The Innovation, Volume *4*

Supplemental Information

Dynamic network biomarker factors orchestrate cell-fate determination at tipping points during hESC differentiation

Lin Li, Yilin Xu, Lili Yan, Xiao Li, Fei Li, Zhuang Liu, Chuanchao Zhang, Yuan Lou, Dong Gao, Xin Cheng, and Luonan Chen

Supplemental Information including:

Materials and Methods

Figures S1 to S6

Note 1

Supplemental Materials and Methods

Analysis of Chu-time dataset based on M-DNB model

To investigate the critical points and factors cell in the process of hESCs' differentiation at single-cell resolution, here we considered a public scRNA-seq dataset, Chu-time(Chu et al., 2016). The dataset consist of 758 cells including 6 times (0h, 12h, 24h, 36h, 72h, 96h) with three cells differentiation states: pluripotent embryonic stem cells, mesendoderm, definitive endoderm. Chu-time dataset contains sufficient time points for analyzing the two transitions, i.e. ES-to-ME and ME-to-DE differentiation processes. We first performed dimensional reduction using Seurat package(Stuart et al., 2019). The dataset was log-transformed with the scale factor for cell-level normalization setting of 1,000,000. And we performed Principal Component Analysis (PCA) using the top 5,000 variable genes. The top 20 PCs were further applied to perform t-SNE. Furthermore, we applied SCPattern which was designed for time course data to identify differential expression genes of two consecutive time points in Chu-time dataset.

To identify the critical points of the differentiation from hESCs to DE, we applied M-DNB model on Chu-time dataset. We considered a total of 19189 genes in this dataset. We first computed the Pearson correlation of genes at each time points. Considering that some of the genes have no expression throughout the differentiation progress and some other genes show no significant correlation (absolute value of Pearson correlation < 0.02) with any other genes during this differentiation, therefore, we obtain 13,971 gene modules. We first calculated composite index (CI) for each gene module at six time points based on M-DNB model. Then, we obtained quantitative indicators (QIs) by averaging top 50 CIs for each time. Based on CI and QI, we

identified two critical states/tipping points (12h and 36h) of the three stages with their M-DNB genes.

After determining the two tipping points, we further identified their important regulators/factors based on the M-DNB genes. We selected gene modules with top 50 CIs at 12h and 36h.

M-DNB factors and M-DNB genes in differentiation process

Since we have obtained M-DNB genes in each tipping point. We further investigated transcription factors (TFs) of M-DNB genes. Here, we applied Ingenuity Pathway Analysis (IPA) to identify the upstream regulators of M-DNB genes, where the molecule type was set as transcription regulator. We found that four transcriptional factors could regulate 86% of the first group of M-DNB genes, that is, genes that regulate the differentiation to ME direction. The four genes are *MYCN*, *FOS, HSF1* and *TP53*. In addition, three transcriptional regulators regulate 88% of the second group of M-DNB genes, controlling cells differentiation to DE direction. The three transcription factors are *MYCN*, *HSF1* and *MYC*.

Human ES cell maintenance and DE differentiation method: human ES cell culture and differentiation process of hESCs towards definitive endoderm

H9 ES cells were maintained on irradiated mouse embryonic fibroblast feeder cells in hESC medium consisting of DMEM/F12 (50:50; Gibco) supplemented with 20% knock-out serum replacement (KOSR) and 5ng/ml basic fibroblast growth factor (bFGF). Prior to the induction of endoderm in the monolayer cultures, hPSCs were passaged onto a Matrigel (1:3 diluted in IMDM) coated surface (typically 6-well dish) for 1 or 2 days. To initiate differentiation (**day 0**), the cells were cultured for 1 day in RPMI-based medium supplemented with glutamine (2 mM), MTG (4.5 × 10−4 M; Sigma), activin A (100 ng/ml), CHIR99021 (2 μM). At day 1, CHIR99021 was removed and cells were cultured for the next 2 d in RPMI supplemented with glutamine (2 mM), ascorbic acid (50 μg/ml: Sigma), MTG (4.5 × 10−4 M: Sigma), basic fibroblast growth

factor (bFGF; 5 ng/ml), activin A (100 ng/ml) and for the after 3 days in serum-free-differentiation (SFD)-based medium with the same supplements. SFD consists of homemade IMDM: Ham's F12(3:1) with N2/B27supplements and 0.05% BSA. The medium was changed every day.

Perturbation of M-DNB factors during differentiation of hESCs towards definitive endoderm

To test M-DNB factors in ES-to-ME and ME-to-DE differentiation, perturbation of each M-DNB factor during the differentiation using lentivirus overexpression/knockdown system (Figure. S2).

