Single-Cell Reprogramming of Mouse Embryo Development Through a Critical Transition State

Masa Tsuchiya[1*], Alessandro Giuliani[2], and Kenichi Yoshikawa[3]

[1]Systems Biology Program, School of Media and Governance, Keio University, Fujisawa, Japan

[2]Environment and Health Department, Istituto Superiore di Sanitá, Rome, Italy

[3]Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe, Japan

Corresponding author: tsuchiya.masa@gmail.com

Abstract

Our statistical thermodynamics approach to the temporal development of the genome-expression profile in single-cell mouse early embryo indicated that reprogramming occurs via a critical transition state, where the critical-regulation pattern of the zygote state disappears. In this report, we unveil the detailed mechanism of how the dynamic interaction of thermodynamic states (critical states) enables the genome system to pass through the critical transition state to achieve genome reprogramming.

Self-organized criticality (SOC) control of overall expression provides a snapshot of self-organization and explains the coexistence of critical states at a certain experimental time point. The time-development of self-organization is dynamically modulated by exchanges in expression flux between critical states through the cell nucleus milieu, where sequential global perturbations involving activation-inhibition of multiple critical states occur from the early state to the late 2-cell state. Two cyclic fluxes act as feedback flow and generate critical-state coherent oscillatory dynamics. Dynamic perturbation of these cyclic flows due to vivid activation of the ensemble of low-variance expression (sub-critical state) allows the genome system to overcome a transition state during reprogramming.

Our findings imply that a universal mechanism of long-term global RNA oscillation underlies autonomous SOC control, and the critical gene ensemble of a critical point drives genome reprogramming. Unveiling the corresponding molecular players will be essential to understand single-cell reprogramming.

Key words: Single-cell early embryo development; Single-cell Reprogramming; Statistical thermodynamics; Transition state; Self-organized criticality; Critical states; Critical gene ensemble

Introduction

In mammalian embryo development, a large number of molecular-level epigenetic studies [1-3] have been conducted to reveal the occurrence of stunning global epigenetic modifications on chromatins (DNA + histones) associated with reprogramming processes. However, the genome-wide principle that drives such extremely complex epigenetic modifications is still unknown.

In our previous studies, based upon transcriptome experimental data for seven distinct cell fates [4], we recognized that a self-organized critical transition (SOC) in whole-genome expression plays an essential role in the change in the genome expression state at both the population and single-cell levels (see attached **Supplementary S1 file**). In early mammalian embryo development, SOC induces (with different timings in mouse and human embryos) a transition event toward a stochastic expression distribution. The massive change in genome expression destroys the SOC control of the initial (only maternal) stage of embryogenesis to allow for a 'new start' that corresponds to activation of the new genetic pattern with both maternal and paternal origins. This global restructuring of gene expression occurs at the two-cell stage in mouse, precisely between the mid and late 2-cell states, and between the 4- and 8-cell stages in humans [5]. The transition period depends on the switch from the use of maternally prepared stable gene transcripts to the initiation of proper embryonic genome transcription activity.

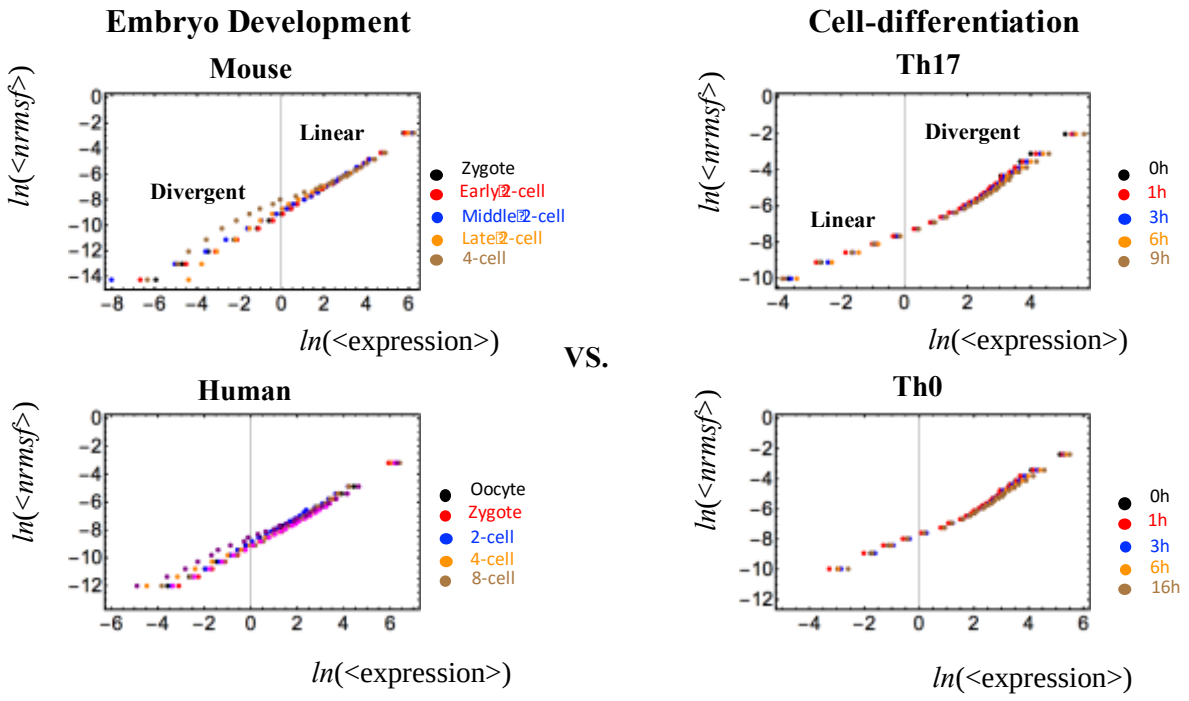

Figure 1: Opposite genome avalanche behaviors in early embryo development and immune cell differentiation: *Genome avalanches, scaling-divergent behaviors in overall expression, as important features of the SOC control of overall expression (see more in* ***Supplementary S1 file****) are evident in the log-log plot of averaging behaviors: (left panel: mouse (first row) and human (second) embryo development; right panel; Th17 (first) and Th0 (second) terminal cell differentiation). Log-log plots represent the natural logarithm of the group average (< >) of expression (x-axis) and nrmsf (y-axis) (n = 685 (mouse), 666 (human), 525 (Th17) and 504 RNAs (Th0) for each dot), where overall expression is sorted and grouped (25 groups) according to the degree of nrmsf. Two distinct biological processes (reprogramming in early embryo development versus immune cell differentiation) show opposite scaling-divergent behaviors. Scaling behavior occurs in the ensemble of high-variance RNA expression (high region of nrmsf: super-critical state; see* ***Figure 3A****) in early embryo development and divergent behavior in the ensemble of low-variance RNA expression (low region of nrmsf: sub-critical state), whereas the T cell terminal cell fate (single cell) has opposite behaviors.*

