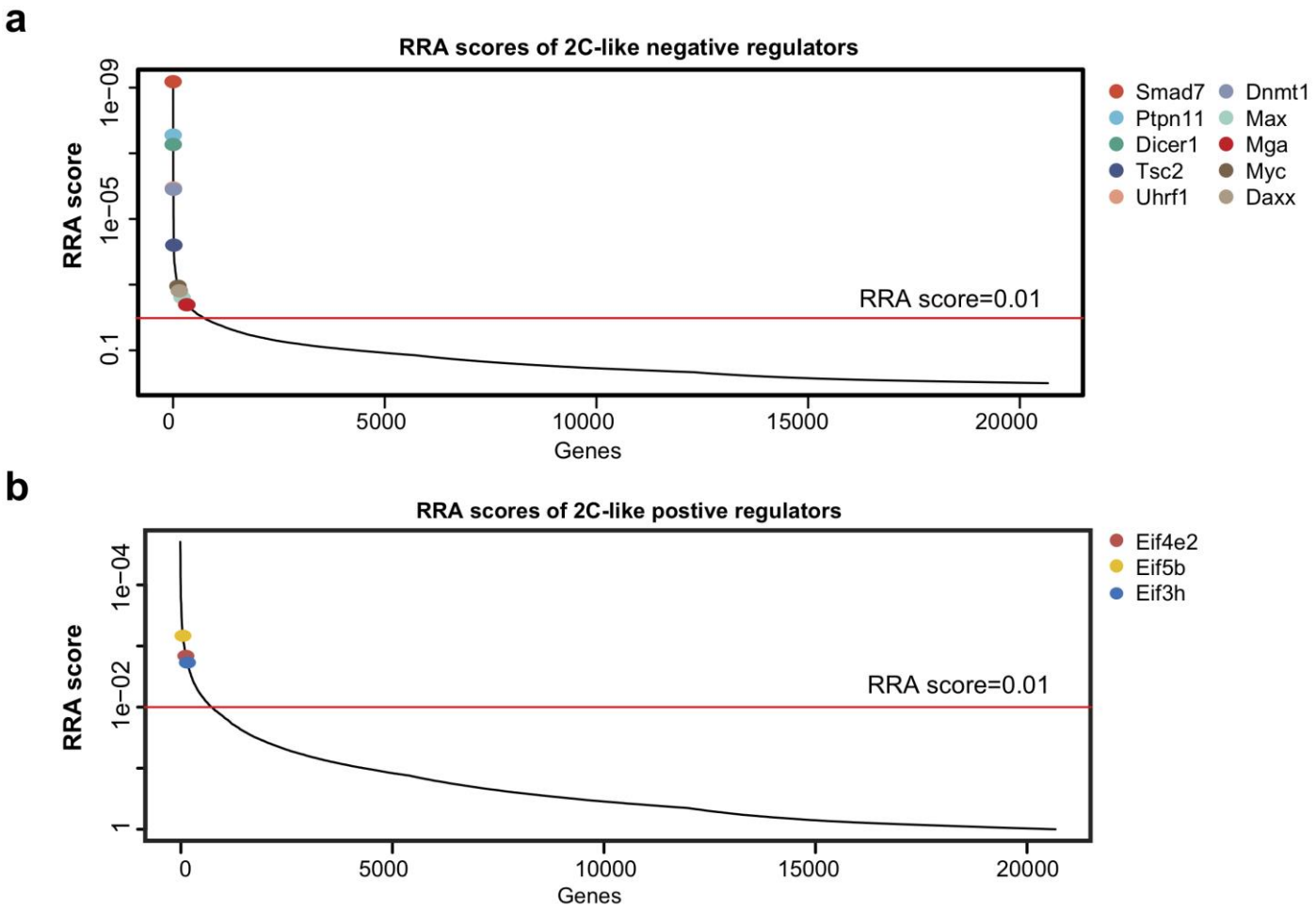
a, Violin plot showing the $\log_2$ expression of representative genes or repeats in each time point (n= 0h: 738 , 12h: 456 cells, 24h: 568, 36h: 871). Violin plots show the kernel density estimation for the distribution of the gene expression in each cluster. The width of the plot represents the proportion of data with the corresponding expression value. **b,** UMAP plot showing the expression of representative genes in each sequenced cells (n=2,633 cells). **c,** UMAP plot showing the subpopulations of of cluster 1 cells (n=1,780 cells). Around 7% pluripotent cells (n=115 cells) belong to the minor subcluster b. **d,** Box plots comparing the gene expression between subcluster a (n=1,665 cells) and subcluster b cells (n=115 cells). The black central line is the median, boxes limits indicate the upper and lower quartiles, whiskers indicate the 1.5 interquartile range, dots represent outliers. *P* values (shown as p) are calculated by two-tailed Mann-Whitney *U*-test and effect-sizes (shown as r) are calculated as $Z/\sqrt{N}$ where Z is the *Z* value of the Mann–Whitney *U*-test and N is the number of samples. **e,** FACS histogram plot showing the expression of Rex1 in mESC culture. About 5% of cells are Rex1-negative. Experiment was repeated independently three times with similar results. **f,** Bar plot showing the proportion of the different cell clusters at different time point of *Dux* induction. **g,** The partial 2-cell embyro-specfic genes and elements activated in D1 2C$^-$ cells leads to less dramatic transcriptional variability between D1 2C$^-$ cells and cluster 2 cells compared to that of cluster 1 cells (n=470 genes/repeats). The red dots show the top 30 genes/repeats that define the variance between D1 2C$^-$ and cluster 1 cells. Correlation was calculated using spearman correlation. Statistical source data can be found in Supplementary Table 10.



