Dynamic network biomarker factors orchestrate cell-fate determination at tipping points during hESC differentiation
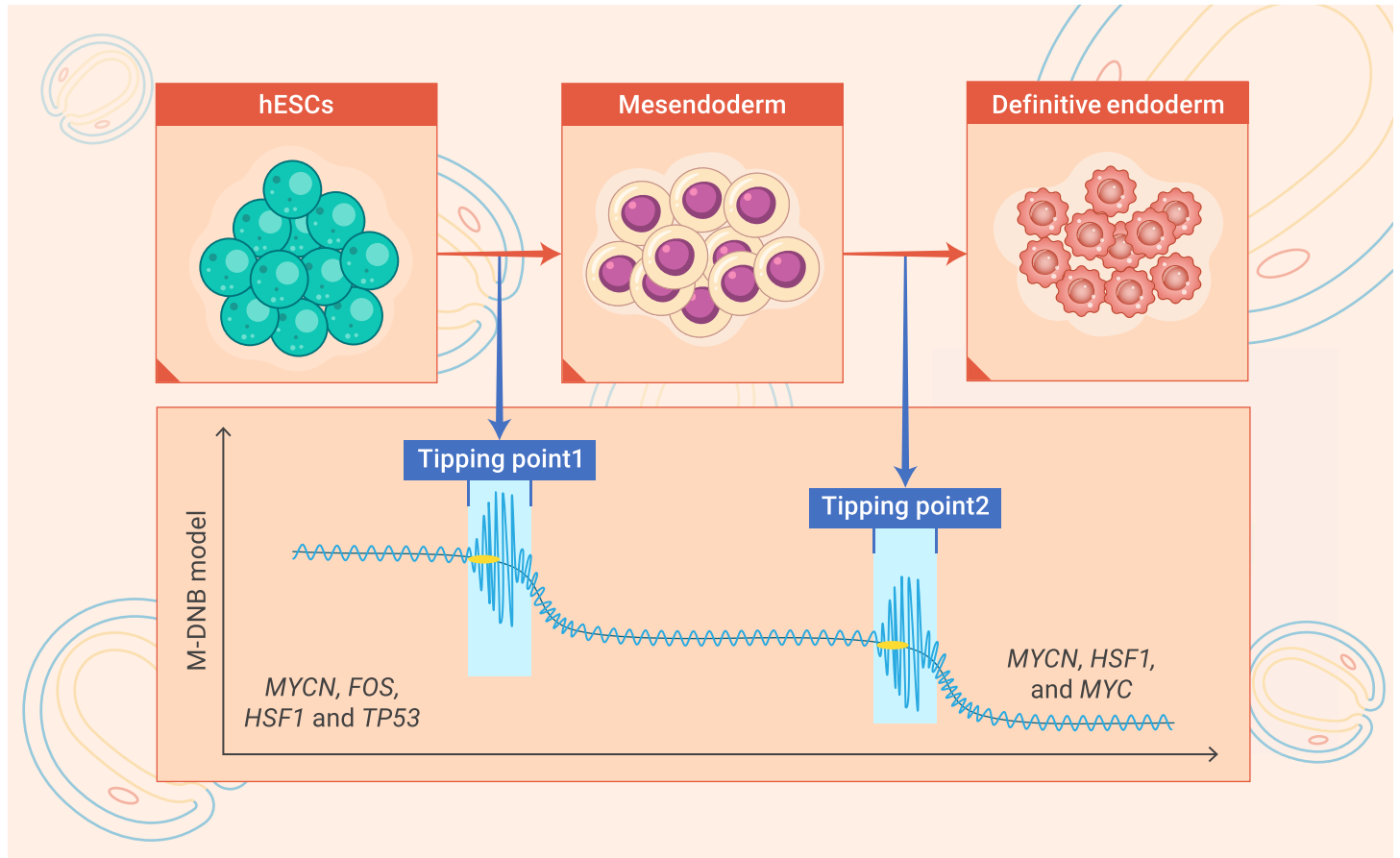
Lin Li,[1,5] Yilin Xu,[1,5] Lili Yan,[1,5] Xiao Li,[1] Fei Li,[1] Zhuang Liu,[1] Chuanchao Zhang,[2] Yuan Lou,[3,*] Dong Gao,[1,*] Xin Cheng,[1,*] and Luonan Chen[1,2,4,*]

*Correspondence: yuanlou@sjtu.edu.cn (Y.L.); dong.gao@sibcb.ac.cn (D.G.); xcheng@sibcb.ac.cn (X.C.); lnchen@sibcb.ac.cn (L.C.)

GRAPHICAL ABSTRACT

PUBLIC SUMMARY

- M-DNB model identifies two tipping points of hESC differentiation.

- Five M-DNB factors are master regulators in hESC differentiation.

- Before tipping points, M-DNB factors orchestrate cell fate determination.

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

Lin Li,[1,5] Yilin Xu,[1,5] Lili Yan,[1,5] Xiao Li,[1] Fei Li,[1] Zhuang Liu,[1] Chuanchao Zhang,[2] Yuan Lou,[3,*] Dong Gao,[1,*] Xin Cheng,[1,*] and Luonan Chen[1,2,4,*]

[1]Key Laboratory of Systems Biology, Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

[2]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou 310024, China

[3]School of Mathematical Sciences, CMA-Shanghai and MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China

[4]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China

[5]These authors contributed equally

*Correspondence: yuanlou@sjtu.edu.cn (Y.L.); dong.gao@sibcb.ac.cn (D.G.); xcheng@sibcb.ac.cn (X.C.); lnchen@sibcb.ac.cn (L.C.)

Citation: Li L., Xu Y., Yan L., et al., (2023). Dynamic network biomarker factors orchestrate cell-fate determination at tipping points during hESC differentiation. The Innovation **4(1)**, 100364.

The generation of ectoderm, mesoderm, and endoderm layers is the most critical biological process during the gastrulation of embryo development. Such a differentiation process in human embryonic stem cells (hESCs) is an inherently nonlinear multi-stage dynamical process which contain multiple tipping points playing crucial roles in the cell-fate decision. However, the tipping points of the process are largely unknown, letting alone the understanding of the molecular regulation on these critical events. Here by designing a module-based dynamic network biomarker (M-DNB) model, we quantitatively pinpointed two tipping points of the differentiation of hESCs toward definitive endoderm, which leads to the identification of M-DNB factors (*FOS*, *HSF1*, *MYCN*, *TP53*, and *MYC*) of this process. We demonstrate that before the tipping points, M-DNB factors are able to maintain the cell states and orchestrate cell-fate determination during hESC (ES)-to-ME and ME-to-DE differentiation processes, which not only leads to better understanding of endodermal specification of hESCs but also reveals the power of the M-DNB model to identify critical transition points with their key factors in diverse biological processes, including cell differentiation and transdifferentiation dynamics.

INTRODUCTION

The generation of three primordial germ layers—the ectoderm, mesoderm, and endoderm—is the most critical biological process during the gastrulation of embryo development.[1,2] It has been clearly demonstrated that human embryonic stem cells (hESCs) can be used as an effective *in vitro* model to recapitulate *in vivo* developmental programs and dissect the molecular mechanism underpinning lineage specification and differentiation.[3] Generally, the differentiation of hESCs can be viewed as a nonlinear dynamical process with multiple transition stages of cell states, during which the regulations on its tipping points immediately before the transition stages are crucial to the cell-fate determination from the perspective of the dynamical system.[4,5] Thus, the identification of the tipping points, as well as the relevant key regulators, is important to improve the understanding of lineage specification and differentiation.[6]

It has been shown that hESCs differentiate into the definitive endoderm (DE) through an intermediate stage known as the mesendoderm (ME),[2] and that the cell-fate determination during this process is tightly controlled by a series of transcription factors, including those of pluripotency factors (*Nanog*, *Oct4*, *Sox2*), the P53 family, the Myc family (MycN, Myc), T box genes (Brachyury/T, Eomesdermin), Mixl1, *Sox17*, and FoxA2.[2,7–9] In addition to classical pluripotency factors, *MycN* has been shown to play a vital role in maintaining cells in a proliferative and undifferentiated state,[10,11] while *Myc* has been shown to maintain pluripotency by inhibiting the primitive endodermal master regulator *GATA6*.[12] Genetic studies have revealed that the loss of the P53 family blocks the mesendodermal differentiation of hESCs,[9] and that Eomes is the master regulator essential for DE specification from the mesendoderm.[13] Being the markers for nascent DE, *Sox17*, or *FoxA2* is required for further differentiation into derivative lineages, not for the DE specification.[9,14] In addition, functional studies have revealed the critical roles of extracellular cues that activate canonical Wnt, Activin/Nodal, and BMP signaling during ME and DE differentiation from hESCs, which synergize in a spatiotemporal fashion to stabilize the transcriptional network.[15] Despite the extensive studies designed to decipher the molecular mechanisms that govern