At tipping point 1, the expression levels of FOS was upregulated, while the expression levels of HSF1 and MYCN was downregulated; at tipping point 2, the expression level of MYC was downregulated. To reserve/perturb the expression level changes of M-DNB factors at tipping points, FOS knockdown, HSF1 and MYCN overexpression should be induced at tipping point 1 (before and near 12h), and MYC overexpression should be induced at tipping point 2 (before and near 36h) without disturbing ES-to-ME differentiation process (0-24h).

To induce the perturbation of M-DNB factors at appropriate timing (at tipping point 1 or 2), four cell lines were established to overexpress or knockdown M-DNB factors using different lentivirus vectors. HSF1 CDS sequence and MYCN CDS sequence were separated cloned to an overexpression lentivirus donor vector pLenti-GIII-CMV, FOS siRNA was cloned to a knockdown lentivirus donor vector pLenti-siRNA-GFP, and *MYC* CDS sequence was and cloned to an all-in-one inducible lentivirus donor vector: pCW-TRE-T2A-dsRed. For lentivirus generation, donor vector was transfected with lentivirus package plasmid pspAX and pMD2.G to 293FT cells cultured in DMEM basic (Gibco) + 10% serum using calcium phosphate. Lentivirus in the supernatant were collected 2 days after transfection. 1ml lentivirus supernatant was added to ES in the 10-cm dish at ~20% confluence for transduction. 4 days after the transduction, 1ug/ml puromycin was added to the medium to select lentivirus transduced ES cells and kill uninfected cells (4 days) to ensure cells were transduced. These cell lines were named ES-HSF1-OE, ES-MYCN-OE, ES-FOS-KD and ME-

MYC-OE.

After establishment of M-DNB perturbation cell lines, ES-to-DE differentiation was carried out using these cell lines, cells were collected after 72h differentiation and compared with wildtype ES, ME (24h) and DE (72h) samples. These cells were further applied for Bulk-seq library preparation and ATAC-seq library preparation.

Specifically, after 72h differentiation, ES-HSF1-OE, ES-MYCN-OE, ES-FOS-KD were used to compared with ES and ME, due to HSF1, MYCN or FOS was already overexpressed(or knockdown) at ES stage (0h), these M-DNB expression levels were perturbed at tipping 1 (12h).

Different from tipping 1, the perturbation of MYC at tipping 2 (36h) in ME-MYC-OE cells should not be start from 0h, but start from 24h to avoid perturbation disturbed ES-to-ME differentiation. Therefore, the inducible lentivirus donor vector was used to overexpress MYC, the overexpression of MYC will only be induced at the presence of doxycycline (2ug/ml). The differentiation of ME-MYC-OE cells with adding doxycycline (24-72h) result in MYC perturbation restricted in ME-to-DE differentiation stage.

Quantitative real-time PCR

Total RNA was prepared using RNAprep Pure Micro Kit (TIANGEN). 500~1000ng RNA was reverse-transcribed into cDNA using random hexamers and Oligo dT with GoScript™ Reserve Transcription System (Promega). qPCR was performed on a QuantStudio6 Flex (ABI) using FastStart Universal SYBR® Green Master (ROX) (ROCHE). Expression levels were normalized to the housekeeping gene TATA box binding protein (TBP). hES genome was prepared using TIANamp Genomic DNA Kit (TIANGEN) as standard samples (100ng, 10ng, 1ng, 0.1ng) to draw standard curves for absolute quantification of qPCR analysis. Primer oligonucleotide sequences are available in Table S1.

Bulk-seq library preparation

Bulk RNA sequencing of sorted cells (ES cells, ME cells, DE cells, ES-FOS-KD cells
ES-HSF1-OE cells, ES-MYCN-OE cells, ME-MYC-OE cells) was performed using RNeasy Mini Kit (QIAGEN, no. 74106) and mRNA-seq V3 Library Prep Kit (Vazyme, no. NR611).

ATAC-seq library preparation

We performed ATAC sequencing of cells ((ES cells, ME cells, DE cells, ES-FOS-KD cells
ES-HSF1-OE cells, ES-MYCN-OE cells). The optimized ATAC-seq protocol (Corces et al., 2017) was performed as follows: 50,000 cells were collected and washed once with PBS. Cells were then lysed for 3 minutes with 50 μL of ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM MgCl2; 0.1% NP-40; 0.1% Tween 20; and 0.01% digitonin). The lysed nuclei were washed immediately with 1 mL of wash buffer (10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM MgCl2; and 0.1% Tween 20) followed by centrifugation at 500 g for 10 minutes at 4 °C. The following steps were used to prepare sequencing libraries using a TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme, TD501).

Preprocessing and analysis of bulk RNA sequence data

Paired-end reads were mapped to the hg19 transcriptome using hisat2(Kim et al., 2015). Read counts and gene length were calculated by feature Counts (1.6.0)(Liao et al., 2014). We used the function fpkm from R package Deseq2(Love et al., 2014) to get the FPKM data matrix.