Supplementary Figure 3

Single-cell transcriptomic analyses revealed the transcriptional dynamics from ESC to 2C-like cell transition.

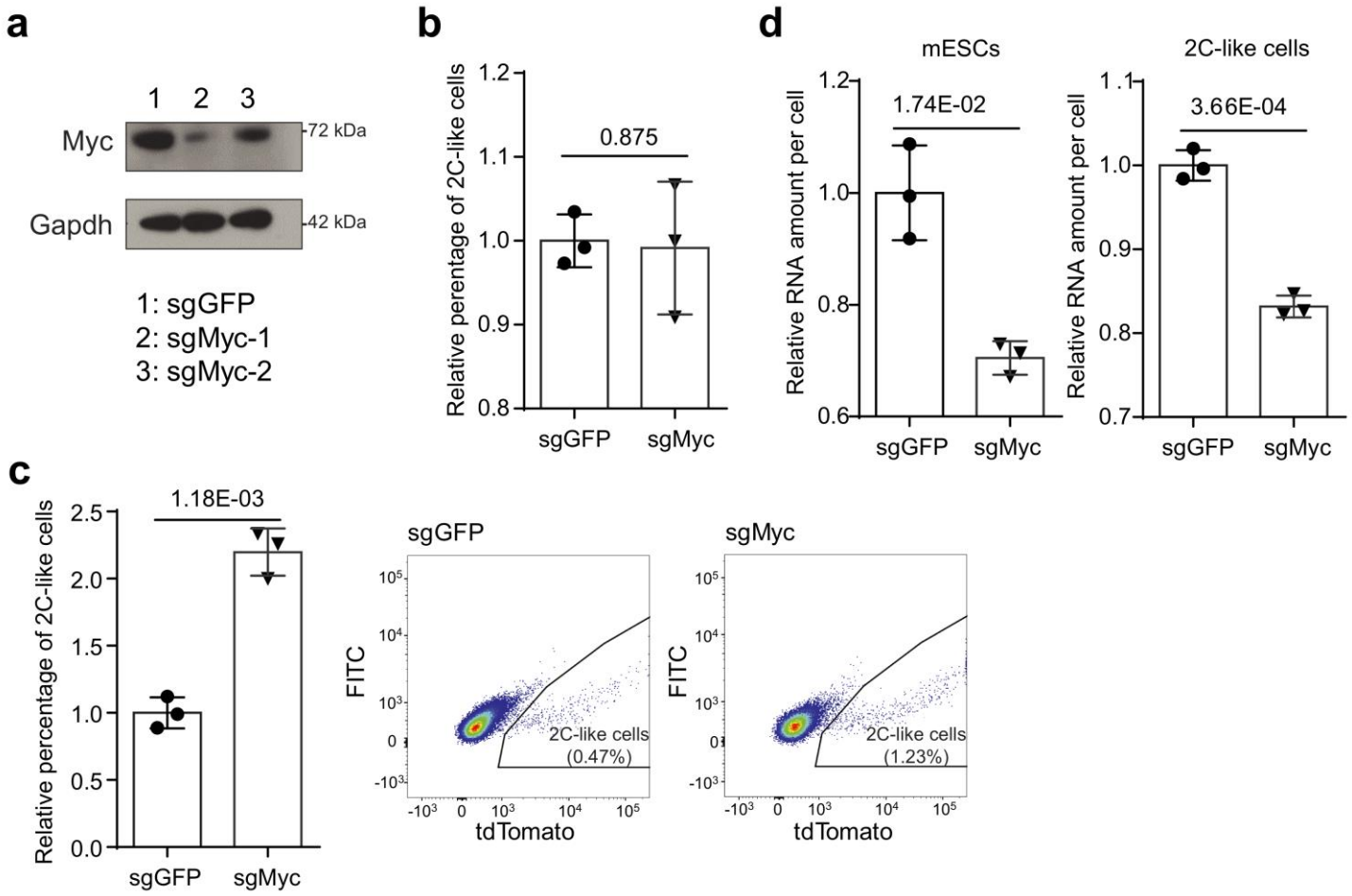
a, Scatter plots showing cells (n=2,633 cells) along the projected pseudo-time following the dynamics of the ESC to 2C-like cell transition. **b,** Cell distribution along with the pseudotime in each time point. Cells collected at the early hours of *Dux*-induction (0 h and 12 h) are located at the beginning of the timeline while cells collected at the late hours (24 h and 36 h) are gradually located towards the end, supporting that the projected timeline captures the transcriptional changes during Dux-induced 2C-like transition. **c,** Mean expression of representative genes in each cell cluster. **d,** Violin plot showing the expression of representative transcripts in each cell cluster (n=1,780 cells, 512 cells and 341 cells, respectively, for clusters 1, 2, and 3). Violin plots show the kernel density estimation of the distribution for the expression of the genes in each cluster. The width of the plot represents the proportion of data with the corresponding expression value. **e,** Box plot showing the expression of the detected $2C^{+}$-downregulated pluripotent genes (n=135 genes), $2C^{+}$-upregulated 2-cell-embyro-specific genes (n=75 genes) and activated repeats (n=501 repeats) in each cell cluster when analyzed after subcluster b cells are excluded. The black central line is the median, boxes limits indicate the upper and lower quartiles, whiskers indicate the 1.5 interquartile range, dots represent outliers. *P* values (shown as p) are calculated by two-tailed Mann-Whitney *U*-test and effect-sizes (shown as r) are calculated as $Z/\sqrt{N}$ where Z is the z-value of the Mann–Whitney *U*-test and N is the number of samples. Statistical source data can be found in Supplementary Table 10.



Supplementary Figure 4

CRISPR-Cas9 screen identified novel regulators mediating ESC to 2C-like transition.