the cell-fate decision during gastrulation, this early developmental process has never been investigated from the perspective of dynamic systems to understand the regulation of the tipping points with their critical transition states, let alone the key modulators of these irreversible changes.
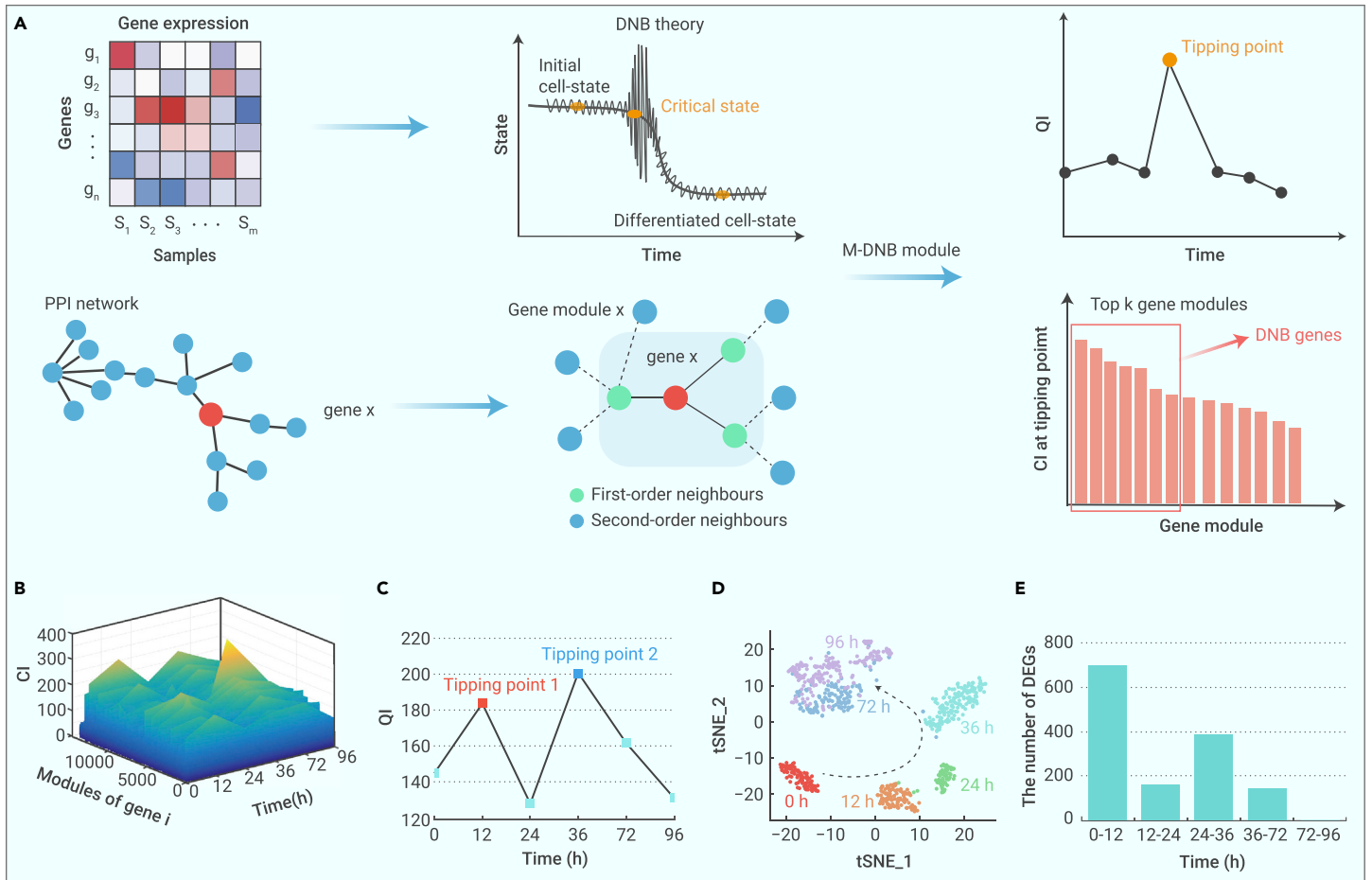
The differentiation of hESCs can generally be viewed as the evolution of a nonlinear dynamical system, during which the molecular events at the tipping points immediately before the downstream transition determinate the cell fate. To further elucidate the dynamic molecular controls that govern the early differentiation of hESCs, we must identify the tipping points and their key modulators, which is challenging as it requires information beyond the traditional differential expression analysis because of high sensitivity to noise and nonlinear dynamics near the tipping points.[4,5] On the basis of nonlinear dynamical systems theory, the dynamic network biomarker (DNB) method, which uses a group of collectively fluctuated genes during a biological process rather than differentially expressed genes, has been developed to identify the critical states and to predict the leading molecules of the biological processes of disease progression by exploiting the dynamical features of the tipping points obtained from bulk omics data.[4,16,17] Among the omics data, those from single-cell transcriptomic analyses (single-cell RNA sequencing [scRNA-seq])[18–20] provide a wealth of unprecedented information for detecting the tipping points and the related leading transcripts. However, the application of sophisticated, dynamics-based methods on single-cell data is limited because of the higher levels of transcript amplification noises and drop-outs than those of RNA sequencing (RNA-seq) performed on bulk cell populations. To overcome this obstacle, we developed a module-based DNB (M-DNB) model by exploring transcriptional network and dynamics information from scRNA-seq data,[21,22] which can provide reliable quantification on the tipping points and their regulators.

Specifically, here, by applying M-DNB to analyze the time course scRNA-seq datasets (Chu-time dataset) from endodermal differentiation of hESCs,[23] we first identified two tipping points (12 and 36 h of differentiation) that divide the whole process into three stages, embryonic stem (ES), ME, and DE and five M-DNB factors (*FOS*, *HSF1*, *MYC*, *MYCN*, and *TP53*) that potentially modulate the two cell-state transitions (ES to ME and ME to DE).

RESULTS

Overview of M-DNB model

As shown in Figure 1A, generally, there is no significant difference between an initial cell state and its critical state, different from the differentiated cell state. Thus, traditional differential gene expression analysis may fail to distinguish the critical state or tipping point. To solve this problem, researchers developed DNB theory, which is a dynamics-based method using bulk omics data to study the tipping points and the related crucial genes (M-DNB genes) in a biological process that involves state/phase transitions. Distinct from the traditional methods that mainly focus on the first-order statistical information (e.g., mean values of individual genes) or "differential gene expression," the DNB method explores the higher order statistical information or "differential gene association" (e.g., correlations or covariances of a gene group) and can identify the tipping points. However, the application of the DNB method in scRNA-seq data analysis is limited because of the severe interference from transcript amplification noises and dropout events. To solve this problem, we

Figure 1. Two tipping points were identified during the differentiation of hESCs based on M-DNB model analysis (A) Overview of M-DNB model. There are three phases during differentiation of hESCs: (1) gradual cell-state changes for preparing cell-fate commitment, (2) the tipping points (just before the irreversible cell-state transitions) for making cell-fate determination, and (3) drastic cell-state transitions for completing the cell-fate determination. Meanwhile, the PPI network from STRING was used to construct the gene module for each gene. Here, only first-order and second-order neighbors are considered in each gene module. Furthermore, to identify the tipping points and critical factors of hESCs' differentiation, we combined traditional DNB theory and gene modules from PPI networks. We first calculate CI for each gene module and choose top *k* gene modules (e.g., *k* = 50) in terms of their CI scores. Then, the average CI of the top *k* genes (i.e., quantitative indicators [QIs]) is obtained to determine whether this time point is the tipping point (i.e., if the QI score of this time point is a peak higher than the two neighboring/adjacent time points, it is a tipping point, and the top *k* genes are also the M-DNB genes of this tipping point). In this way, we can identify all tipping points with their major regulators during the hESCs' differentiation. (B) CI scores of gene modules during the process of the differentiation of hESCs. The CI scores at 12 and 36 h are higher than those at other time points, which indicates that 12 and 36 h may be the critical states or tipping points during the differentiation of hESCs. (C) QI scores at 12 and 36 h of hESCs' differentiation were higher than those at other time points, which indicated that 12 and 36 h were the tipping points during the differentiation of hESCs. (D) t-SNE plot showed the distribution of cells during the differentiation of hESCs. (E) The number of differential expression genes between each two adjacent times which were identified by SCPattern.

designed a module-based DNB model that transforms gene expression information into gene modules/networks on the basis of a protein-protein interaction (PPI) network (Figure 1A). First, instead of identifying gene groups on the basis of clustering, our model transforms a certain gene (namely gene *x*) into a gene module/network constructed on the basis of the associations or covariances between this specific gene and other genes, where gene *x* is the core of the module (Figure 1A). The template network is based on the PPI network available in STRING (https://string-db.org/). A pair of genes are considered to have direct interaction if the two genes are connected at the template network and have the co-expressed association judging from the scRNA-seq data. In this study, we focused on the first-order neighbors that connect directly to the core gene *x*, as well as second-order neighbors that have direct interaction with any of the first-order neighbors.

The original DNB criterion was established on the basis of the principle that "collective fluctuation of a group/cluster of genes/molecules implies the imminent transition" from the bifurcation theory.[4] However, distinct conclusions of the DNB analysis can be achieved because of differential clustering methods. In this study, we developed the M-DNB model, in which each given gene has one local network without a clustering procedure, providing a reliable quantification for tipping point detection. Specifically, the M-DNB criterion, also known as "composite indicator" (CI), establishes that each gene (e.g., gene *i*) is defined by Equation 1, with the three components SD_{in}, PCC_{in}, and PCC_{out}, estimated using scRNA-seq data:

$$CI = \sqrt{n} * SD_{in} * \frac{PCC_{in}}{PCC_{out}}, \qquad \text{(Equation 1)}$$

where the first component SD_{in} is the average variance of the internal genes in a gene module *i*, which describes the volatility of the gene expressions and can be calculated by the following equation:

$$SD_{in} = \frac{1}{n} \sum_{j=1}^{n} \sqrt{E\left[|x_{ij} - E(x_i)|^2\right]}, \qquad \text{(Equation 2)}$$

where x_{ij} represents the expression of gene *i* for cell *j*, *n* is the number of genes in the module, and *E* is the operation of average value.