We computed the Pearson correlation to compare the similarity among samples using FPKM data matrix. To identify signatures of ES cells, ME cells, and DE cells, we performed differential expression tests by making pairwise comparisons (cut-off: log(fold change) ≥ 1). We also obtained the signatures of ES-FOS-KD cells, ES-HSF1-

OE cells, ES-MYCN-OE cells and ME-MYCN-OE cells by making comparisons (ES-FOS-KD cells vs ES cells, ES-HSF1-OE cells vs ES cells, ES-MYCN-OE cells vs ES cells, ME-MYCN-OE cells vs ME cells). The cut-off was set as log(fold change) ≥ 1.

To more accurately measure the differentiation potency of these samples, we used the Markov-Chain entropy (MCE) method(Shi et al., 2018). The normalized data and protein–protein interaction networks were used to compute MCE values of each samples. We normalized data in the same way as in MCE method. we added an offset value of 0.1 before log-scale transformation (log2(1.1)≈0.13) and obtained the protein-protein interaction networks from the dataset(Teschendorff and Enver, 2017).

Preprocessing and analyzing ATAC-seq data

We first removed adaptors from raw fastq files using TrimGalore-0.5.0 and mapped the trimmed fastq files to hg19 genome using Bowtie2(Langmead and Salzberg, 2012). Sambamba(Tarasov et al., 2015) was further conducted to remove duplicates. Then we obtained the normalized CPM.bw files using function bamCoverage from DeepTools(Ramirez et al., 2016). The normalized CPM.bw files were further applied for heatmap visualization with the function of computeMatrix and plotHeatmap. MACS2-2.1.1 was utilized for peak calling. To obtain the count matrix of all samples, we first generated the consensus peak list using an R package, diffbind, with the parameter minOverlap = 0.8. Based on consensus peak list, we used bedtools multicov to count the number of overlaps in each BAM file. To identify specific open peaks in ES cells, ME cells and DE cells, we applied edgeR using count matrix of these samples.

We also identified specific open peaks of ES-FOS-KD cells, ES-HSF1-OE cells, ES-MYCN-OE cells making pairwise comparisons (ES-FOS-KD cells vs ES cells, ES-HSF1-OE cells vs ES cells, ES-MYCN-OE cells vs ES cells). The specific open peaks were obtained by a maximum FDR of 0.05 and a minimum log2(fold change) of 0.5. Since the samples of each cell types are highly consistent, we show one of

them in our main figure (Figure.4). We further annotated these peaks by homer (http://homer.ucsd.edu/homer/motif/rnaMotifs.html).

Gene ontology and gene set enrichment analyses

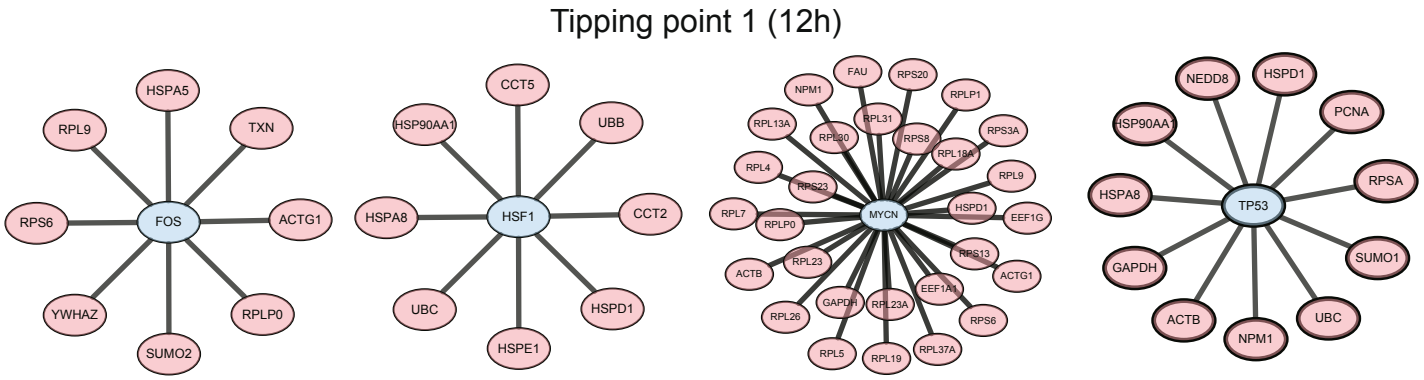
The R package clusterProfile was utilized to perform Gene Ontology functional analysis. The signature genes of these samples from RNA-seq and the annotated peaks of samples from ATAC-seq were collected for subsequent enrichment analyses of functions and pathways by annotating Gene Ontology biological processes.