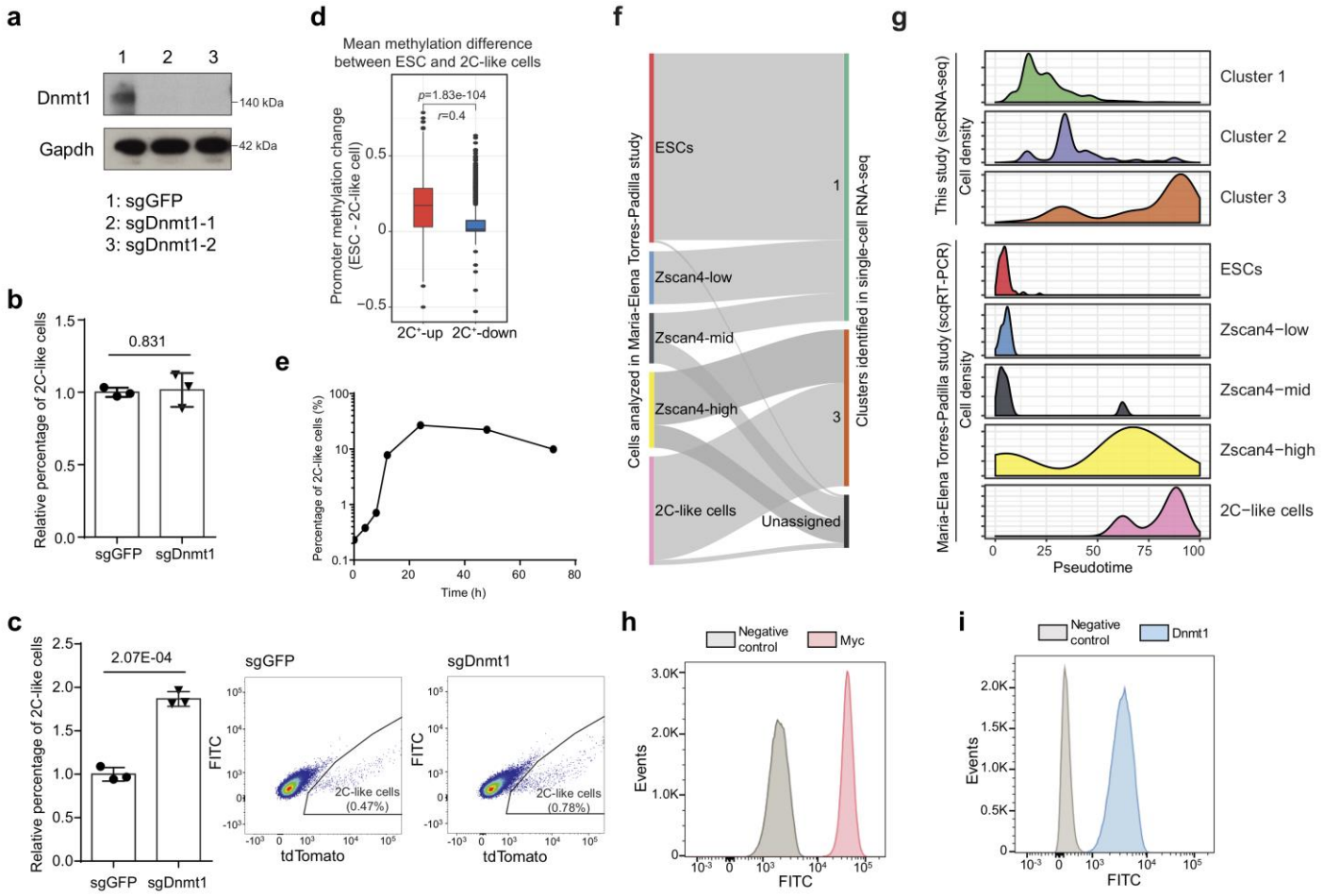
a-b, RRA score as calculated by MAGeCK tool of negative effect **(a)** and positive effect **(b)** of genes on the ESC to 2C-like cell transition from the screen. The results were based on two biologically independent screens.



Supplementary Figure 5

Myc perturbation facilitats ESC to 2C-like cell transition.

a, Western-blot showing the sgRNA efficiency. Unprocessed raw blots can be found in Supplementary Figure 7. **b,** Relative percentage of 2C-like cells after 24-hour culture of D1 $2C^+$ cells isolated from ESC infected with sgGFP or sgMyc. Shown are mean ± SD, n=3 biologically independent cell cultures. **c,** Relative percentage of spontaneous 2C-like cells and representative FACS plot of ESC culture. Shown are mean ± SD, n=3 biologically independent cell cultures. **d,** Myc knockdown decreases total RNA amount per cell in ESCs (left panel) and D1 $2C^+$ cells (right panel), confirming the transcrition amplification effect of Myc in mESCs. Shown are mean ± SD, n=3 biologically independent cell cultures. **b-d,** *P* values (indicated as numbers in the graphs) are calculated by unpaired t-test, two-tailed, two-sample unequal variance. Statistical source data can be found in Supplementary Table 10. **a-d,** experiments were repeated independently twice with similar results.