The second component (PCC_{in}) is the average of the internal correlation of a gene module *i*, which is defined as follows:

$$PCC_{in} = \frac{1}{m} \sum_{j=1}^{m} cov(x_i, x_j) \Big/ \sqrt{var(x_i)var(x_j)}, \qquad \text{(Equation 3)}$$

where *m* is the number of edges in the gene module, x_i, x_j are the genes *i* and *j* that interact in the module, respectively; and *cov* and *var* represent covariance and variance operations across all cells, respectively. Equation 3 reflects the
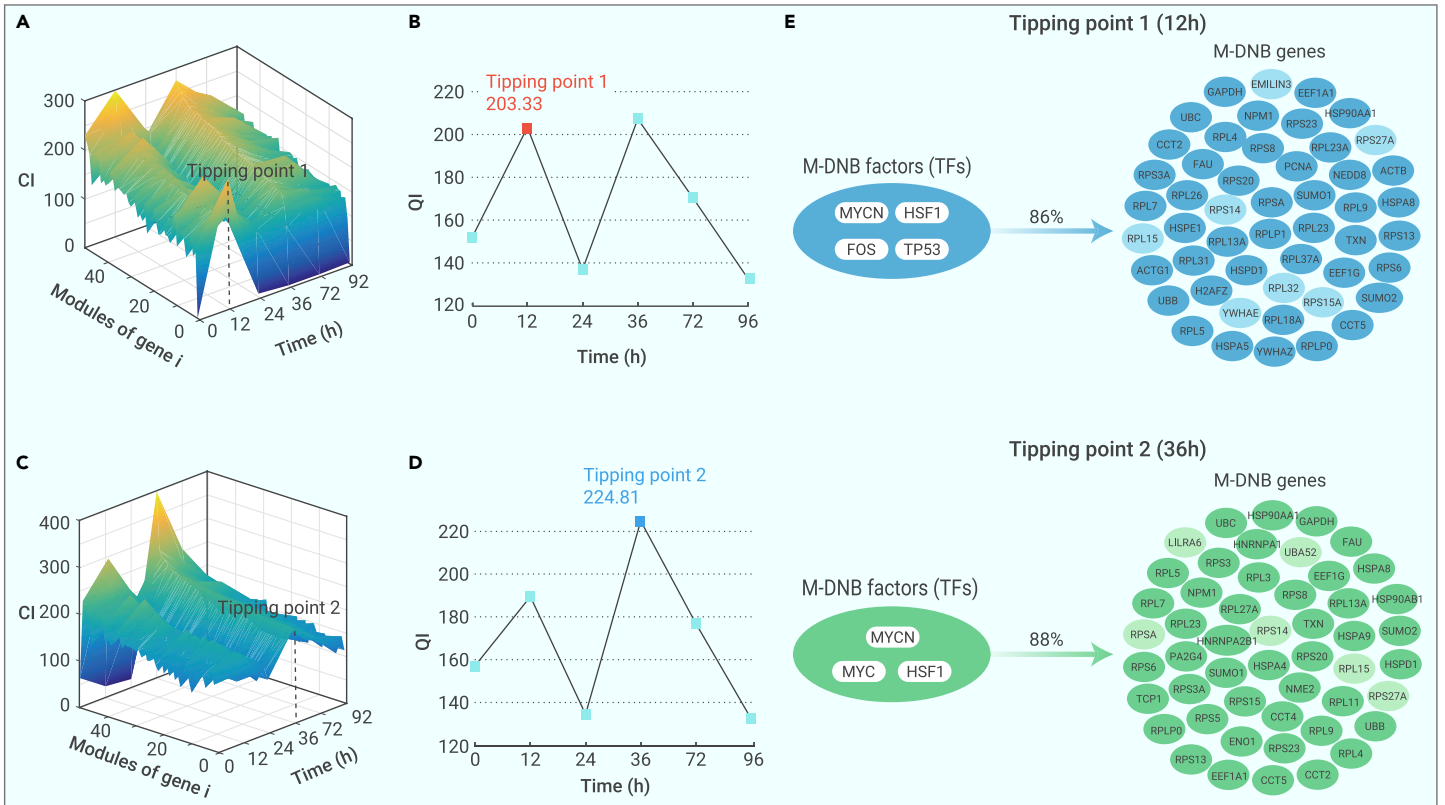
Figure 2. The tipping points of hESCs' differentiation process revealed by the M-DNB model (A) Behavior of M-DNB gene modules (M-DNB genes) with the top 50 CIs at 12 h in the hESCs' differentiation process. (B) QI scores of the gene modules with the top 50 CIs at 12 h in the hESCs' differentiation process. (C) Behavior of M-DNB gene modules (M-DNB genes) with the top 50 CIs at 36 h in the hESCs' differentiation process. (D) QI scores of the M-DNB gene modules (M-DNB genes) with the top 50 CIs value at 36 h in the hESCs's differentiation process. (E) Four hub upstream transcription factors (MYCN, HSF1, TP53, and FOS) could regulate 86% of M-DNB genes that were identified at 12 h (the first tipping point). Three hub upstream transcription factors (MYCN, MYC, and HSF1) could regulate 88% of M-DNB genes at 36 h (the second tipping point).

average correlation/association between the internal genes in a gene module *l* at a certain time point.

The third component (PCC_{out}) is the average correlation between the internal (inside the gene module *i*) and the external (outside the gene module *i*) genes, defined as follows:

$$PCC_{out} = \frac{1}{p} \sum_{j=1}^{p} cov(x_i, x_j) \bigg/ \sqrt{var(x_i)var(x_j)}, \quad \text{(Equation 4)}$$

where *p* is the number of edges between the gene module *i* and its first-order (or the second-order) neighbors; x_i is the gene inside the module, and x_j is gene outside of the module. Equation 4 reflects the average external association between the gene module *i* and all other genes at a certain time point. Note that we can also use CI socre by only the first and second components (Equations 2–3) without this third component (Equation 4). Thus, we can quantify CI score for each gene or each gene module, by Equation 1.

As the M-DNB criterion CI effectively combines the aforementioned three indices of the original DNB model, we can quantify the collective fluctuation by focusing on each gene. For each time point of the time course scRNA-seq data, we first calculate the CI for each gene and choose the top-*k* genes according to their CI scores. Then, the "quantitative indicator" (QI), which represents the average CI of the top *k* genes, is used to determine whether a certain time point is the tipping point. The time point with the highest QI score is designated as the tipping point. Genes with the top *k* CI value (e.g., *k* = 30) at this tipping point are defined as M-DNB genes. In summary, we developed an M-DNB model to analyze the time course scRNA-seq data, which can identify the tipping points with their major regulators in biological processes.

**Tipping points and DNB factors in endodermal differentiation of hESCs revealed by M-DNB analysis**

To interrogate the tipping points and their major regulators in the process of endodermal differentiation of hESCs at single-cell resolution, we analyzed a time course scRNA-seq dataset (Chu-time dataset),[23] which profiled a total of 758 cells at 0, 12, 24, 36, 72, and 96 h of differentiation (Note S1), representing three differentiation stages: ES, ME, and DE.

First, on the basis of the M-DNB model, we obtained the CIs of each gene module for six time points of differentiation (Figure 1B). The quantitative indicator of each time point was computed by averaging the top 50 CIs (Figure 1C). The peak QIs at 12 and 36 h indicate that these two time points were tipping points. The tipping points at 12 and 36 h may represent the critical points of ES-to-ME and ME-to-DE transitions. Next, to confirm the biological relevance of these two tipping points, we performed t-distributed stochastic neighbor embedding (t-SNE) to differentiate cells from the six time points (Figure 1D). SCPattern[24] was further applied to identify the differentially expressed genes (DEGs) of these time points, which revealed that the numbers of DEGs were significantly increased at 12 and 36 h compared with the rest of time points (Figure 1E). These results are consistent with the identification of tipping points by the M-DNB model. Taken together, 12 and 36 h were tipping points during the endodermal differentiation of hESCs.

Then, we identified M-DNB genes on the basis of the M-DNB model at the two tipping points. The gene modules with the top 50 CI values at 12 and 36 h were defined as M-DNB genes. As expected, the M-DNB genes at 12 h showed higher CIs and QI than those of the two neighboring time points (0 and 24 h) (Figures 2A and 2B), similarly to the M-DNB genes at 36 h (Figures 2C and 2D).