To identify the relationship of cells, we performed gene set enrichment analysis (GSEA). The ES cells' signature genes, ME cells' signature genes, and DE cells' signature genes of bulk RNA-seq were set as the gene set database for further analysis. We selected the top 100 ES markers, top 100 ME markers and top 100 DE markers (ordered by log(Fold Change)) as the gene set database. GSEA was performed using GSEA v4.1.0 software with 1000 gene-set permutations.

Statistical Analysis.

Statistical ananlysis was carried out using R v4.1. For bulk RNA-seq analysis, we used the log(Fold change) to identify the signature genes of each group and the threshold was set as log(Fold change)>1. For ATAC-seq analysis, we applied edgeR to defined the specific open peaks based on an R package, edgeR. Statistical significance for the analyses conducted was set at a maximum FDR of 0.05 and a minimum log2(fold change) of 0.5.

A



B

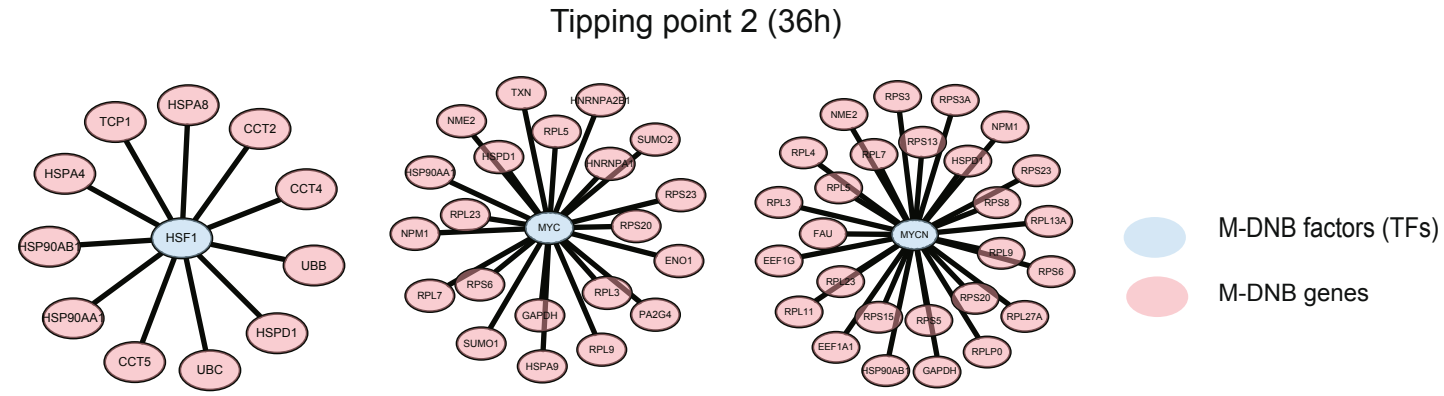


Figure. S1 Networks showed the relation of regulation from M-DNB factors to M-DNB genes in tipping point 1 (A); and tipping point 2 (B).

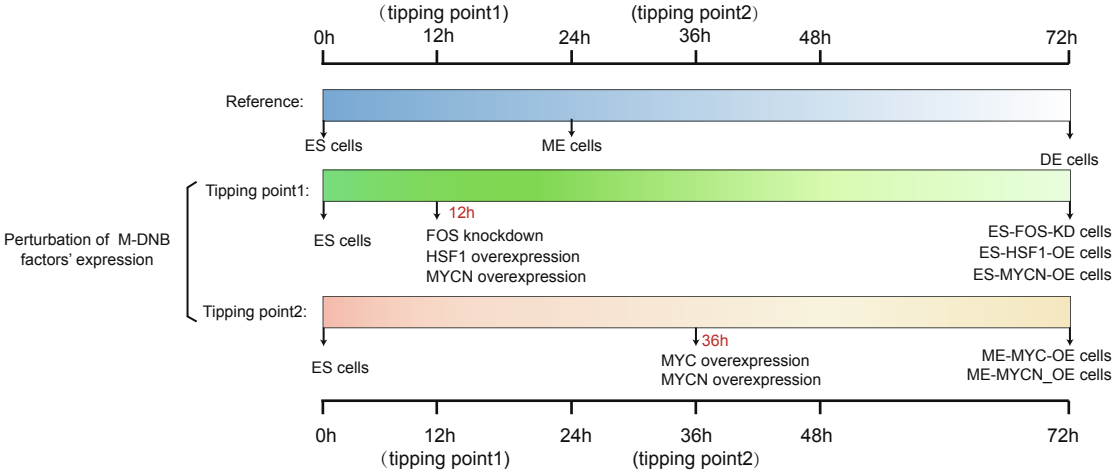


Figure. S2 Experimental flow of M-DNB factors' perturbation and hESCs' differentiation. Reference is the process of hESCs' differentiation, and we obtained ME cells at 24h, DE cells at 72h of hESCs' differentiation. Perturbation of M-DNB factors' expression consists of two parts. In tipping point 1, we knockdown FOS and overexpress HSF1, MYCN,respectively at 12h (nearly) and obtained ES-FOS-KO cells, ES-HSF1-OE cells, and ES-MYCN-OE cells at 72h of differentiation. In tipping point 2, we overexpress MYC,respectively at 36h (nearly) and obtained ME-MYC-OE cells at 72h of differentiation.