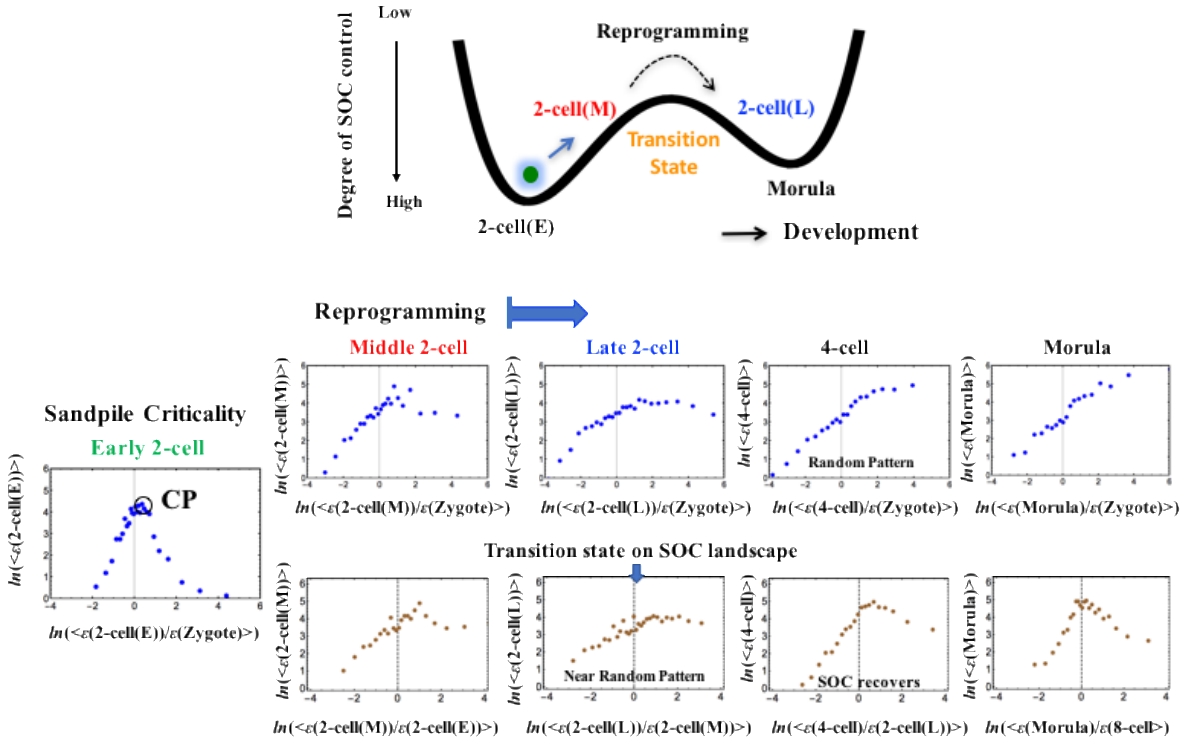

Figure 2: Timing of genome-state change on the SOC-control landscape through a transition state in a single cell: *The timing of the genome-state change was revealed in the erasure of sandpile-type criticality of the zygote state. The erasure of zygote criticality (second row) occurs after the middle 2-cell state to reveal a stochastic expression pattern as a linear correlative behavior (refer to S2 Fig. in [4]). The transition of SOC control through non-SOC control (stochastic pattern) shows an SOC-control landscape: a valley (SOC control: zygote - early 2-cell states) - ridge (non SOC control: middle - late 2-cell states) - valley (SOC control: 8-cell - morula states); a high degree of SOC control for a well-developed shape of the sandpile-type criticality, an intermediate degree of SOC control for a*

weakened (broken) sandpile, and a low degree for non-SOC control. At the middle - later 2-cell states (third row), sandpile-type criticality disappears and recovers thereafter; this indicates the existence of a transition state in the SOC control landscape (see the detailed mechanism in ***Figure 7****). Sandpile-type critical behaviors, which exhibit diverging up- and down-regulation at a critical point (CP), emerge in overall expression sorted and grouped according to the fold-change in expression (see stochastic resonance in [6]). Each group (net 25 groups) contains n = 685 RNAs. The x- and y-axes show the natural logarithm of the fold-change in each group average of RNA expression between the* s_i *and* s_j *states (*i^{th} *and* j^{th} *cell states;* ***Methods****) represented by* $ln(<\varepsilon(s_i)/\varepsilon(s_j)>)$ *and of RNA expression (*$ln(<\varepsilon(s_i)>)$*), respectively.*

Recently, we demonstrated that essentially the same critical-state dynamics that we observed for cell differentiation processes [4,6] are also present in overall RNA expression in single-cell mouse embryo development, which is particularly relevant to give ~~a~~ further proof of SOC control as a universal characteristic. Overall RNA expression and its dynamics exhibit typical features (genome avalanche and sandpile type criticality) of self-organized criticality (SOC) control in mouse embryo development [7]; **Figure 1** shows a genome avalanche, i.e., scaling-divergent behavior in a log-log-scale plot between expression and the temporal variance of expression (*nrmsf*), where the onset of divergent behavior occurs near the critical point (CP) of the sandpile critical behavior, and **Figure 2** shows that sandpile-type criticality (critical behavior) of the zygote state survives at the early 2-cell state and disappears after the middle 2-cell state to reach a stochastic pattern in the 4-cell state (linear pattern revealed in randomly shuffled overall expression: S2 Fig. in [4]).

Importantly, genome avalanches reveal that the regions of scaling and divergent regions in terms of *nrmsf* are opposite between embryo development and cell differentiation at a single-cell level, which is opposite the order of self-organization [7]; note that similar scaling-divergent order for cell differentiation is also observed at a cell-population level in cancer cells [4,6]. Furthermore, since chromosomes exhibit fractal organization, the power law behavior may reveal a quantitative relation between the aggregation state of chromatin through *nrmsf* and the average expression of an ensemble of genes, which is opposite in embryo development and cell differentiation.

On the other hand, sandpile-type criticality indicates that i) the memory of early embryogenesis in the zygote is lost after the middle 2-cell state: a significant change in the genome-state, i.e., the occurrence of reprogramming, and ii) the SOC control landscape

exhibits a critical transition state at the middle - late 2-cell states through a stochastic pattern, in which SOC control (sandpile-type criticality) disappears (see more in [4] and [7]).

In this report, we investigate the self-organizing dynamics of whole RNA expression in mouse embryo development [8] to unveil how genome reprogramming occurs by passing through a critical transition state via the dynamic interaction of critical states. This is elucidated by a dynamic expression flux analysis (quantitative evaluation of perturbation based on self-organization through SOC), and the results suggest that the critical gene ensemble of the critical point plays an essential role in reprogramming. Our results suggest that the SOC control mechanism of genome dynamics is rather universal among several distinct biological processes (refer to **Supplementary S1 file**). Thus, in the field of genome dynamics, our findings may provide a universal classification scheme for phenomena with far-from-equilibrium phase transitions, which has been missing in past studies [9].

Results

I. Dynamic Interaction of Critical States

Sloppiness of Mouse RNA Expression Dynamics: Coherent-Stochastic Behaviors in Critical States

As has been demonstrated in cell differentiation [4,6,10], mouse embryo overall RNA expression is also self-organized into distinct response domains (critical states). This can be seen as a transitional change in the bimodality of groups of expression according to *nrmsf* (**Figure 3A**), which further confirms SOC control of overall expression in mouse single-cell embryo development.