Last, we set out to identify the potential upstream transcriptional regulators of the M-DNB genes of the two tipping points. We focused on transcription factors (TFs) because they are key players that define the cell identity and drive cell-fate transitions.[25–27] We used an Ingenuity Pathway Analysis (IPA) to predict the TFs of M-DNB factors, which are defined as M-DNB factors, and identified four TFs/M-DNB factors (*MYCN, FOS, HSF1,* and *TP53*) for the first tipping point (12 h) and three TFs (*MYCN, HSF1,* and *MYC*) for the second tipping point (36 h) (Figures 2E and S1). These two groups of M-DNB factors could regulate 86% or 88% of the M-DNB genes of each tipping point, respectively (Figures 2E and S1). Among these five M-DNB factors, Myc, MycN, and P53 have been reported
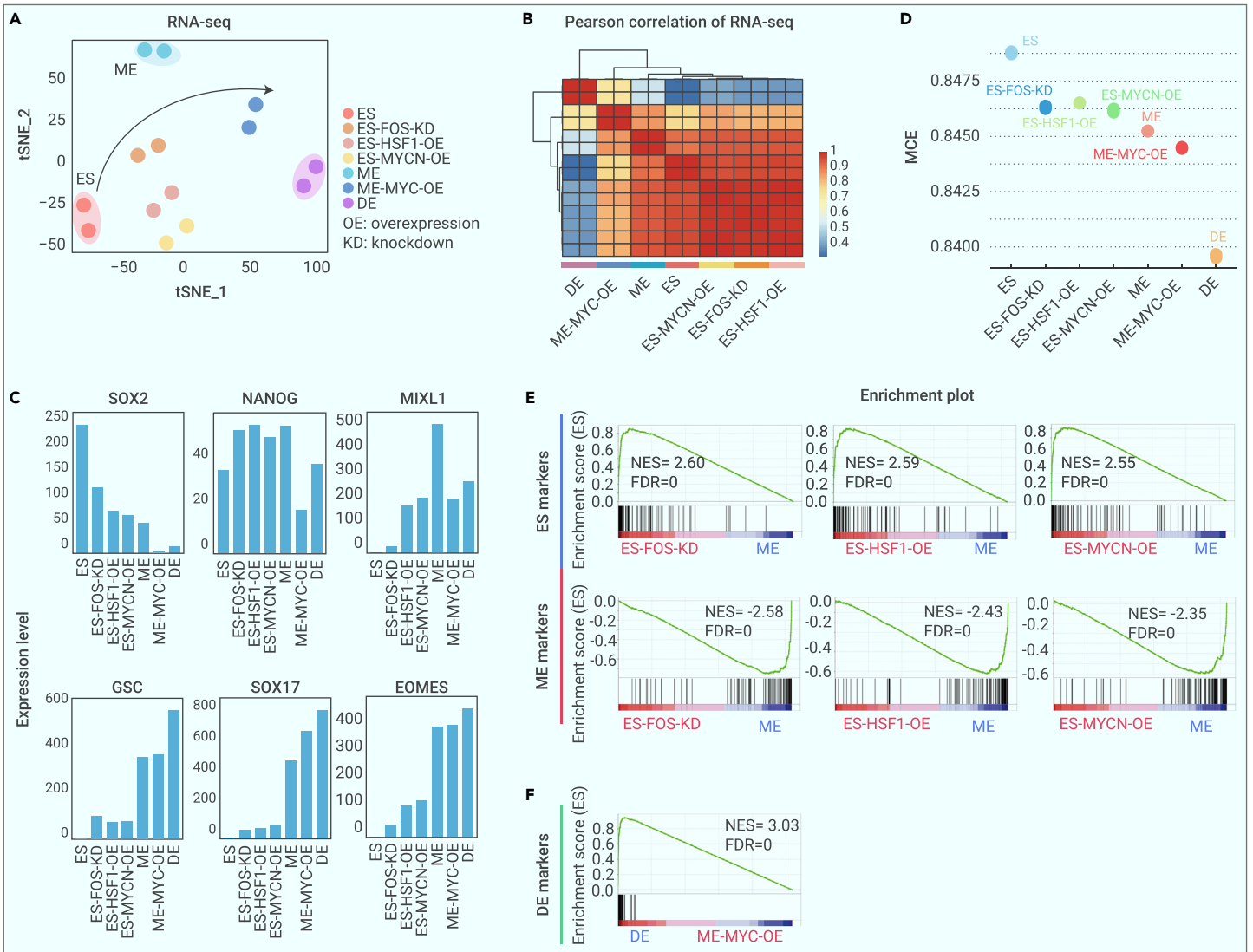
Figure 3. Transcriptomic analysis reveals the functional roles of M-DNB factors in hESCs' differentiation (A) t-SNE plot shows the distribution of all samples on the basis of bulk RNA-seq. (B) Heatmap showed the Pearson correlation of all samples on the basis of bulk RNA-seq. (C) The relative expression levels of markers of ES, ME, and DE. (D) The MCE values of ESCs, ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells, ME cells, ME-MYC-OE cells, and DE cells. (E) GSEA enrichment analysis of ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE versus ME cells using signatures of ESCs and ME cells. (F) GSEA enrichment analysis of ME-MYC-OE cells versus DE cells using signatures of DE cells.

to play essential roles in the maintenance and proliferation of ESC[10,12,28,29] or ME differentiation from hESCs,[9] while the functions of *FOS* and *HSF1* in the early differentiation of hESCs have never been investigated.

Taken together, using the M-DNB model, we identified two tipping points during the endodermal differentiation of hESCs and predicted five M-DNB factors that may play key regulatory roles in the cell-fate determination in this process.

### Validation of M-DNB factors at the tipping points by RNA-seq data

After identifying the M-DNB factors using the M-DNB model (Figures 1 and 2), we further validated the roles of the five M-DNB factors in endodermal differentiation of hESCs at the transcriptomic and chromosome accessibility levels. We performed gene perturbation assays by using knockdown or overexpression and obtained related bulk RNA-seq and assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (Figure S2; Note S1). We first performed RNA-seq on the wild-type cells harvested at time points along the endodermal differentiation (ES, ME, and DE cells), as well as on the cells in which the expression of M-DNB factors had been disrupted before the tipping points (ES-HSF1-OE, ES-MYCN-OE, and ES-FOS-KD for ES-to-ME transition; ME-MYC-OE for ME-to-DE transition).

t-SNE analysis of the RNA-seq data revealed that ES, ME, and DE cells were distantly positioned, which represents the distinct developmental stages where they arrived, and that ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells were distributed between ES and ME cells (Figure 3A). ME-MYC-OE cells were

located close to DE cells. Pearson correlation analysis also revealed that ES-FOS-KD, ES-HSF1-OE, and ES-MYCN-OE cells had significantly stronger correlations with one another and with ESCs than with ME cells (Figure 3B). Moreover, there were relatively high correlations among ME-MYC-OE cells with DE cells.

We then examined the expression patterns of the hallmark genes of ES, ME, and DE stages across the wild-type and the manipulated cells. For the ES-to-ME transition, we observed that ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells displayed higher expression levels of ES signatures, including *POU5F1*, *SOX2*, and *NANOG*, and lower expression levels of ME signatures, including *MIXL1*, *GSC*, and *EOMES* (Figure 3C). Moreover, ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE were enriched for Gene Ontology (GO) terms related to tissue development and cell proliferation, suggesting undifferentiated or progenitor properties of these cells compared with the nondividing, differentiated, wild-type ME cells (Figures S3A–S3C). During the ME-to-DE transition, ME-MYC-OE cells expressed higher levels of ME's signature gene, *GSC*, rather than those of DE's, *SOX17* (Figure 3C).

Furthermore, we performed Markov-chain entropy (MCE) analysis[30] to evaluate the differentiation potency of the manipulated cells. MCE analysis was developed to predict the developmental potential of the cells on the basis of Markov process. MCE analysis revealed a higher potency of ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells than that of wild-type ME cells, as well as a higher potency of ME-MYC-OE cells than that of the wild-type DE cells (Figure 3D).
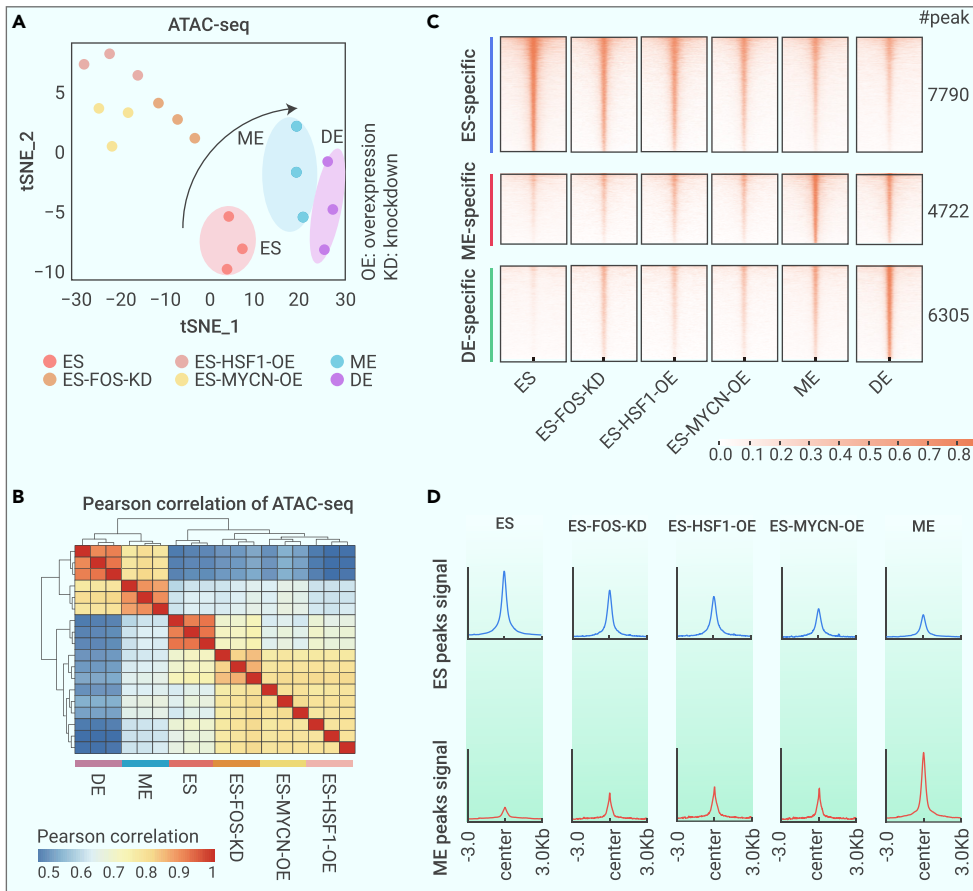
Figure 4. The specific accessible chromatin regions of ESCs, ME cells, and DE cells reveal the cell states of ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells (A) t-SNE plot shows the distribution of all samples on the basis of ATAC-seq. (B) The heatmap shows the Pearson correlation of all samples on the basis of ATAC-seq. (C) Heatmap of ATAC-seq data around the peak center of ES-generated binding sites, ME-generated binding sites, and DE-generated binding sites. The numbers of peaks for ESCs, ME cells, and DE cells are also shown. (D) Profile plot of ES-specific binding site and ME-specific binding site showed in ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells and ME cells.