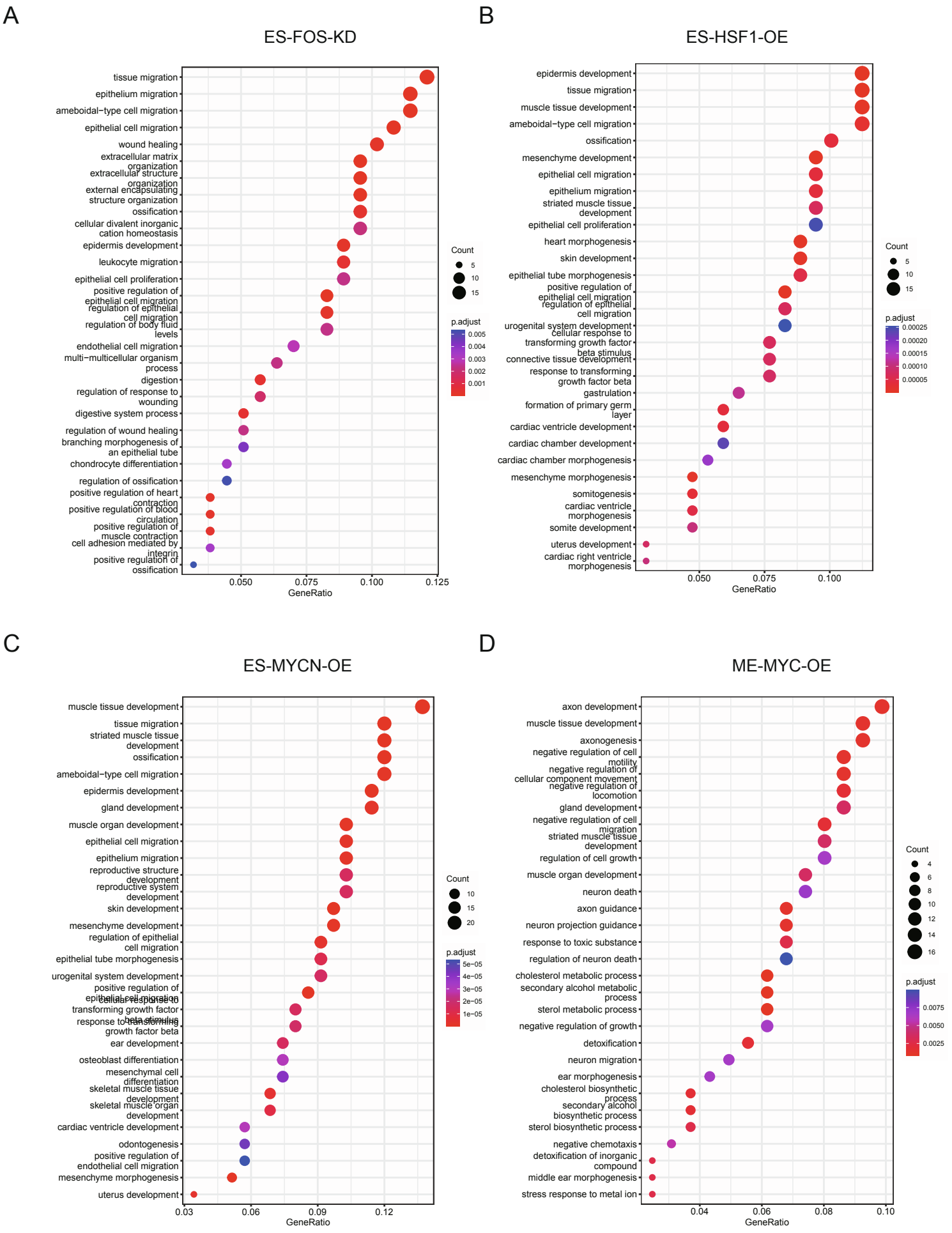