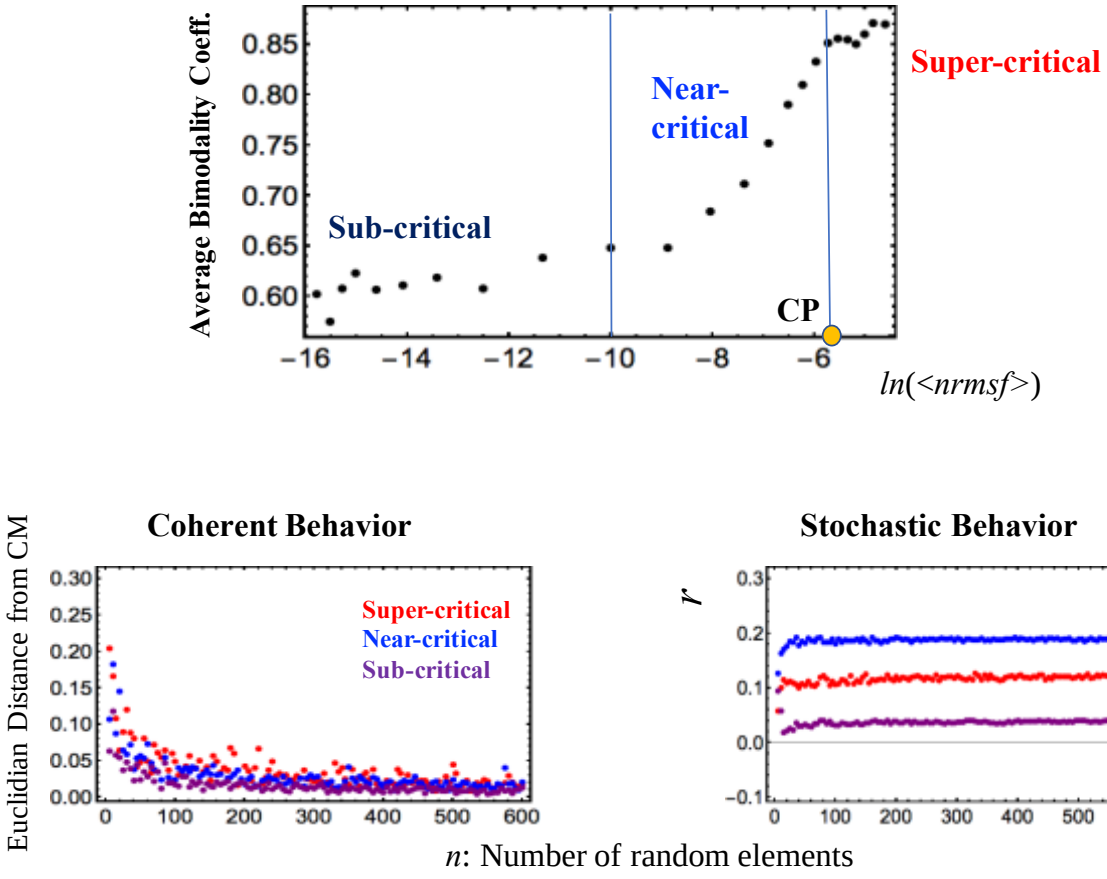

Figure 3: Critical states and their coherent stochastic behaviors:

A) Estimation of the bimodality coefficients (averaged over all cell states) according to nrmsf. Overall RNA expression is sorted and grouped according to the degree of nrmsf. The nrmsf grouping is made at a given sequence of discrete values of nrmsf (x_i: integer and integer + 0.5 from -15 to -4) with a fixed

*range (x_i - 0.5 < x_i < x_i + 0.5), and the corresponding average of the bimodality coefficient over the number of cell states is evaluated. The result clearly exhibits a transitional behavior to distinguish (averaged) critical states with respect to the value of ln<nrmsf>: sub-critical state (N= 4913 RNAs)< -8.8; -8.8 <near-critical state (N = 9180)< -5.4; -5.4 <super-critical state (N= 3408), where < > indicates the ensemble average of the nrmsf grouping. A critical point (CP: summit of sandpile criticality) exists around the edge between the near- and super-critical states, where the CP exists around ln(<nrmsf>) ~ -5.5 (**Figure 2**).*

*B) Coherent stochastic behaviors are revealed: stochastic expression within a critical state (right panel) is confirmed by the low Pearson correlation of ensembles of RNA expression between samples randomly selected from a critical state (averaged over the number of cell states with 200 repeats). The degree of stochasticity further supports the existence of distinct critical states. In the left panel, the Euclidian distance (averaged over 200 repeats and the number of cell states) of the center of mass (CM) of samples randomly selected from each critical state converges to the CM of the whole critical state (y = 0), which shows that the CM of a critical state (see the definition of CM in **section I**) represents its coherent dynamics that emerge in the ensemble of stochastic expression (refer to Figure 6 in [6]).*

As shown in **Figure 3B**, mouse embryo single-cell development exhibits clearly coherent-stochastic behavior (CSB: emergent coherent behavior in the ensemble of stochastic expression): i) stochastic expression within a critical state, in which a distinct degree of stochasticity within critical states further verifies the critical states, and ii) in the ensemble of stochastic expression within a critical state, coherent expression dynamics represented by the CM (center of mass) of the critical states emerges (e.g., Figure 7 in [10] and Figure 6 in [6]).

This emergent coherent behavior shows how a single cell can overcome the problem of stochastic fluctuation in local genes by gene expression regulation, which further affirms the statistical significance of averaging (mean-field) behaviors of genome avalanche and sandpile-type criticality (**Figures 1 and 2**). This macroscopic control in early embryo development points, notably, to the fact that bewildering epigenomic reprogramming molecular processes are regulated by a few hidden control parameters through SOC even at a single-cell level (for examples of control at the cell-population level, see [4]).

Expression Flux Dynamics Representing the Exchange of Genetic Activity

Here, we elucidate a statistical mechanism to explain how a single mouse embryo cell achieves genome reprogramming by passing through the transition state (**Figure 2**). **Figure 3B** reveals that the CSB in a critical state corresponds to the scalar dynamics of its CM, $X(s_j)$, where $X \in \{\text{Super}, \text{Near}, \text{Sub}\}$: Super, Near and Sub represent the corresponding critical states

at the j^{th} cell state, s_j. Thus, the change in the one-dimensional effective force acting on the CM determines the dynamics of $X(s_j)$. **Figure 4A** clearly confirms this point, in that the trend of the dynamics of the CM of a critical state follows its effective force (net self-flux dynamics: see the definition below). We now consider that the respective average values of the effective force can serve as baselines, and how perturbation from these baselines occurs dynamically.

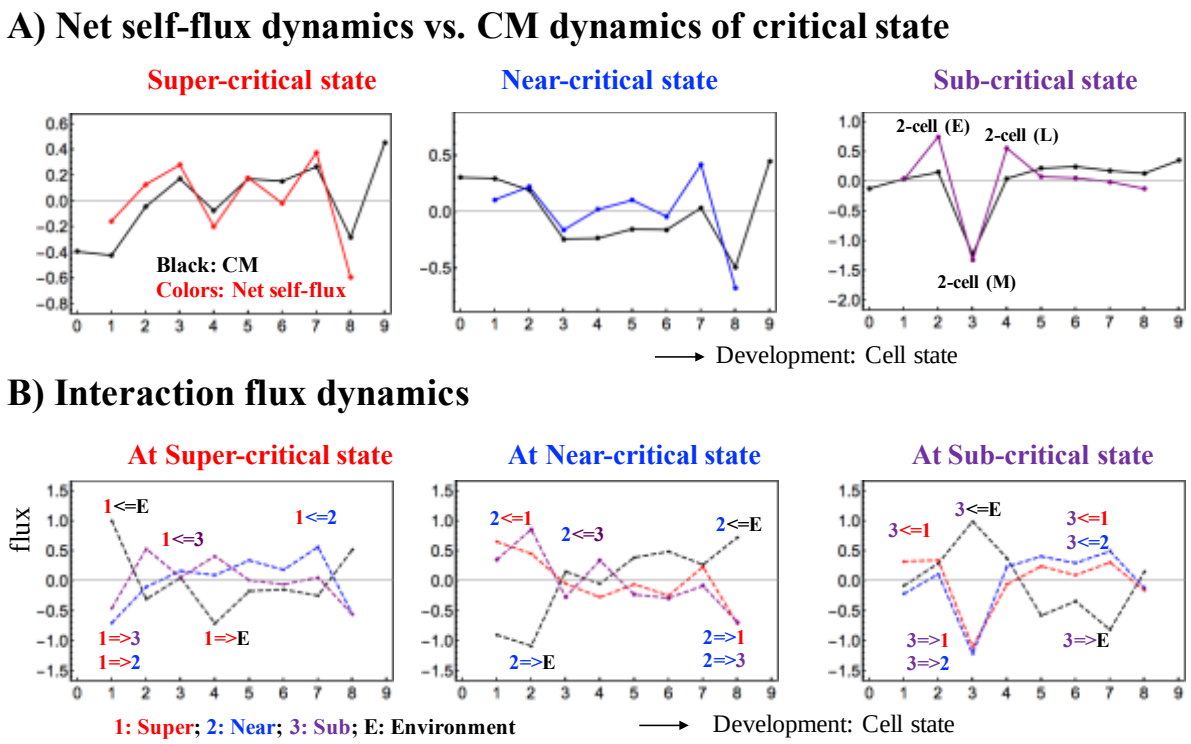

Figure 4: Flux dynamics among interacting critical states:

A) The plots show that the net self-flux dynamics (net effective force acting on CM) follow the dynamics of up- or down-regulated CM, where the sign of the net self-flux (i.e., IN and OUT) corresponds to activation (up-regulated flux) for positive responses and inactivation (down-regulated flux) for negative responses. The net self-fluxes of critical states (red: super-critical; blue: near-critical; purple: sub-critical) and the dynamics of the CM of critical states (black lines) are evaluated in terms of time averages. The x-axis represents the development of cell states and the y-axis represents net self-fluxes (black) and their CM dynamics (color).