Last, to further establish the relationship of manipulated cells with the wild-type ES, ME and DE cells, we next performed gene set enrichment analysis (GSEA) using GSEA version 4.1.0 software with 1,000 gene-set permutations. The ES, ME, and DE signature genes were set as the gene set database for further analyses. Remarkably, GSEA revealed that the ES signature genes were significantly and positively correlated with the ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells, but not ME cells, while the ME signature genes were negatively correlated with the ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells (Figure 3E). Moreover, the DE signature genes were negatively correlated with the ME-MYC-OE cells (Figure 3F). These results indicate that the state of ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells are developmentally closer to ESCs than ME cells, while ME-MYC-OE cells have not yet reached the DE stage (Figures 3E and 3F).

Collectively, these results clearly show that the perturbation of the expression of M-DNB factors before the tipping points delayed the ME or DE differentiation from hESCs.

Validation of M-DNB factors at the tipping points by ATAC-seq data

We next attempted to clarify the modulatory effect of M-DNB factors on chromosome accessibility in endodermal differentiation of hESCs (Note S1). ATAC-seq was performed on the wild-type cells and the manipulated cells, in which the expression of M-DNB factors wasas disrupted before the tipping points, as mentioned earlier.

Consistent with what was observed with RNA-seq, t-SNE and Pearson correlation analyses on ATAC-seq data revealed that ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells showed higher similarity and correlation with wild-type ESCs than with ME cells (Figures 4A and 4B). We first compared chromatin accessibility among ES, ME, and DE cells using edgeR to distinguish cell populations. We observed that the ES-specific chromosome accessibility regions reduced in intensity as the cells differentiated into ME or DE (Figures 4C and S4) but retained similar intensities in ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells as those of ESCs, which indicates that the manipulated cells resembled ES rather than ME cells. However, the intensity of ME-specific regions

(Figures 4D and S4B) reduced in wild-type ES, ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells, as expected. We further identified the differential peaks of ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells, and applied ChIPseeker[31] to retrieve the nearest genes around these peak sets. The enrichment analysis showed that ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells were enriched in tissue development (Figure S5).

In summary, data analyses of ATAC-seq on the wild-type and the manipulated cells further confirmed the important stage-specific roles of M-DNB factors in regulating the chromosome accessibilities prior to the tipping points of the endodermal differentiation of hESCs.

Validation of M-DNB factors at the tipping points by loss-of-function or gain-in-function assays

To validate the roles of the five M-DNB factors in endodermal differentiation of hESCs, we first performed loss-of-function or gain-of-function assays of the three M-DNB factors (HSF1, FOS, and MYCN) (Figure S6; Note S1). Our hypothesis is that the perturbations of the expression of M-DNB factors prior to the tipping points prevent cell-fate transitions. The expression patterns of three M-DNB factors revealed by qRT-PCR (Figure 5A) confirmed the upregulations of FOS and HSF1 and the downregulation of MYCN during ES-to-ME transition prior to tipping point 1 (12 h). Similarly, the downregulation of MYCN were observed during the ME-to-DE transition before tipping point 2 (36 h).

To induce the loss-of-function or gain-of-function of M-DNB factors, we performed lentiviral-mediated short hairpin RNA (shRNA) knockdown or overexpression on the hESCs 12 h before the tipping points. For ES-to-ME transition, HSF1 (ES-HSF1-OE), MYCN (ES-MYCN-OE) were overexpressed, and FOS (ES-FOS-KD) were knocked down at 0 h (Figure 5A). As expected, the upregulations of pluripotent markers (OCT4 and SOX2) and the downregulations of ME markers (EOMES, MIXL1, and GSC) were observed when the expressions of HSF1, MYCN, or FOS were disrupted 12 h ahead of the tipping point 1 compared with those of the wild-type control (Figure 5B). For ME-to-DE transition, the overexpression of MYCN (ME-MYCN-OE) was induced at 24 h. We observed the maintenance of expression of ME marker subsets (EOMES, MIXL1, GSC) and the
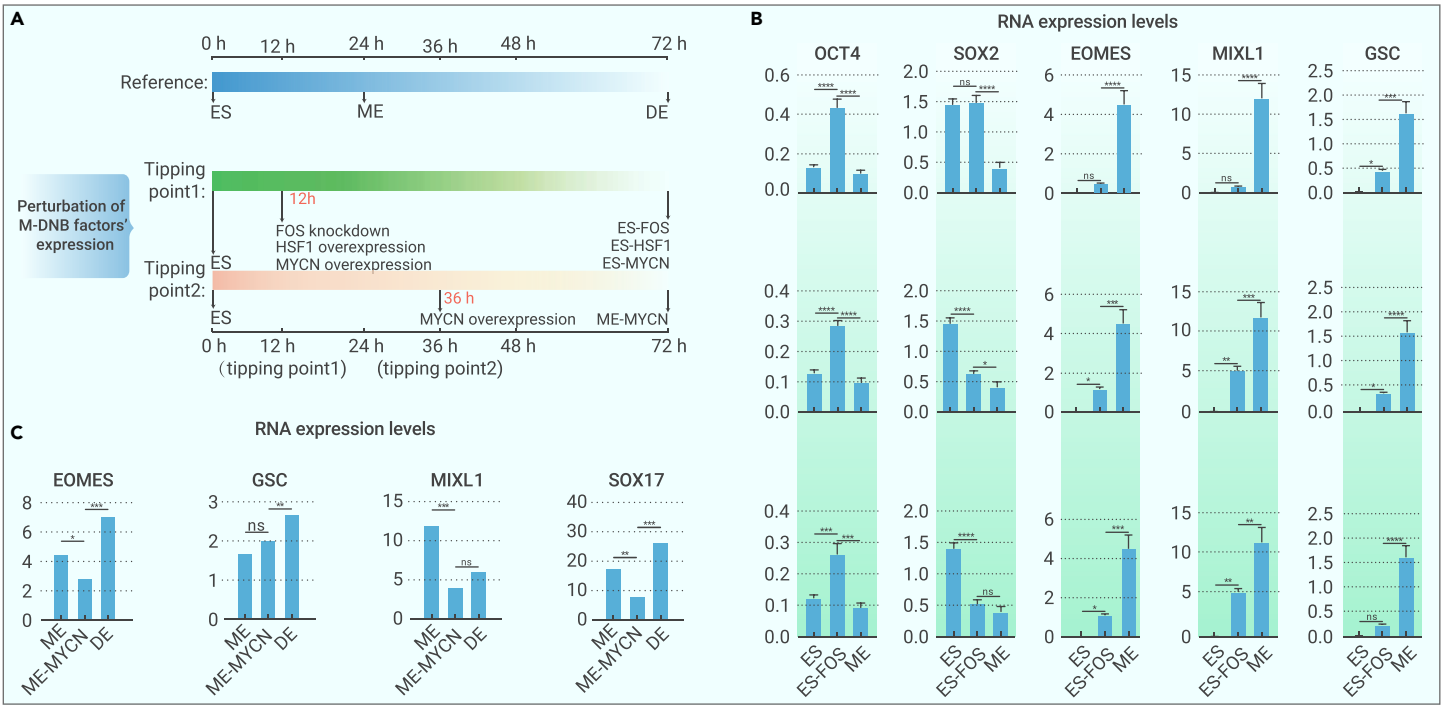
Figure 5. Perturbation experiment of M-DNB factors in DE differentiation system (A) Experimental flow of M-DNB factors' perturbation and hESCs' differentiation. Reference is the process of hESCs' differentiation, and we obtained ME cells at 24 h, DE cells at 72 h of hESCs' differentiation. Perturbation of M-DNB factors' expression consists of two parts. In tipping point 1, we knock down FOS and overexpress HSF1 and MYCN, respectively, at 12 h (nearly) and obtained ES-FOS-KO cells, ES-HSF1-OE cells, and ES-MYCN-OE cells at 72 h of differentiation. In tipping point 2, we overexpress MYCN, respectively, at 36 h (nearly) and obtained ME-MYCN-OE cells at 72 h of differentiation. (B) RNA expression levels of ES and ME stage marker genes. (C) RNA expression levels of ME and DE stage marker genes. OCT4, and SOX2 as ES marker genes; EOMES, MIXL1, and GSC as ME markers; SOX17 as DE marker genes.

downregulation of DE markers (SOX17) in the manipulated differentiating cells (Figure 5C).

These results show that the perturbations in the expression of the M-DNB factors prior to the tipping points disrupted the normal differentiation and resulted in delays in cell-fate transitions, which is in line with the previous studies that revealed the essential roles of the P53, *MYCN*, and *MYC* in endodermal differentiation of hESCs.[9,11,12,28,29]

DISCUSSION

It has been well established that hESCs can be used as an effective *in vitro* model to recapitulate *in vivo* developmental programs and to dissect the molecular mechanism underpinning lineage specification and differentiation. Induced pluripotent stem cells (iPSCs) provide a novel method to investigate the mechanisms of stem cell pluripotency and self-renewal. Some transcriptional factors have been shown to be pivotal for ESCs' differentiation.[9,12,28,29] From the viewpoint of dynamical systems, molecular events at the tipping points of ESCs' differentiation are considered to determine cell-fate determination. Thus, identifying the tipping points and their major regulators (TFs) is of great importance for understanding of the mechanisms of stem cell pluripotency and ESCs' differentiation. Despite the progress made on elucidating the mechanisms underlying the differentiation of hESCs, it remains unclear how to identify the tipping points of these processes and further reveal their molecular regulations on these critical events. Here, we identified the two tipping points of the differentiation of hESCs toward DE through an intermediate stage of ME by designing an M-DNB model and further found the key regulators/M-DNB factors (*FOS, HSF1, MYCN, TP53*, and *MYC*) of this process on the basis of time course single-cell transcriptomic analyses. The stage-specific and essential roles of these M-DNB factors were confirmed by the differentiation experiments involving knockdown or overexpression of these transcription factors, which were analyzed with time course RNA-seq and ATAC-seq. In particular, we demonstrated that M-DNB factors were could maintain the cells states and orchestrate cell-fate determination before the tipping points during ES-to-ME and ME-to-DE differentiation processes. The results also demonstrate the power of the M-DNB model for analyzing massive single-cell RNA sequencing data, which may to identify critical points and their key factors in diverse biological processes, including cell differentiation and trans-differentiation dynamics.