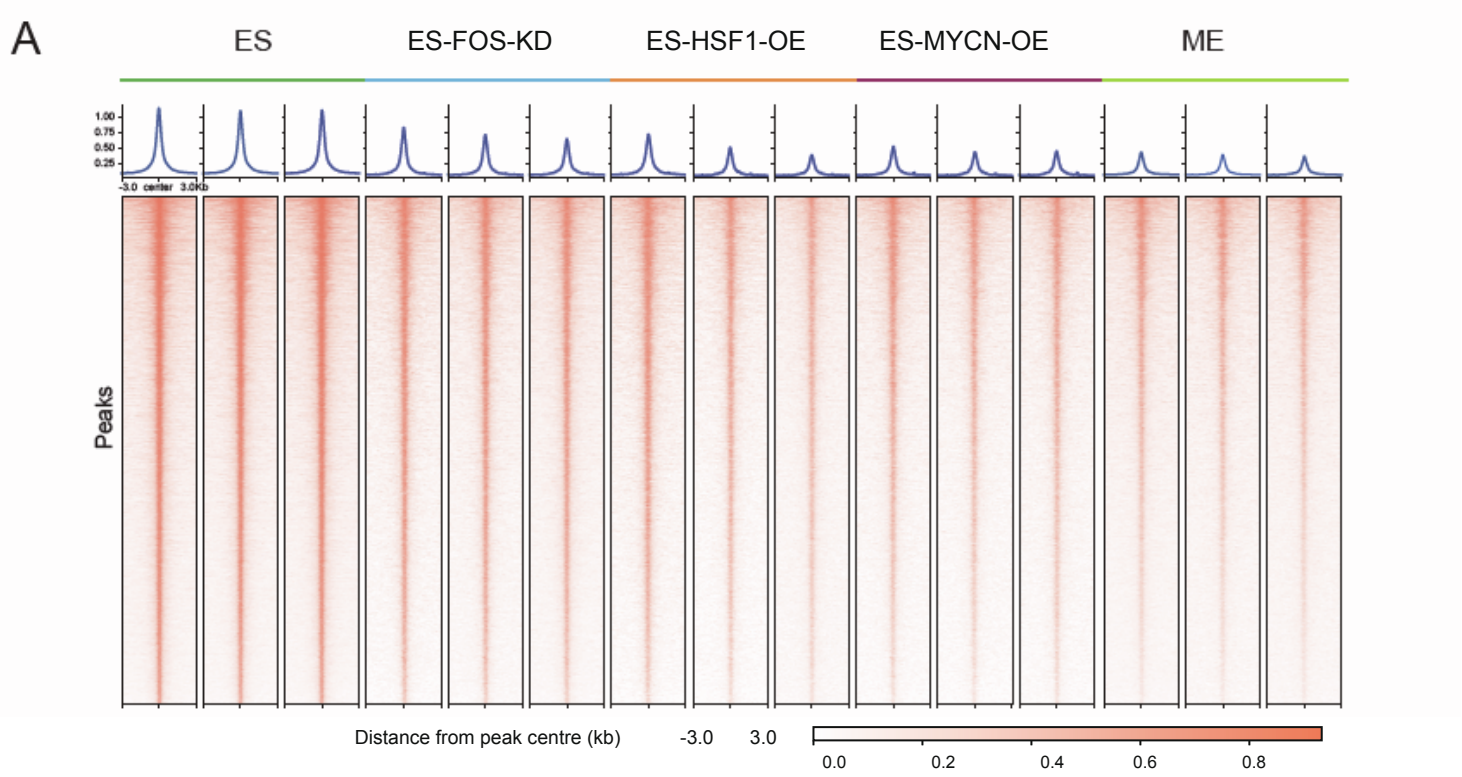