*B) The plot shows flux dynamics through crosstalk with the environment. Interaction flux dynamics i<=j (or i=>j; color based on that of the j^{th} critical state) represent the interaction flux from the j^{th} critical state to the i^{th} critical state or vice versa. The results clearly show how global perturbations occur to guide genome reprogramming around the middle 2-cell state (3^{rd} point), based on the activity of the near-critical state and sub-critical state before and after reprogramming (see more in **Figure 7**).*

The genome is embedded into the intra-nuclear environment, where the expression flux represents the exchange of genetic energy or activity - the effective force produces work, and thus causes a change in the internal energy of critical states. This model shows a statistical thermodynamic picture of self-organized overall expression under environmental dynamic perturbations; the regulation of RNA expression is managed through the mutual interaction between critical states and the external connection with the cell nucleus milieu.

The effective force can be interpreted as a combination of incoming flux from the past to the present and outgoing flux from the present to the future cell state [4]:

$$f\left(X(s_j)\right) = \frac{\Delta P}{\Delta s} = \frac{1}{\Delta s}\left\{\frac{\left(X(s_j)-X(s_{j-1})\right)}{\Delta s_j} - \frac{\left(X(s_{j+1})-X(s_j)\right)}{\Delta s_{j+1}}\right\} - \langle f(X)\rangle$$
$$= (\text{IN flux} - \langle \text{IN flux}\rangle) - (\text{OUT flux} - \langle \text{OUT flux}\rangle), \qquad (2)$$

where ΔP is the change in momentum with a unit mass (i.e., the impulse: $F\Delta s = \Delta P$) and the development of the embryo cell-state is considered as the time-development with an equal time interval (a unit): $\Delta s_j = s_j - s_{j-1} = 1$ and $\Delta s = s_{j+1} - s_{j-1} = 2$; s_j corresponds to a specific cell state (such as zygote); the CM of a critical state is $X(s_j) = \frac{1}{N}\sum_{i=1}^{N} ln\left(\varepsilon_i(s_j)\right)$ with the natural log of the i^{th} expression $\varepsilon_i(s)$, $ln\left(\varepsilon_i(s_j)\right)$ at the j^{th} cell state, $s = s_j$ (N = the number of RNAs; **Methods**); the average of net self-flux over the number of critical states, <f(X)> = <INflux> - <OUTflux>.

The effective force, $f(X(s_j))$, is called the *net self-flux of a critical state* at the j^{th} cell state s_j. The net self-flux, IN flux - OUT flux, has a positive sign for incoming force (net IN self-flux) and a negative sign for outgoing force (net OUT self-flux). Thus, the CM from its average over all cell states represents up- (down-) regulated expression for the corresponding net IN (OUT) flux.

The *interaction flux* of a critical state $X(s_j)$ with respect to another critical state or the environment (milieu) Y (**Figure 4B**) can be defined as:

$$f\left(X(s_j);Y\right) = \frac{1}{\Delta s}\left\{\frac{\left(X(s_j)-Y(s_{j-1})\right)}{\Delta s_j} - \frac{\left(Y(s_{j+1})-X(s_j)\right)}{\Delta s_{j+1}}\right\} - \langle f(X;Y)\rangle, \qquad (3)$$

where, again, the first and second terms represent IN flux and OUT flux, respectively, and the net value, IN flux- OUT flux, represents incoming (IN) interaction flux from Y for a positive sign and outgoing (OUT) interaction flux to Y for a negative sign. $Y \in \{\text{Super}, \text{Near}, \text{Sub}, \text{E}\}$ where $Y \neq X$: E represents the environment.

Due to the law of force, the net self-flux of a critical state is the sum of the interaction fluxes with other critical states and the environment:

$$f\left(X(s_j)\right) = \sum_{i=1}^{M=2} f(X(s_j); A_i) + f(X(s_j); E), \quad (4)$$

where $A_i \in \{\text{Super}, \text{Near}, \text{Sub}\}$ with $A_i \neq X$, and M is the number of internal interactions (M = 2), i.e., for a given critical state, there are two internal interactions with other critical states. Equation (4) tells us that the sign of the difference between the net self-flux and the overall contribution from internal critical states, $f\left(X(s_j)\right) - \sum_{i=1}^{M=2} f(X(s); A_i)$, reveals incoming flux (positive) from the environment to a critical state or outgoing flux (negative) from a critical state to the environment.

Regarding average flux, an *average flux balance* exists in terms of the net average fluxes coming in and going out at each critical state (near-zero) through the environment: average expression fluxes: $\langle f(X) \rangle \approx 0$ and $\langle f(X;Y) \rangle \approx 0$ (Equations (2) and (3)) for the mouse RNA-Seq data (**Methods**). Note: the balance does not hold at each time point/cell state.

This model of expression flux dynamics shows environmental dynamic perturbations in the self-organization of overall expression; the regulation of RNA expression is managed through the mutual interaction between critical states through the cell nucleus. Based on this concept, it becomes possible to evaluate the temporal change in the open-thermodynamic genetic system, where the expression flux represents the exchange of genetic energy or activity.

Sub-Critical State as a Generator of Perturbation in Genome-Wide Self-Organization

Average net IN and OUT flux flows show how the internal critical states and the external cell nucleus milieu mutually interact (**Figure 5)**. Two cyclic expression flux flows form among critical states: between super- and near-critical states, and *a dominant cyclic expression flux between super- and sub-critical states* in the genomic system.

These cyclic fluxes are considered in light of self-regulatory gene expression by means of a complex epigenetic machinery (methylation processes, long non-coding RNAs, miRNAs, etc): the super-critical state (high variance RNAs) acts as a sink internally to receive genetic information and send it back to the other critical states through the cell environment. On the other hand, the sub-critical state (low-variance RNAs) acts as an internal source of information and sustains (like a generator in an electrical circuit) the interaction with cyclic fluxes.

This implies that the collective behavior of an ensemble of low-variance RNA expression (sub-critical state) plays an essential role in reprogramming in single cells (see a

more dynamical discussion below). This dominant cyclic flow also shows that the dynamics of the sub-critical and super-critical states are anti-phase with respect to each other (**Figure 6A**) to form a strong coupling between them [6].

The above model (formation of a dominant cyclic flux between a source and sink) provides a genome-engine metaphor for SOC control mechanisms [4], suggesting that the genome engine may be a **universal mechanism** in the genome system of mammalian cells.