A wealth of research has led to an increasingly finely tuned understanding of the tipping points of complex dynamical systems. Moreover, DNB theory has been applied to multiple biology research areas, such as detecting the tipping point of the endocrine resistance process in breast cancer[32] and identifying the tipping points of metastasis of hepatocellular carcinoma.[5] However, it is a challenge to identify tipping points on the basis of scRNA-seq data because of noisy data and dropout problems. Here, we first developed the M-DNB model, which can quantitatively pinpoint the tipping points and their key regulators in biological processes from single-cell RNA sequencing data. On the basis of DNB theory, our M-DNB model can be applied not only in such a differentiation process but also in many other biological processes that contain multiple stages, states, or time points. For instance, the M-DNB model can be applied to identify the pre-resistance state and the pre-metastatic state of cancer and to detect associated M-DNB genes, which helps in the treatment of complex diseases.

Then, we applied the M-DNB model to reliably and accurately identify the tipping points and their critical factors during the hESCs' differentiation process on the basis of a time course scRNA-seq dataset (Chu-time dataset).[23] Two tipping points and five M-DNB factors (*MYCN, MYC, FOS, HSF1,* and *TP53*) were identified, which regulate two cell-state transitions, ES-to-ME differentiation/transition and ME-to-DE differentiation/transition, in hESCs' differentiation process. Interestingly, the three factors (*MYCN, MYC,* and *TP53*) among the five M-DNB factors have been reported to be essential for stem cell differentiation, which is consistent with our analysis results of the M-DNB model and data.

We further performed an integration analysis of our RNA-seq and ATAC-seq data to demonstrate the functional roles of M-DNB factors by conducting overexpression or knockdown experiments for the identified M-DNB factors at the two tipping points in hESCs' differentiation, which validated the functional roles of the M-DNB factors. Specifically, ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells showed higher potency and more similarity of ESCs compared with ME cells. Moreover, there was a higher intensity of ESCs' regions in ES-FOS-KD cells, ES-HSF1-OE cells, and ES-MYCN-OE cells. In contrast, these cells showed a lower intensity of ME cells' regions. As a whole, M-DNB factors (HSF1, FOS, MYCN) can effectively maintain the state of cells in ES-to-ME differentiation and ME-to-DE differentiation.

Moreover, we established M-DNB factors expression in reversed hESC lines in ES-to-ME differentiation and two reversed hESC lines in ME-to-DE differentiation

to validate the regulatory functions of the M-DNB factors at the tipping point. Cells after reversing the expression trends of *MYCN*, *HSF1*, or *FOS* in ES-to-ME differentiation showed high expression of ES markers and low expression of ME markers. Cells after reserving expression trends of *MYCN* in ME-to-DE differentiation showed low expression of DE markers. These results suggest that four M-DNB factors orchestrate cell-fate determination and maintain the cell state in ES-to-ME differentiation and ME-to-DE differentiation before the tipping points. Such analysis of hESCs' differentiation on the basis of the tipping points not only helps unveil tipping points and key factors in diverse biological processes, but also enhances the understanding of the cell-fate decision.

MATERIALS AND METHODS

See Supplemental Information for details on Materials and Methods.

REFERENCES

1. Lewis, S.L., and Tam, P.P.L. (2006). Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function. Dev. Dyn. **235**, 2315–2329.
2. Wang, L., and Chen, Y.G. (2016). Signaling control of differentiation of embryonic stem cells toward mesendoderm. J. Mol. Biol. **428**, 1409–1422.
3. Sumi, T., Tsuneyoshi, N., Nakatsuji, N., and Suemori, H. (2008). Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. Development **135**, 2969–2979.
4. Chen, L., Liu, R., Liu, Z.P., et al. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci. Rep. **2**, 342.
5. Yang, B., Li, M., Tang, W., et al. (2018). Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat. Commun. **9**, 678.
6. Cliff, T.S., Wu, T., Boward, B.R., et al. (2017). MYC controls human pluripotent stem cell fate decisions through regulation of metabolic flux. Cell Stem Cell **21**, 502–516.e9.
7. Loh, K.M., Ang, L.T., Zhang, J., et al. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. Cell Stem Cell **14**, 237–252.
8. Tosic, J., Kim, G.J., Pavlovic, M., et al. (2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. Nat. Cell Biol. **21**, 1518–1531.
9. Wang, Q., Zou, Y., Nowotschin, S., et al. (2017). The p53 family coordinates Wnt and Nodal inputs in mesendodermal differentiation of embryonic stem cells. Cell Stem Cell **20**, 70–86.
10. Zhang, J.T., Weng, Z.H., Tsang, K.S., et al. (2016). MycN is critical for the maintenance of human rmbryonic stem cell-derived neural crest stem cells. PLoS One **11**, e0148062.
11. Knoepfler, P.S., Cheng, P.F., and Eisenman, R.N. (2002). N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. Genes Dev. **16**, 2699–2712.
12. Varlakhanova, N.V., Cotterman, R.F., deVries, W.N., et al. (2010). Myc maintains embryonic stem cell pluripotency and self-renewal. Differntiation **80**, 9–19.
13. Nelson, A.C., Cutty, S.J., Niini, M., et al. (2014). Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. BMC Biol. **12**, 81.
14. D'Amour, K.A., Agulnick, A.D., Eliazer, S., et al. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. **23**, 1534–1541.
15. Zorn, A.M., and Wells, J.M. (2009). Vertebrate endoderm development and organ formation. Annu. Rev. Cell Dev. Biol. **25**, 221–251.
16. Richard, A., Boullu, L., Herbach, U., et al. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLoS Biol. **14**, e1002585.
17. Liu, R., Chen, P., Aihara, K., et al. (2015). Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. Sci. Rep. **5**, 17501.
18. Grün, D., Lyubimova, A., Kester, L., et al. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature **525**, 251–255.
19. Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science **343**, 776–779.
20. Wen, L., Li, G., Huang, T., et al. (2022). Single-cell technologies: from research to application. Innovation **3**, 100342.
21. Dai, H., Li, L., Zeng, T., et al. (2019). Cell-specific network constructed by single-cell RNA sequencing data. Nucleic Acids Res. **47**, e62.
22. Li, L., Dai, H., Fang, Z., et al. (2021). c-CSN: single-cell RNA sequencing data analysis by conditional cell-specific network. Genom. Proteom. Bioinform. **19**, 319–329.
23. Chu, L.F., Leng, N., Zhang, J., et al. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. **17**, 173.
24. Leng, N., Chu, L.-F., Choi, J., et al. (2016). SCPattern: a statistical approach to identify and classify expression changes in single cell RNA-seq experiments with ordered conditions. Preprint at bioRxiv. https://doi.org/10.1101/046110.
25. Fong, A.P., and Tapscott, S.J. (2013). Skeletal muscle programming and re-programming. Curr. Opin. Genet. Dev. **23**, 568–573.
26. Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. **17**, 183–193.
27. Lambert, S.A., Jolma, A., Campitelli, L.F., et al. (2018). The human transcription factors. Cell **172**, 650–665.
28. Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., et al. (2016). Myc depletion induces a pluripotent dormant state mimicking diapause. Cell **164**, 668–680.
29. Smith, K.N., Singh, A.M., and Dalton, S. (2010). Myc represses primitive endoderm differentiation in pluripotent stem cells. Cell Stem Cell **7**, 343–354.
30. Shi, J., Teschendorff, A.E., Chen, W., et al. (2018). Quantifying Waddington's epigenetic landscape: a comparison of single-cell potency measures. Brief. Bioinform. **21**, 248–261.
31. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics **31**, 2382–2383.
32. Liu, R., Wang, J., Ukai, M., et al. (2019). Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. J. Mol. Cell Biol. **11**, 649–664.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

L.C., X.C., D.G., and Y.L. conceived and designed this work. L.L. and L.Y. designed the algorithm and analyzed data. Y.X., F.L., and Z.L. performed experiments. L.L. and L.Y. wrote the source code used in this work. L.L., L.C., X.C., and D.G. drafted the manuscript. L.C. and X.C. revised the manuscript. F.L., X.L., and C.Z. proposed valuable advice for revising the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

All raw sequencing data created in this study have been uploaded to the National Omics Data Encyclopedia (NODE; https://www.biosino.org/node/project/detail/OEP003324) with accession number OEP003324.

Code of M-DNB model and related analysis are available at https://github.com/LinLi-0909/M-DNB-model.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.1016/j.xinn.2022.100364.

LEAD CONTACT WEBSITE

http://cemcs.cas.cn/sourcedb_cemcs_cas/zw/pi/202008/t20200823_5670080.html.

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 ($4.5 \times 10{-4}$ 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 ($4.5 \times 10{-4}$ 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^{TM} 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: $\log(\text{fold change}) \geq 1$). 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 $\log(\text{fold change}) \geq 1$.

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 ($\log2(1.1) \approx 0.13$) 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.