Figure. S3 Dot plots show the significantly enriched gene ontology terms of ES-FOS-KD (A), ES-HSF1-OE (B), ES-MYCN-OE (C), and ME-MYC-OE (D) for RNA-seq data.



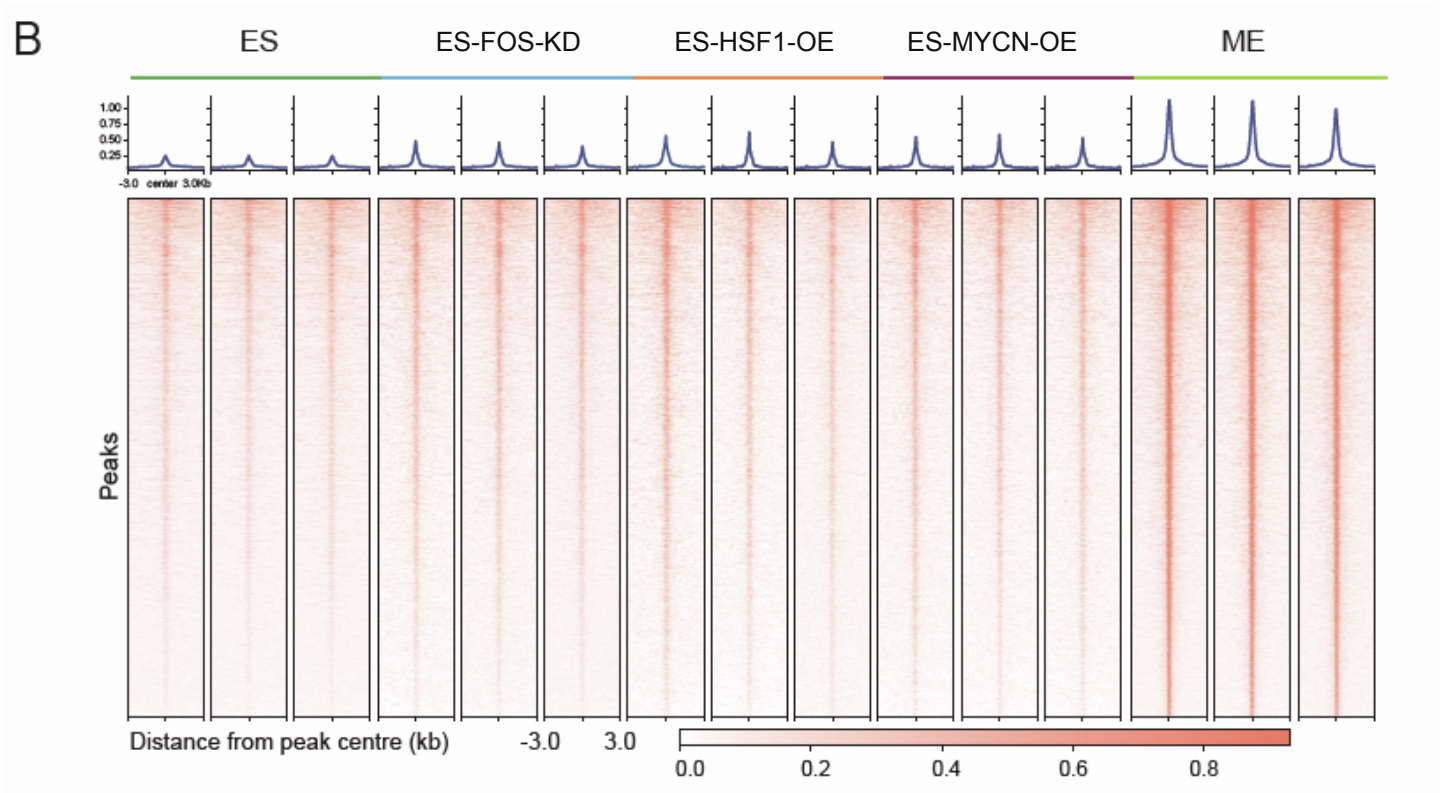


Figure.S4 (A) Profiles (top) and Heatmap of ES, ES-FOS-KD, ES-HSF1-OE, ES-MYCN-OE and ME cells around the peak center of ES binding sites.

(B) Profiles (top) and Heatmap of ES, ES-FOS-KD, ES-HSF1-OE, ES-MYCN-OE and ME cells around the peak center of ME binding sites.