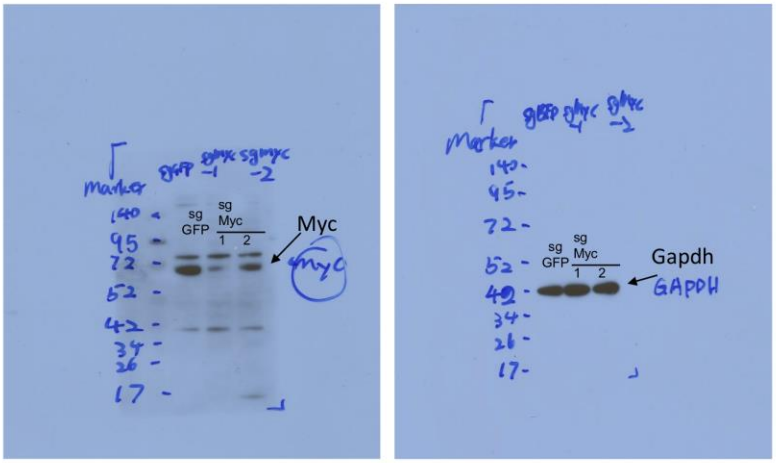


Supplementary Figure 6

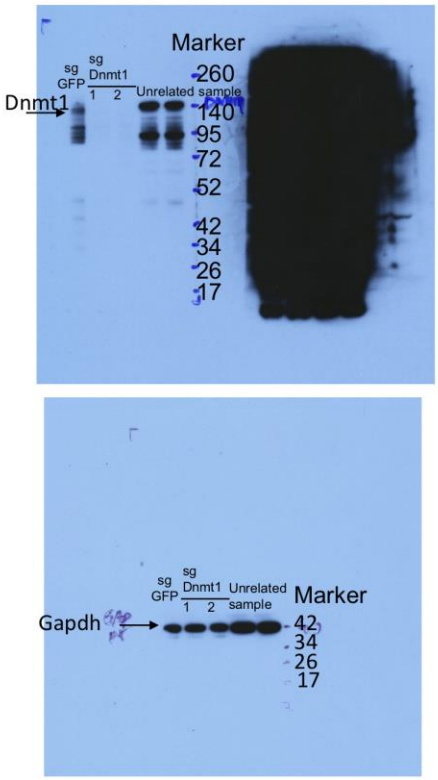
Dnmt1 and Myc perturbation facilitates ESC to 2C-like cell transition.

a, Western-blot showing the sgRNA efficiency. Unprocessed raw blots can be found in Supplementary Figure 7. **b,** Relative percentage of 2C-like cells after 24-hour culture of D1 $2C^+$ cells isolated from ESC infected with sgGFP or sgDnmt1. Shown are mean ± SD, n=3 biologically independent cell cultures. **c,** Relative percentage of spontaneous 2C-like cells and representative FACS plot of ESC culture. Shown are mean ± SD, n=3 biologically independent cell cultures. **b-c,** *P* values (indicated as numbers in the graphs) are calculated by unpaired t-test, two-tailed, two-sample unequal variance. **d,** Mean promoter methylation change from ESC to 2C-like cell of $2C^+$-upregulated (n=1,611 genes) and $2C^+$-downregulated genes (n=1,398 genes). The black central line is the median, boxes limits indicate the upper and lower quartiles, whiskers indicate the 1.5 interquartile range, dots represent outliers. *P* values (shown as p) are calculated by two-tailed Mann-Whitney *U*-test and effect-sizes (shown as r) are calculated as $Z/\sqrt{N}$ where *Z* is the *Z* value of the Mann–Whitney *U*-test and N is the number of samples. **e,** The percentage of 2C-like cells with prolonged *Dux* induction. Prolonged *Dux* induction did not induce complete 2C-like transition as *Dux* induction initiates an unsynchronized 2C-like transition (Figure 1b) and cannot maintain 2C-like state (Supplementary Figure 1g). **f**, Bootstrapped correlation analysis showing that the $Zscan4^+$ intermediate cells reported by Rodriguez-Terrones et al. are assigned to clusters 1 and 3 of our single-cell analysis but cannot be firmly assigned to cluster 2, the intermediate state. **g,** Density plot showing the distribution of cells from the different clusters generated in our study and a study from Rodriguez-Terrones et al. along the pseudotime. Cells were ordered using the 93 genes selected by Rodriguez-Terrones et al. after technical batch removal using the MNN method. We note that the date from Rodriguez-Terrones et al did to capture the intermediate state (Cluster 2) identified in our study. **h-i,** FACS histogram plot showing the expression of Myc **(h)** or Dnmt1 **(i)** in mESCs. **a-c, e, h-i,** experiments were repeated independently twice with similar results. Statistical source data can be found in Supplementary Table 10.