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).
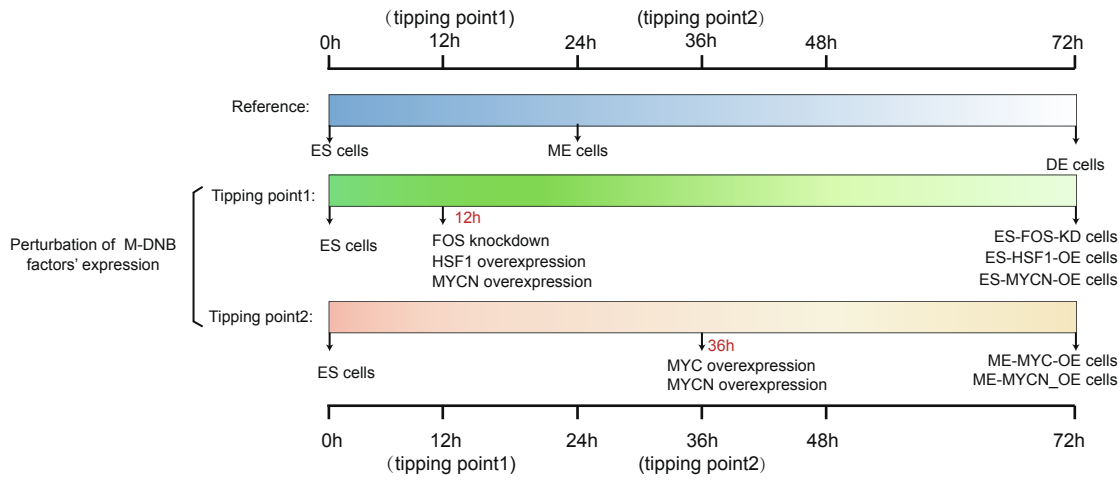
Figure. S2 Experimental flow of M-DNB factors' perturbation and hESCs' differentiation. Reference is the process of hESCs' differentiation, and we obtained ME cells at 24h, DE cells at 72h of hESCs' differentiation. Perturbation of M-DNB factors' expression consists of two parts. In tipping point 1, 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.
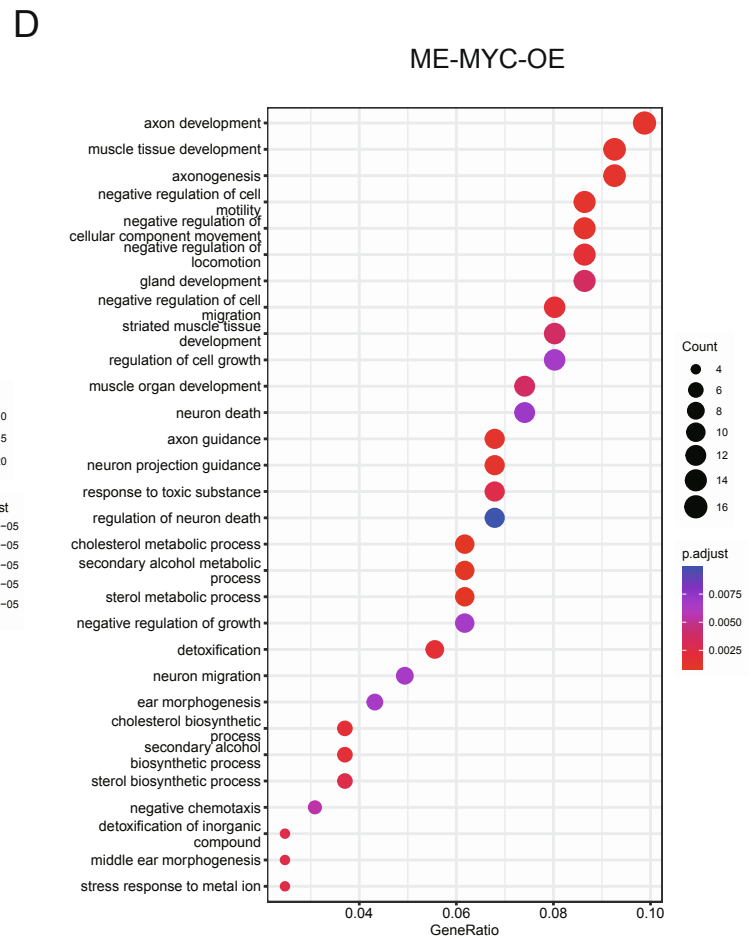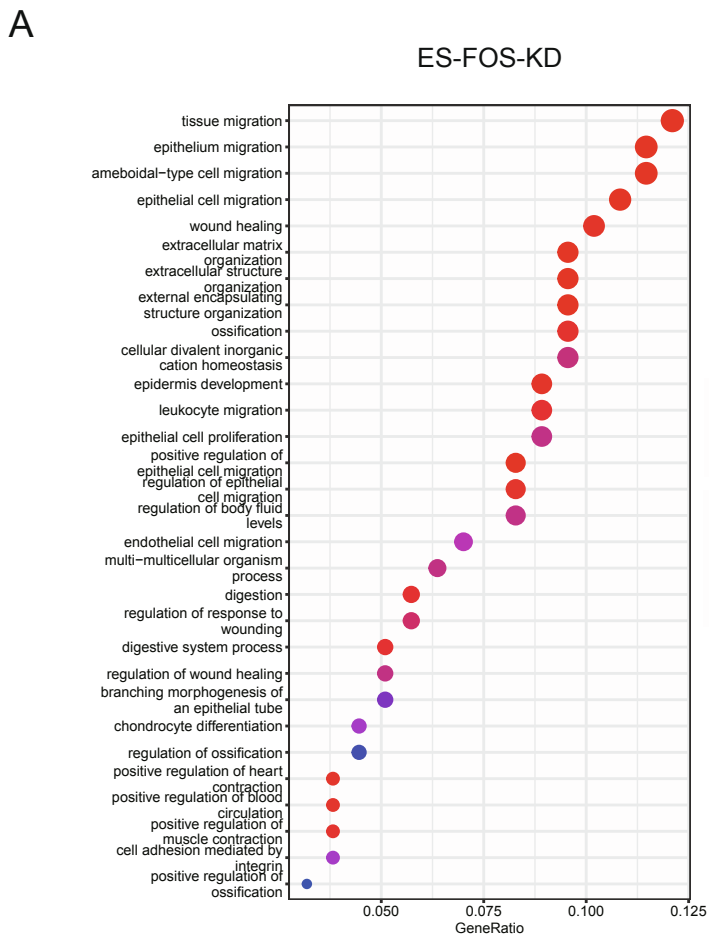
Figure. S3 Dot plots show the significantly enriched gene ontology terms of ES-FOS-KD (A), ES-HSF1-OE (B), ES-MYCN-OE (C), and ME-MYC-OE (D) for RNA-seq data.
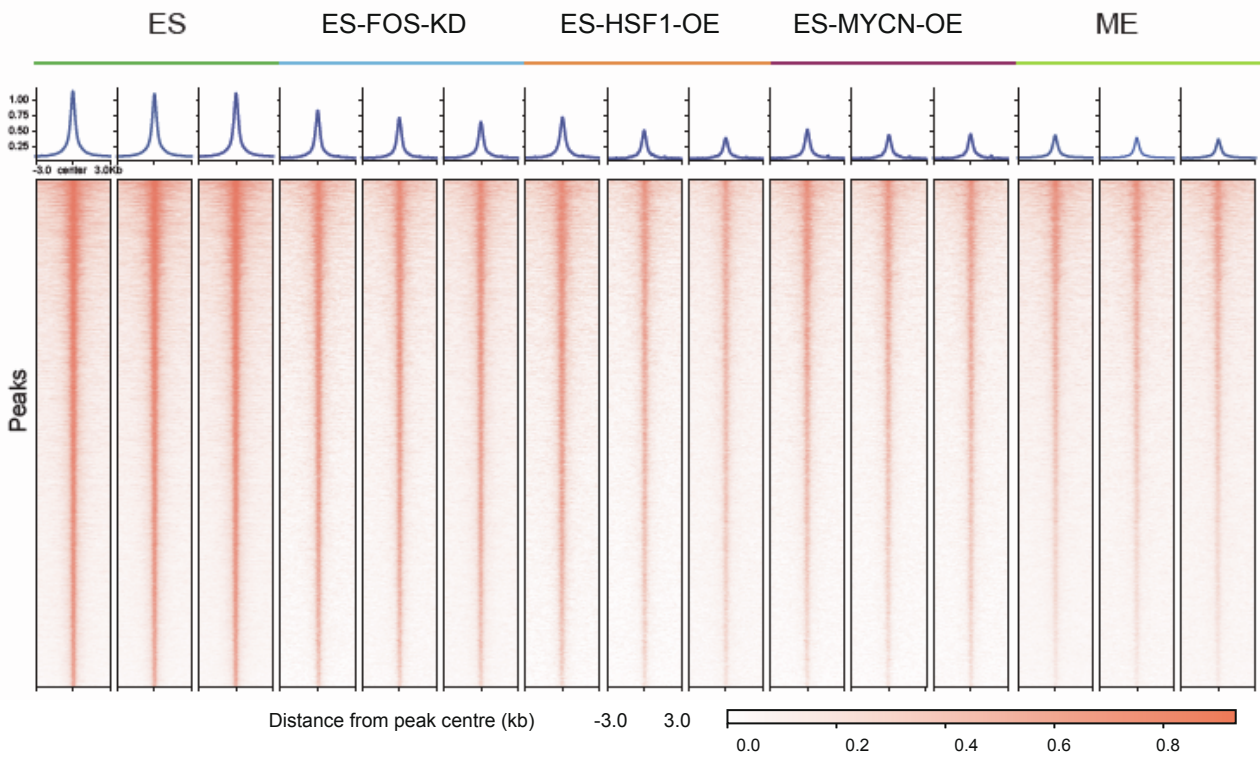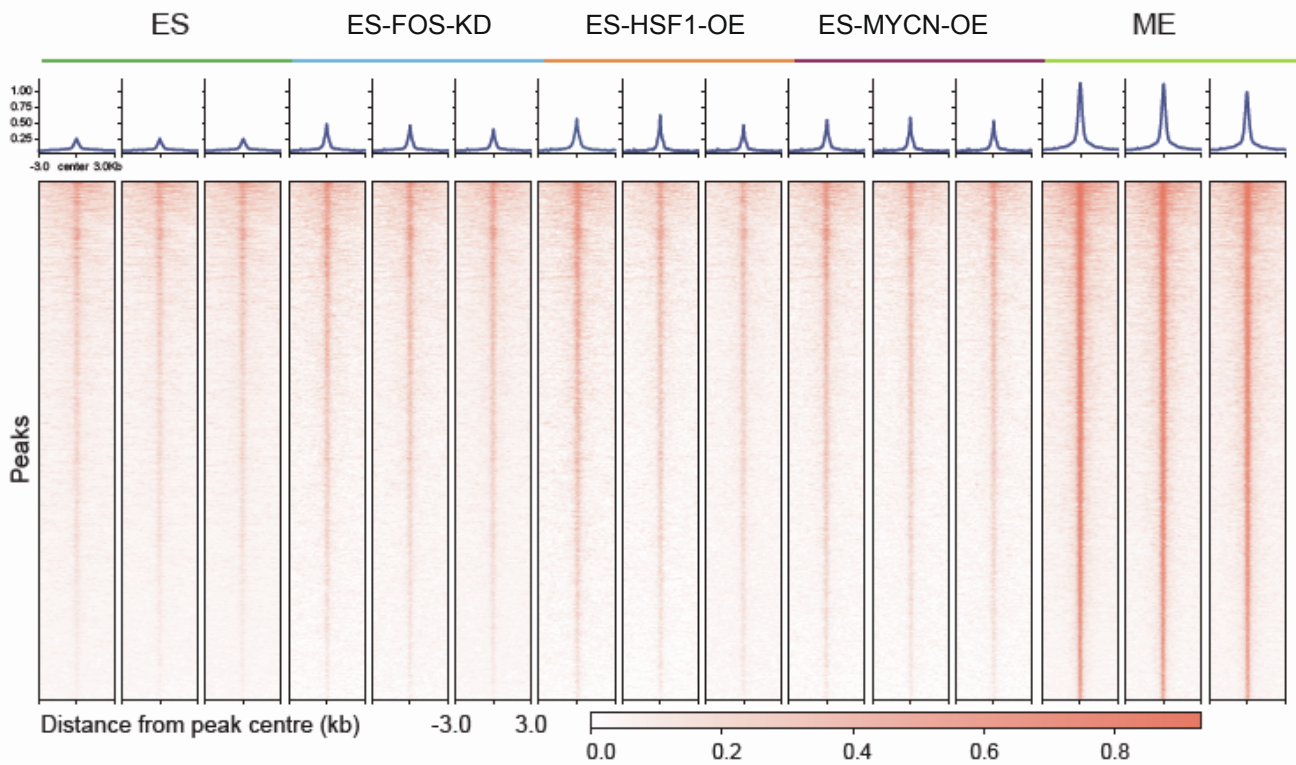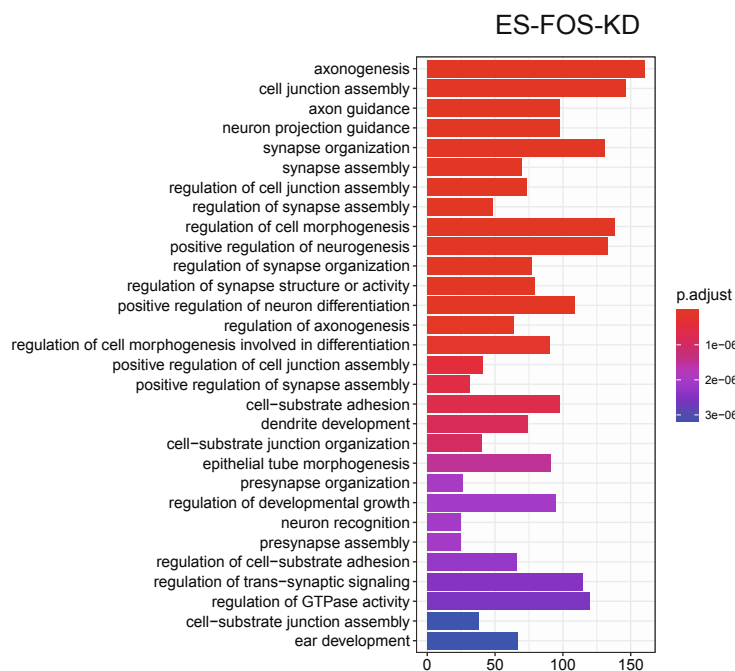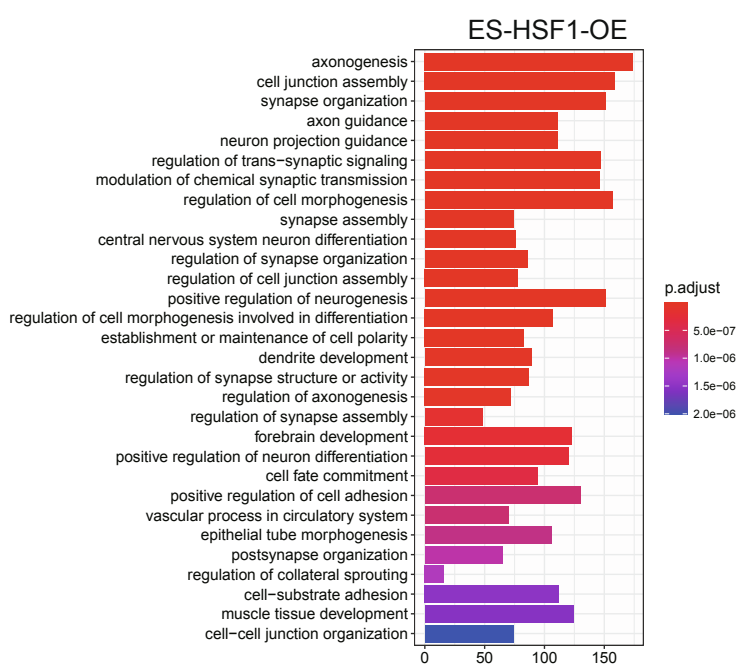
Figure.S4 (A) Profiles (top) and Heatmap of ES, ES-FOS-KD, ES-HSF1-OE, ES-MYCN-OE and
ME cells around the peak center of ES binding sites.
(B) Profiles (top) and Heatmap of ES, ES-FOS-KD, ES-HSF1-OE, ES-MYCN-OE and ME cells
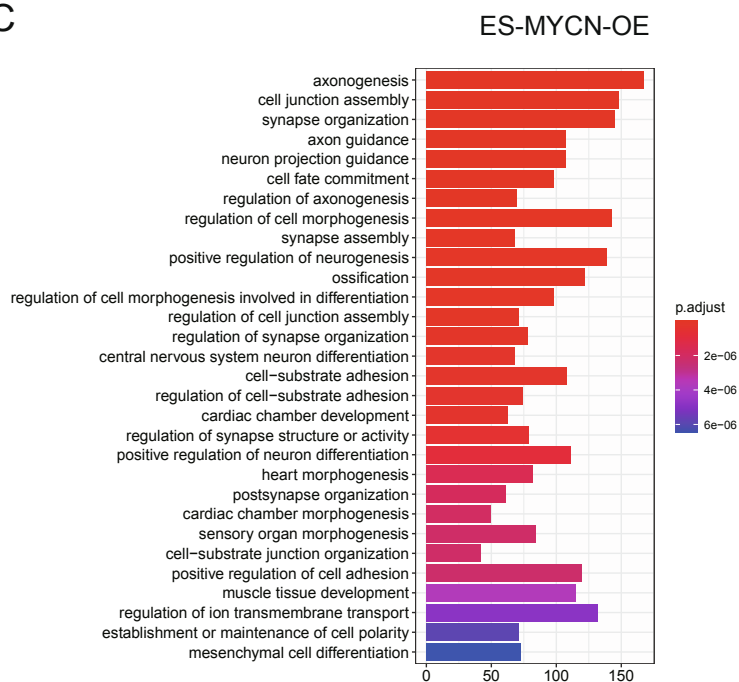around the peak center of ME binding sites.