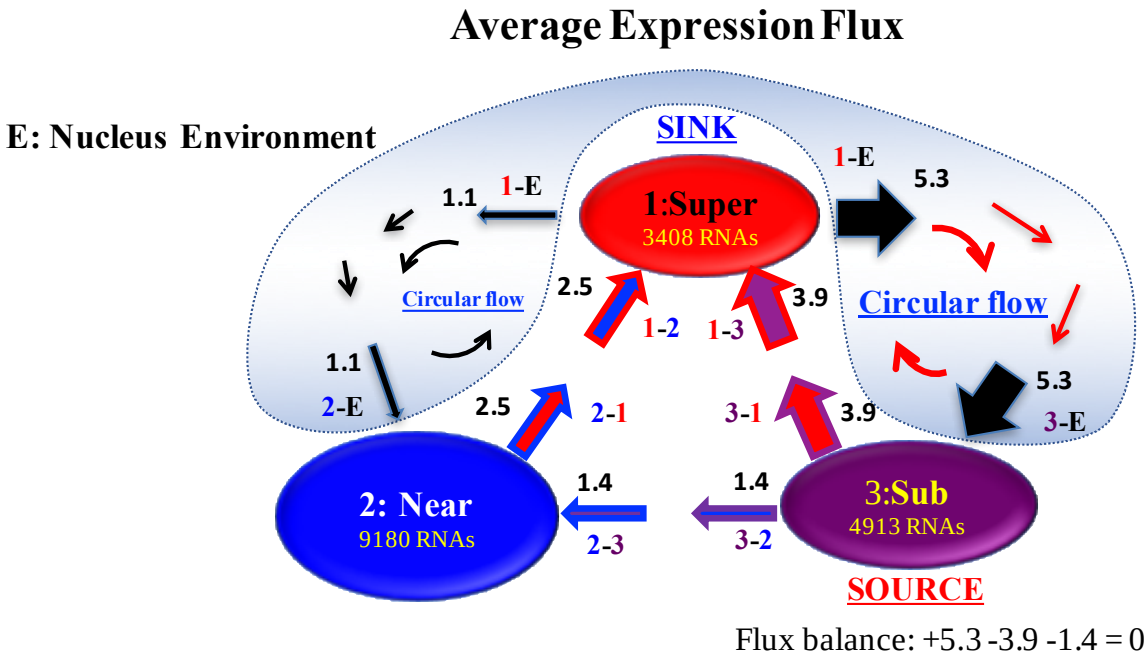

Figure 5: Scheme of SOC-control through the representation of average expression flux in mouse embryo development: *Average expression flux network in the mouse embryo genome system reveals that a sub-critical state acts as an internal 'source', where IN flux from the environment (shaded blue) is distributed to other critical states. In contrast, a super-critical state acts as an internal 'sink' that receives IN fluxes from other critical states, and the same amount of expression flux is sent to the environment, due to the average flux balance (an example is shown at the sub-critical state). Two cyclic fluxes (Near-Super and Sub-Super) through the environment are seen. Sub-Super cyclic flux forms a dominant flux flow, which generates strong coupling between the super- and sub-critical states accompanied by their anti-phase expression dynamics [6], makes its change oscillatory feedback, and thus sustains autonomous SOC control of overall gene expression. This formation of a dominant cyclic flux with a source and sink provides a universal genome-engine metaphor of SOC control mechanisms, as in terminal cell fates (refer to the Discussion in [4]). Numerics represent average net flux values, where the average net interaction flux of the super-critical state to the environment (black arrow) is decomposed into two parts for the flux-balance at other critical states.*

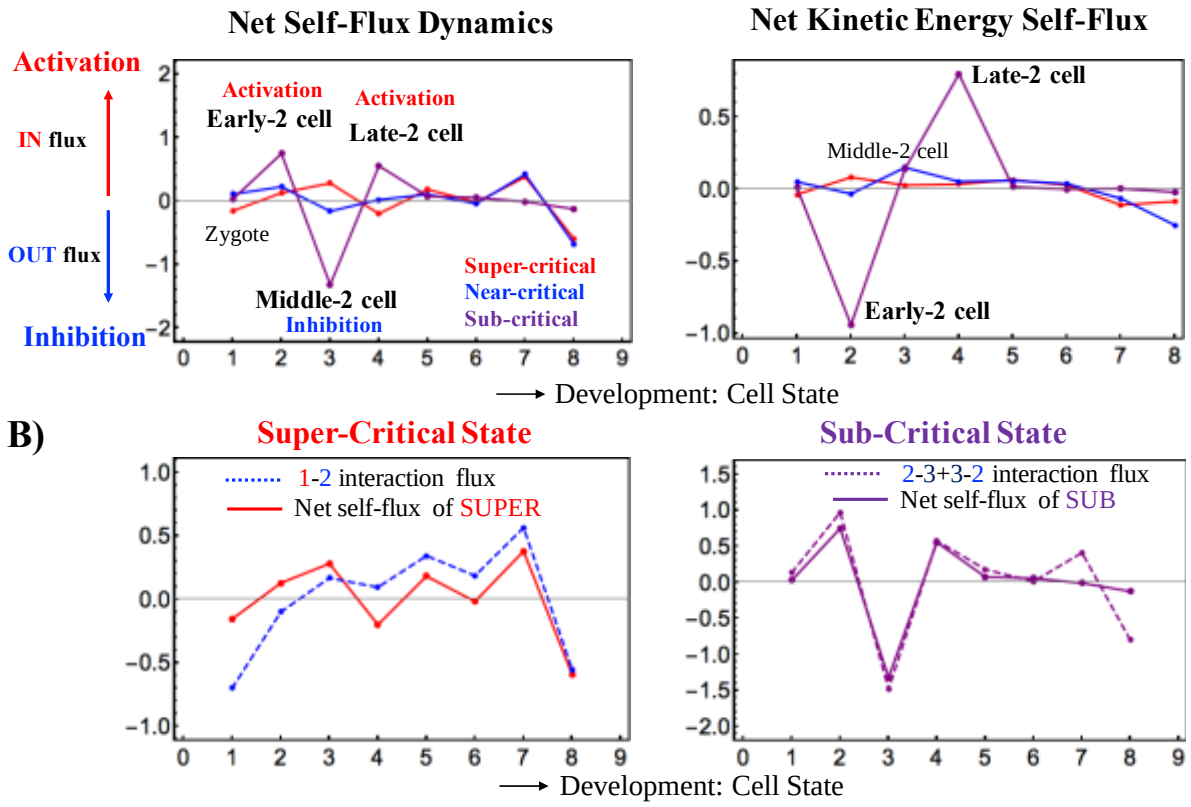

Figure 6: Sequential global perturbations revealed through net- self-flux and kinetic-energy self-flux, and critical-state expression dynamics determined by the interaction flux:

A) Net self-flux dynamics (left panel: dynamic effective force acting on the CM) show that sequential global perturbations occur at the early - middle 2-cell states (activation → inhibition) and at the middle- late 2-cell states (inhibition → activation), which is supported by the change in the net kinetic energy flux (right panel) from OUT to IN flux (see more in ***section II****).*

B) The interaction dynamics reveal that the expression dynamics of the sub-critical state (the generator of perturbation) are determined by the net interaction flux between the near- and sub-critical states, and the dynamics of the super-critical state are determined by its interaction flux from the near-critical state (super to near interaction). This image shows how the critical dynamics (temporal change in criticality exists at the boundary of the near- and sub-critical states) affect the entire genome expression system, since the sub-critical state generates autonomous SOC control of overall expression.

II. SOC Control Mechanism of Genome Reprogramming Through a Critical Transition State

The erasure of initial-state criticality (e.g., in the zygote state) points to the onset of genome reprogramming after the middle 2-cell state. As noted, an initial state can be the early 2-cell state instead of the zygote state; this independent choice of the initial state further confirms the timing of the genome-state change [4].