Original blots
in Supplementary Fig. 5a



Original blots
in Supplementary Fig. 6a



Supplementary Figure 7
Unprocessed original scans of blots.
Unprocessed images of immunoblots shown in Supplementary Figure 5a and Supplementary Figure 6a are provided.

Titles and legends for Supplementary Tables:

Supplementary Table 1
Gene expression and differential gene expression summary of D0 $2C^-$, D1 $2C^-$ and D1 $2C^+$ samples.
Table S1-1: TMM normalized gene expression of D0 $2C^-$, D1 $2C^-$ and D1 $2C^+$ samples.
Table S1-2: List of the list of Up and Down regulated genes after Dux-induction (D1 $2C^+$ Vs D0 $2C^-$).

Supplementary Table 2
The category of the dynamically changing genes/repeats between D0 $2C^-$, D1 $2C^-$, and D1 $2C^+$ shown in Fig.1f.
This Table summarize the direction of the change of the genes whose expression is altered during ESC to 2C-like transition shown in Fig. 1f.

Supplementary Table 3
log_2(mean expression+1) of genes and repeats in the identified scRNA-seq clusters.
Table showing the log_2-scaled mean expression of genes/repeats in the scRNA-seq clusters.

Supplementary Table 4
CRISPR screen results obtained using MAGeCK.
Table S4-1: Results of batch1.
Table S4-2: Results of batch2.

Supplementary Table 5
Summary of the bulk RNA-seq results after Myc knock-down.
Table S5-1: TMM normalized gene expression of Myc knock-down experiment.
Table S5-2: List of D1 $2C^+$ Down-regulated genes directly regulated by Myc.
Table S5-3: List of D1 $2C^+$ up-regulated genes directly regulated by Myc.

Supplementary Table 6
RRBS CpG methylation ratios around the promoters (TSS±1kb) of D1 $2C^+$ up-/down-regulated genes.
Table S6-1: RRBS CpG methylation at 1kb around the promoters of D1 $2C^+$ up-regulated genes
Table S6-2: RRBS CpG methylation at 1kb around the promoters of D1 $2C^+$ down-regulated genes

Supplementary Table 7
Summary of the bulk RNA-seq results after Dnmt1 knock-down.
Table S7-1: TMM normalized expression of D1 $2C^+$ up-regulated genes after Dnmt1 knock-down (Dnmt1 silenced genes/repeats are highlighted in red).
Table S7-2: TMM normalized expression of D1 $2C^+$ down-regulated genes after Dnmt1 knock-down.

Supplementary Table 8
Summary of the sequences used in this study.
Table S8-1: sgRNA used in this study.
Table S8-2: synDux sequence.

Supplementary Table 9
Sequencing data statistics.
Table S9-1: Bulk RNA-seq sequencing statistics.
Table S9-2: RRBS samples sequencing statistics.
Table S9-3: Drop-seq samples sequencing statistics.

Supplementary Table 10
Numerical source data for all figures.
Each panel in the excel sheet contains the raw numerical data used to generate the corresponding panel. The tables names correspond to the panels they represent.