Figure.S5 Bar plots show the significantly enriched gene ontology terms of ES-FOS-KD

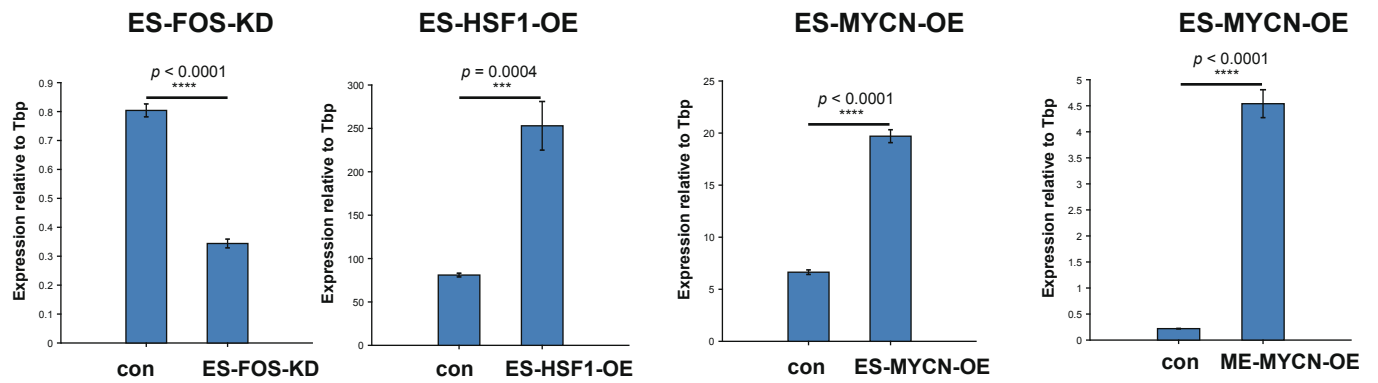(A), ES-HSF1-OE (B), ES-MYCN-OE (C) for ATAC-seq data.

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.