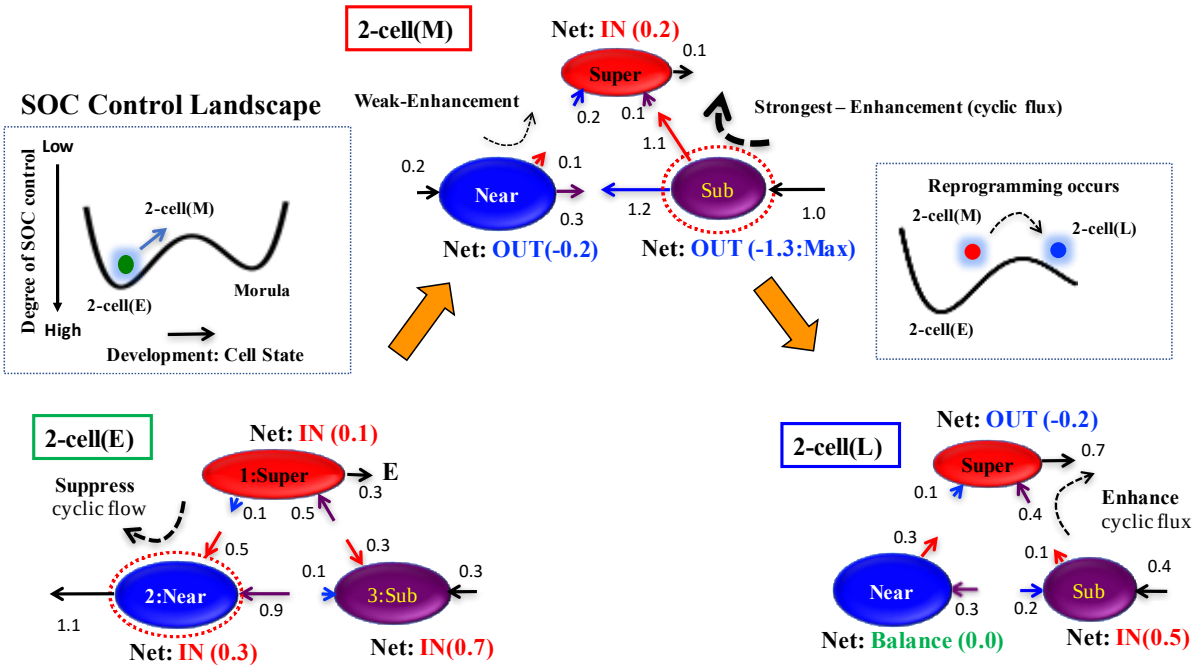

Figure 7: Time-development of SOC control of genome reprogramming revealed through expression flux dynamics: *Reprogramming occurs at the middle - late 2-cell states through the transition state of the SOC control landscape (**Figure 2B;** more in [4]; **Supplementary S1 file**), which is supported by sequential global perturbations at the early - late 2-cell states (**Figure 6**). Interaction flux dynamics (**Figure 4B**) reveal a SOC control mechanism of genome reprogramming:*

i) At the early 2-cell state, interaction flux between Super and Near suppresses cyclic flux. All critical states receive the net IN flux, which indicates that, inside the nucleus, the mouse genome system is activated through the cell milieu (environment), which makes the cell state move up from a valley on the SOC control landscape, as shown in the inset,

*ii) At the middle 2-cell state, a substantial change in the net interaction flux at critical states occurs from the early 2-cell state to induce a major change in two cyclic fluxes: a change from suppression to the enhancement of cyclic flux (Super - Near), and the strongest enhancement of cyclic flux (Super - Sub). This leads to a reverse change in the net interaction fluxes at the near-critical state (IN to OUT and vice versa) to enhance the cyclic flux (Super - Near) from major suppression in the early 2-cell state. This activation of the sub-critical state (the maximum net OUT flux sending to other critical states) enables the system to pass through the transition state (erasure of zygote criticality: **Figure 2**) to reprogram the mouse embryo genomic state as described in the inset, and*

iii) In the late 2-cell state, while the enhancement of cyclic flux between super- and sub-critical states becomes weak, the perturbation of average flux activity almost disappears. While the near-critical state becomes balanced, the sub-critical state receives the net expression flux and the super-critical state sends the net flux.

i)-iii) show the occurrence of sequential global perturbations to pass through the transition state to achieve single-cell reprogramming. Numerics represent net interaction flux values.

Interaction flux dynamics (**Figures 4B**, **7**) describes the erasure as a thermodynamical event that passes through a critical transition state, which shows how the genome system can pass through the critical transition state - dynamic perturbation in the average flux in terms of the enhancement-suppression of two cyclic flows around the reprogramming event:

i) In the early 2-cell state, interaction flux (Super - Near) suppresses cyclic flux, where the near-critical state acts as a major suppressor (**Figure 7**). All critical states receive the net IN flux, which indicates that inside the nucleus, the mouse genome system is activated through the cell milieu (environment).

ii) A substantial change in the net interaction flux at critical states occurs beginning in the early 2-cell state and into the middle 2-cell state. These changes induce a major change in the two cyclic fluxes: a change from suppression to the enhancement of cyclic flux (Super - Near), and the strongest enhancement of cyclic flux (Super - Sub).

This leads to the reverse change in the net interaction fluxes at the near-critical state (IN to OUT and vice versa) to enhance the cyclic flux (Super - Near) from the major suppressor at the early 2-cell state. Thus, these cyclic enhancements make the genome system pass through the critical transition state. This interaction model reveals that

a) A major biological event in genome reprogramming through a stochastic pattern is guided by the sub-critical state at the middle 2-cell state, and

b) A detailed open thermodynamic mechanism regarding the erasure of criticality of the early 2-cell state in the middle 2-cell state.

iii) In the late 2-cell state, while the enhancement of cyclic flux (Super - Sub) becomes weak, the perturbation of average flux activity almost disappears due to passage through the transition state.

Therefore, two major global perturbations, which involve the activation-inhibition of multiple critical states, occur between the early and middle 2-cell states, and between the middle and late 2-cell states during genome reprogramming. This global perturbation event is clearly seen in the net kinetic energy flux [4] in a critical state (**Figure 6A**):

$$K\left(X(s_j)\right) = \frac{1}{2}\left\{\left(\frac{\left(X(s_j)-X(s_{j-1})\right)}{\Delta s_j}\right)^2 - \left(\frac{\left(X(s_{j+1})-X(s_j)\right)}{\Delta s_{j+1}}\right)^2\right\} - \langle K(X)\rangle, \qquad (5)$$

where the kinetic energy of the CM for the critical state with unit mass at $s = s_j$ is defined as $1/2 \cdot v(s_j)^2$ with average velocity: $v(s_j) \equiv \frac{X(s_j)-X(s_{j-1})}{\Delta s_j}$.

The perturbation of self-organization shows a temporal change in the degree of SOC control: a high (low) degree of SOC control points to a weak-local (strong-global) perturbation of the average flux flow with a high (low) degree of SOC control. As described in the above, the interaction dynamics among critical states (insets in **Figure 7**) show that reprogramming occurs by climbing the hill of the SOC landscape. This process encompasses the following steps (**Figure 7**): 1). Release from a high degree of SOC control in the early 2-cell state due to the activation of the near-critical state, 2). Passage through the transition state - the lowest degree of SOC control (i.e., non-SOC control) due to the strongest global perturbation at the middle 2-cell state stemmed from the activation of the sub-critical state, which makes SOC control from a high degree to non-control, 3). Return to SOC control in the late 2-cell state.

Figure 6B shows that the expression dynamics of the sub-critical state (the generator of perturbation) are determined by the net interaction flux between near- and sub-critical states, and the dynamics of the super-critical state are determined by its interaction flux from the near-critical state (super to near interaction). Biologically, this between-states interaction serves as the underlying mechanism of self-regulatory gene expression through the orchestrated cooperation of myriads of epigenetic modifications, transcriptional factors and non-coding RNA regulations to determine the critical-state coherent oscillatory behavior. The two cyclic fluxes act as feedback flow for the change in criticality to generate coherent oscillatory dynamics of critical states. The above dynamic model provides a mechanism for long-term global RNA oscillation underlying autonomous SOC control generated by the sub-critical state [4,6,12].

Discussion and Conclusion

Notwithstanding that, view point with current thermodynamics allows for incredibly precise predictions thanks to the ‘generality of its premises’ (Einstein’s words [13]), this allowed us to grasp the essentials by skipping the (largely unknown) detailed biological mechanisms. The erasure of epigenetic marks (refer to the Discussion in [4]) is consistent with thermodynamics perspective beside its actual mechanism.