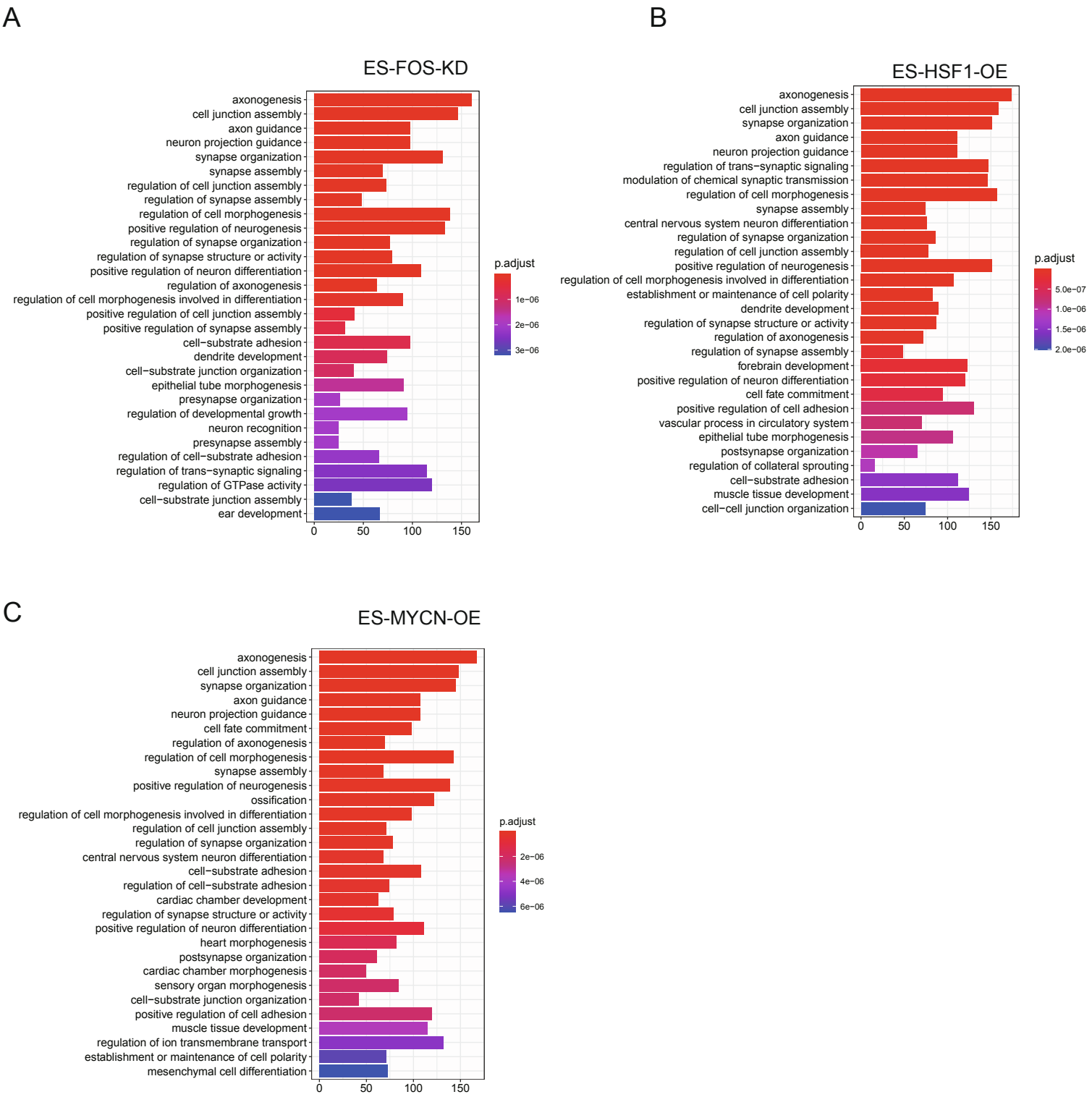


Figure.S5 Bar plots show the significantly enriched gene ontology terms of ES-FOS-KD (A), ES-HSF1-OE (B), ES-MYCN-OE (C) for ATAC-seq data.

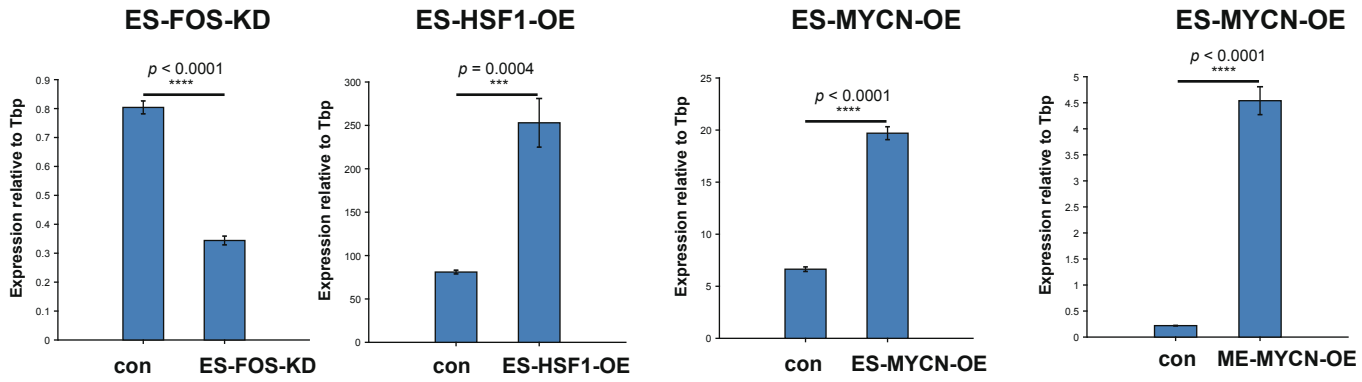


Figure. S6 M-DNB factors expression level changes in hES after lentivirus transfection. At tipping point 1, overexpression lentivirus vector pLenti containing FOS-shRNA, HSF1, MYCN sequence was used to generate lentivirus, and transfected into H9 ES, the RNA expression level changes were tested after lentivirus transection. At tipping point 2, inducible expression lentivirus vector pCW-TRE-T2A-dsRed containing MYCN sequence was used to generate lentivirus, and transfected into H9 ES, overexpression of MYCN were induced by adding doxcycline and tested via qPCR.