The time-development of sandpile criticality reveals the existence of a transition state which in turn suggests that programming of mouse embryo occurs through the transition state. As a proof of this concept, it was demonstrated that EGF-stimulated MCF-7 cells do not erase sandpile criticality (see Figure 5A in [4]), i.e., no genome-state change (consistent with the experiment no cell-differentiation occurs [14]). This shows that the event of single-cell

programming is related to overcoming of the transition state, which is a typical event in thermodynamic reaction mechanism. Moreover, the timing of the reprogramming does not depend on the selection of an initial cell state (see **Supplementary S1 file**). The result helps us to obtain a quantitative appreciation of the still largely qualitative notion of the epigenetic landscape.

Intriguingly, our statistical thermodynamics approach reveals that the collective behavior of an ensemble of low-variance RNA expression (sub-critical state), which shows only marginal changes in expression and consequently are considered to be devoid of any interest, guides the genome to pass through the transition state (erasure of initial-state sandpile-type criticality) to reprogram the mouse embryo. The sub-critical state gene ensembles act as a driving force to transmit their potentiality, or energy of coherent transcription fluctuations, to high-variance genes (the genome-engine mechanism [4]). We observed also this generator role of the sub-critical state in cell differentiation [4] (MCF-7 human cancer cells and HL-60 human promyelocytic leukemia cells). Thus, genome-engine mechanism may provide a universal SOC control mechanism in the genome system.

Sandpile-type critical point (CP) exists around the edge between the near- and super-critical states in the mouse genome. The erasure of sandpile-type criticality through embryo development induces dynamic change in interaction flux dynamics between near- and sub-critical states, which determines the activation-inhibition dynamics of the sub-critical state. This suggests that a critical gene ensemble of sandpile-type criticality should exist to affect the entire genome expression. Therefore, elucidation of the molecular mechanism that guides the genome system through a transition state is expected to unveil molecular clues as to how a single cell can succeed or fail at reprogramming, such as in an iPS single cell.

The establishment of a material basis for the observed phenomenology by unveiling the molecular mechanism on the criticality of gene ensemble is awaited as the next study, which may lead us toward a comprehensive understanding of single-cell reprogramming.

Methods

Biological Data Sets:

We analyzed mammalian RNA-Seq data:

i) Early embryonic development in human and mouse developmental stages in RPKM (Reads Per Kilobase Mapped) values; GEO ID: GSE36552 (human: N = 20286 RNAs) and GEO ID:

GSE45719 (mouse: N = 22957 RNAs), which have 7 and 10 embryonic developmental stages (experimental details in [8] and [15], respectively):

Human: oocyte (m=3), zygote (m=3), 2-cell (m=6), 4-cell (m=12), 8-cell (m=20), morula (m=16) and blastocyst (m=30),

Mouse: zygote (m=4), early 2-cell (m=8), middle 2-cell (m=12), late 2-cell (m=10), 4-cell (m=14), 8-cell (m=28), morula (m=50), early blastocyst (m=43), middle blastocyst (m=60) and late blastocyst (m=30), where m is the total number of single cells.

ii) T helper 17 cell differentiation from mouse naive CD4+ T cells in RPKM values, where Th17 cells are cultured with anti-IL4, anti-IFNγ, IL-6 and TGF-β, and Th0 cultures provide control cells that receive TCR activation in the absence of exogenous polarizing cytokines (IL-6 + TGF-β) (details in [11]); GEO ID: GSE40918 (mouse: N = 22281 RNAs), which has 9 time points: t_0 = 0, t_1 = 1,3,6,9,12,16,24, $t_{T=8}$ = 48h for Th17, and 6 time points: t_0 = 0, t_1 = 1,3,6,16, $t_{T=5}$ = 48h for Th0.

RNAs that had RPKM values of 0 over all of the cell states were excluded. Random real numbers in the interval [0-1] generated from a uniform distribution were added to all expression values for the natural logarithm in the analysis of sandpile criticality (**Figure 2**; no addition of random numbers in the rest of the Figures). This procedure avoids the divergence of zero values in the logarithm. The robust sandpile-type criticality through the grouping of expression was checked by multiplying the random number by a positive constant, a (a< 10), and we set a = 0.01. Note: The addition of large random noise (a>>10) destroys the sandpile CP.

References

1. Reik W.; Dean W.; Walter J. Epigenetic reprogramming in mammalian reprogramming. *Science* **2001**, 10, 1089-1093.
2. Santos, F.; Dean, W. Epigenetic reprogramming during early development in mammals. *Reproduction* **2004**, 127, 643-651.
3. Cantone, I.; Fisher, A.G. Epigenetic programming and reprogramming during development. *Nat. Struct. Mol. Biol.* **2013**, 20, 282-289.
4. Tsuchiya, M. Giuliani, A.; Hashimoto, M.; Erenpreisa, J.; Yoshikawa, K. Self-organizing global gene expression regulated through criticality: mechanism of the cell-fate change. *PLoS One* **2016**, 11, e0167912

5. Wennekamp, S.; Mesecke, S.; Nédélec, F.; Hiiragi, T. self-organization framework for symmetry breaking in the mammalian embryo. *Nat. Rev. Mol. Cell Biol.* **2013**, 14, 452-459.
6. Tsuchiya, M.; Giuliani, A.; Hashimoto, M.; Erenpreisa, J.; Yoshikawa, K. Emergent Self-Organized Criticality in gene expression dynamics: Temporal development of global phase transition revealed in a cancer cell line. *PLoS One* **2015**, 11, e0128565.
7. Giuliani, A.; Tsuchiya, M.; Yoshikawa, K. Single-cell genome dynamics in early embryo development: a statistical thermodynamics approach, *bioRxiv* **2017**. doi: https://doi.org/10.1101/123554
8. Deng, Q.; Ramsköld, D.; Reinius, B.; Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **2014**, 343, 193–196.
9. Marković, D.; Gros, C. Power laws and self-organized criticality in theory and nature. *Phys. Rep.* **2014**, 536, 41-74.
10. Tsuchiya, M.; Hashimoto, M.; Takenaka, Y.; Motoike, I.N.; Yoshikawa, K. Global genetic response in a cancer cell: Self-organized coherent expression dynamics. *PLOS One* **2014**, 9: e97411.
11. Ciofani, M.; Madar, A.; Galan, C.; Sellars, M.; Mace, K.; Pauli, F.; Agarwal, A.; Huang, W.; Parkhurst, C.N.; Muratet, M.; et al. A validated regulatory network for Th17 cell specification. *Cell* **2012**, **151**, 289-303.
12. Tsuchiya, M., Wong, S.T.; Yeo, Z.X.; Colosimo, A.; Palumbo, M.C.; Farina, L.; Crescenzi, M.; Mazzola, A.; Negri, R.; Bianchi, M.M.; Selvarajoo, K.; Tomita, M.; Giuliani, A. Gene expression waves: cell cycle independent collective dynamics in cultured cells. *FEBS J.* **2007**, 274, 2874–2886.
13. Einstein, A.; Schilpp, P.A. Autobiographical notes. Open Court Publishing Company, 1949.
14. Nagashima, T.; Shimodaira, H.; Ide, K.; Nakakuki, T.; Tani, Y.; Takahashi, K.; Yumoto, N.; Hatakeyama, M. Quantitative transcriptional control of ErbB receptor signaling undergoes graded to biphasic response for cell differentiation. *J. Biol. Chem.* **2007**, 282: 4045–4056.
15. Yan, L.; Yang, M.; Guo, H.; Yang, L.; Wu, J.; Li, R.; Liu, P.; Lian, Y.; Zheng, X.; Yan, J.; et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **2013**, 20, 1131-1139.

Acknowledgements:

MT sincerely thanks the Institute for Advanced Biosciences, Keio University, Tsuruoka City, the Yamagata prefectural government, Japan, and Mr. Fumiaki Kikuchi for allowing him to complete this research project at Keio University. The authors are thankful to Drs. Jekaterina Erenpreisa and Mesut Tez for fruitful and critical discussions.

Supplementary S1 file:

Summary of Our Previous Results

SOC control of genome expression

We have investigated the dynamics of collective gene behavior in several biological processes associated with changes in the cell fate from early embryo development in human and mouse (single cell), and terminal cell fates (helper T17 cell (single cell); human leukemia HL-60 cells; human breast cancer MCF-7 cells) at the single-cell and cell-population levels.

Regarding the mechanism of the cell-fate change through the self-organization of whole-genome expression (mRNA: microarray data and RNA: RNA-Seq data), we have demonstrated that, in all of the models analyzed, a self-organized critical transition (SOC) in whole-genome expression [Tsuchiya, M. et al., 2015 and 2016; Giuliani, A. et al., 2017] plays an essential role in the change in the genome state at both the population and single-cell levels: **1) Self-Organization Through Criticality**: In the cell-fate change at the terminal phase (determination of differentiation for the cell population), critical dynamics self-organizes whole-genome expression into a few distinct response expression domains (critical states). Two distinguished critical behaviors, sandpile-type transitional behavior and scaling-divergent behavior (genomic avalanche), are observed when overall expressions are grouped according to the fold-change in expression and the temporal variance of expression, *nrmsf* (normalized root mean square fluctuation), respectively (see details in Methods [Tsuchiya, M. et al., 2016]). Sandpile critical behaviors (criticality) are evident in the fold-change in expression for both human and mouse embryo development. A physical explanation for such critical dynamics in the fold-change is that amplification in a critical state stemmed from a stochastic resonance effect in the change in the ensemble of stochastic expression [Tsuchiya, M. et al., 2015]. Sandpile-type criticality shows diverging behaviors of up- and down-regulation at a critical point (CP), whereas in a genomic avalanche order (scaling) and disorder (divergence) are balanced at the CP (similar to the characteristics at the edge of chaos [Langton, 1990; Waldrop, 1992; de Oliveira, 1992]). Importantly, SOC does not correspond to a phase transition in overall expression from one critical state to another. Instead, it represents self-organization of the coexisting critical states through a critical transition, i.e., SOC consolidates critical states into a genome expression system. In each critical state, coherent (collective/coordinated) behavior emerges in ensembles of stochastic expression by more than 50 elements (coherent-

stochastic behavior). The characteristics of the self-organization through SOC become apparent only in the collective behaviors of groups with an average of more than 50 genes (mean-field approach).

2) **Timing of the Genome-State Change**: At both the population and single-cell levels, the cell-fate change in the genome is determined through the erasure of initial-state criticality (as a boundary condition of SOC control):

a) **In terminal cell fates**, the cell-fate change (commitment to cell differentiation) occurs at the end of a dissipative global perturbation in self-organization through the erasure of an initial-state criticality. As a proof of this idea, the EGR-stimulation of MCF-7 cells (cell population), where the cell fates do not change over time (cell proliferation and no differentiation), does not exhibit a transitional change (i.e., no erasure of criticality; see Figure 5 in Tsuchiya, M. et al., 2016) in critical dynamics. The independence of the choice of the initial state at $t = t_0$ for the breakdown of sandpile-type criticality at $t = t_b$ (condition: $t_0 < t_b$) further confirms the timing of the genome-state change. The time point (t_0) of the initial state is earlier or equal to the onset of the pulse-like global perturbation; after the global perturbation, this independence does not hold, which suggests that a pulse-like global perturbation may be related to the first stage of cell-fate determination (process for autonomous terminal differentiation). The result suggests that the terminal cell fate is determined through two stages: the first stage by means of a pulse-like global perturbation, and the commitment stage through the erasure of initial-state criticality. Quantitative evaluation of global perturbation reveals that dynamic interactions between critical states determine the critical-state coherent regulation. The occurrence of a temporal change in criticality perturbs this between-states interaction, which directly affects the entire genomic system, and thus suggests that the erasure of criticality is associated with a loss of the initial SOC control mechanism for the dynamical change in the entire genomic system (i.e., pruning procedure for the regulation of global gene expression at the initial state).

b) The reprogramming of early human and mouse embryo cells destroys the zygote SOC control to initiate self-organization in the new embryonal genome, which passes through a stochastic overall expression pattern. This timing of reprogramming was further confirmed by the facts that i) the Pearson correlation between overall RNA expression (RNA-Seq data) in the zygote and developed cell states revealed a critical transition between the middle and late 2-cell states; the low correlation showed the loss of zygote embryogenesis after the middle 2-cell state, and ii) the choice of the initial

state is independent of the breakdown of sandpile-type criticality; the choice of the initial state before the middle 2-cell state does not change the erasure event of criticality. The breakdown of early SOC zygote control in overall expression indicates that significant global perturbation occurs to destroy SOC zygote control in early embryo development.

The results suggest the existence of specific molecular-physical routes for the erasure of critical dynamics for the cell-fate decision.

3) **SOC-Control Landscape**: The development of SOC control in early mouse embryo development exhibits a transition from a sandpile-type criticality to another one through a stochastic pattern, i.e., an SOC-control landscape: a valley (SOC control: the zygote single-cell stage to the early 2-cell state) - ridge (non-SOC control: the middle to the late 2-cell state) - valley (SOC control: the 8-cell state to the morula state). This should provide a qualitative image of the epigenetic landscape framed in broad terms of the global activation-deactivation dynamics of the genome that is generally consistent with the DNA de-methylation-methylation landscape [Guo, H. *et al*., 2014].

4) **Existence of a Critical Transition of Collective Expression Dynamics in the Ensemble of Cells**: The onset of the genome-state change (at the breakdown of initial-state SOC control) exhibits a clear difference between single cells and a cell population: cell populations do not exhibit a stochastic pattern, in contrast to single cells in early human and mouse embryonic development. The stochastic pattern is confirmed by a low Pearson correlation between the zygote and early embryo single-cell states at the onset, whereas in cell populations, the Pearson correlation for overall expression at different time points is close to unity. This suggests that there is a transition from single-cell stochastic to highly correlated cell-population behavior at the genome-state change in overall expression as the emergent layer of a relevant collective regulation behaviors starting from a given threshold number of cells. The near-stochastic pattern of helper Tell 17 cell differentiation at the onset (single cell) supports such a transition.

5) **Long-term global mRNA oscillation underlies SOC control**: The sub-critical state (ensemble of low-variance gene expressions) sustains critical dynamics in SOC control of terminal cell fates, where a sub-critical state forms a robust cyclic state-flux with a super-critical state (ensemble of high-variance gene expressions) through the cell nuclear environment. The results show that there is no fine-tuning of an external driving parameter to maintain critical dynamics in SOC control and therefore, the oscillatory expression dynamics

of sub-critical states generates a long-term global mRNA oscillation [Tsuchiya, M. et al., 2007] to sustain the self-control of SOC.

References:

1. de Oliveira, P. M. Why do evolutionary systems stick to the edge of chaos? *Theo. Biosci.* **120**, 1-19 (2001).
2. Giuliani, A. *et al*. Single-cell genome dynamics in early embryo development: a statistical thermodynamics approach, *bioRxiv* (2017). doi: https://doi.org/10.1101/123554
3. Guo, H. *et al*. The DNA methylation landscape of early human embryos. *Nature* **511**, 606-610 (2014).
4. Langton, C. G. Computation at the edge of chaos - phase transitions and emergent computation. *Physica D* **42**, 12–37 (1990).
5. Tsuchiya, M. *et al*. Gene expression waves: cell cycle independent collective dynamics in cultured cells. *FEBS J* **274**, 2874–2886 (2007).
6. Tsuchiya, M. *et al*. Emergent Self-Organized Criticality in gene expression dynamics: Temporal development of global phase transition revealed in a cancer cell line. *PLoS One* **11**: e0128565 (2015).
7. Tsuchiya, M. *et al*. Self-organizing global gene expression regulated through criticality: mechanism of the cell-fate change. *PLoS One* **11**: e0167912 (2016).
8. Waldrop, M. M. Complexity: The Emerging Science at the Edge of Chaos. Simon and Schuster, New York (1992).