Self-Organizing Global Gene Expression Regulated Through Criticality: The Mechanism of Cell Fate Change

Masa Tsuchiya[1*], Alessandro Giuliani[2], Midori Hashimoto[3] Jekaterina Erenpreisa[4] and Kenichi Yoshikawa[5]

[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]Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Japan

[4]Latvian Biomedical Research & Study Centre, Riga, Latvia

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

*Corresponding author: tsuchiya.masa@gmail.com

Key words: Cell fate; Self-organized criticality; Genome-wide expression; Global & local perturbations; Flux dynamics; Embryo development; Low-variance genes; Collective behavior.

Abstract

Background

A fundamental issue in bioscience is to understand the underlying mechanism of the dynamic control of genome-wide expression through the complex temporal-spatial self-organization of the genome regulating cell fate change. We address this issue by elucidating a physically motivated self-organizing mechanism.

Principal Findings

Building upon transcriptome experimental data for seven distinct cell fates, including early embryonic development, we demonstrate that self-organized criticality (SOC) plays an essential role in the dynamic control of global gene expression regulation at both population and single cell levels. The novel findings are:

i) **Mechanism of cell fate changes**: A sandpile-type critical transition self-organizes overall expression into a few transcription response domains (critical states). Cell fate change occurs by means of a dissipative pulse-like global perturbation in self-organization through the erasure of an initial-state critical behaviors (criticality). Most notably, reprograming of the early embryo cells destroys the zygote SOC control to initiate self-organization in the new embryonal genome, which passes through a stochastic overall expression pattern.

ii) **Perturbation mechanism of SOC controls**: Global perturbations of the SOC controls involve the temporal regulation of critical states. Elucidation of the dynamic interaction of critical states in terminal cell fates reveals that sub-critical states (ensembles of genes for which expression undergoes only very limited changes during the process) act as a 'source' to sustain global perturbations, whereas super-critical states (ensembles of genes for which expression varies greatly) behave as a 'sink' to form a dominant cyclic state-flux with sub-critical states through the cell nuclear environment.

Conclusion and Significance

The 'whole-genome' level of gene expression regulation, where the collective behavior of low-variance genes plays a central role in genome-wide self-organization, complements the microscopic gene-by-gene fine tuning, and represents a still largely unexplored thermodynamically regulated mechanism responsible for massive genome expression reprogramming.

Introduction

In mammalian mature stem cells, the cell fate/state can be reprogrammed to provoke a shift between two stable (and very different) gene expression profiles through a pluripotent state that involves tens of thousands genes by means of a few reprogramming stimuli [1-5].

The coordinated control of a large number of genes must overcome several difficulties, such as the substantial instability of genetic products due to the stochastic noise stemming from the low copy number of specific gene mRNAs per cell and the lack of a sufficient number of molecules to reach a thermodynamic limit [6,7]. Due to the complexity of the interaction between molecular effectors and chromatin structural changes, it has been a challenging issue to understand how globally coordinated control can determine the cell fate/state from a genomic point of view. In this respect, it is important to gain a comprehensive understanding of dynamic control mechanisms that could help us to obtain a quantitative appreciation of the still largely qualitative notion of the epigenetic landscape [8]. The existence of global gene regulation implies that a driving force of the genomic expression is acting through a few number of control parameters underpinning highly complex molecular genetics reaction mechanisms. The hypothesis of the possibility to get a reliable model of a complex system by the use of few relevant parameters was aptly commented under the heading of 'sloppiness' by Transtrum et al [9]:

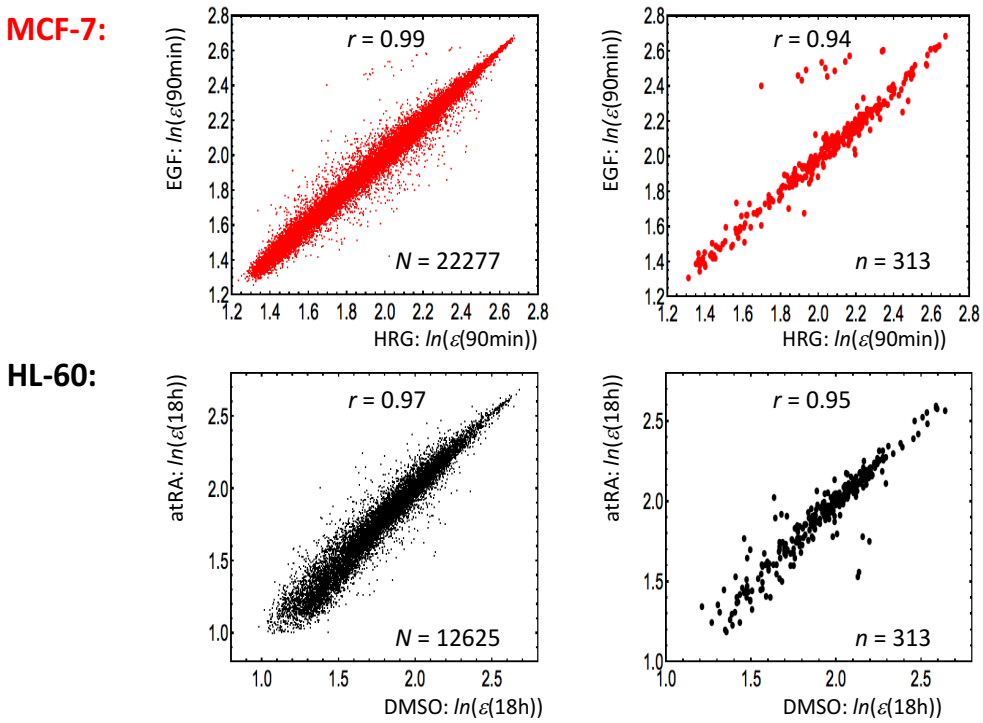
"., in spite of the large number of parameters, complex biological systems typically exhibit simple behavior that requires only a few parameters to describe, analogous to how the diffusion equation can describe microscopically diverse processes. Attempting to accurately infer all of the parameters in a complex biological model is analogous to learning all of the mechanical and electrical properties of water molecules in order to accurately predict a diffusion constant. ... Sloppiness suggests that an effective model that is microscopically inaccurate may still be insightful and predictive."

Here we explore the potentiality of a coarse grain statistical metrics on expression levels of gene ensembles [10,11], to sketch a sort of biological regulation with the framework of 'statistical mechanics'. This ends up into the clarification of the ability of a small number of hidden control parameters to provoke a global change in expression profile involving thousands of genes.

Scientists working with microarray and recent *RNA sequencing* technologies are familiar with the strict profile-invariance of independent samples relative to the same type of tissue. The expression vectors for two independent samples of the same tissue, which consist of about 20,000 ORFs, show a near-unity Pearson correlation, which points to a very strict global integration of cell populations in terms of gene expression.

The presence of a global tissue-level control is evident when we compare two samples from different tissues (**Figure 1**). A near-unity correlation is observed between the profiles of two independent samples of the same kind of tissue (**Figure 1A**). In contrast, when consider samples from different tissues, the near-unity correlation breaks down (**Figure 1B**). This simple plot suggests that global self-organization supports the phenotype that corresponds to different cell types and involves the whole expression profile.

A) Same cell type by different types of molecular stimulation



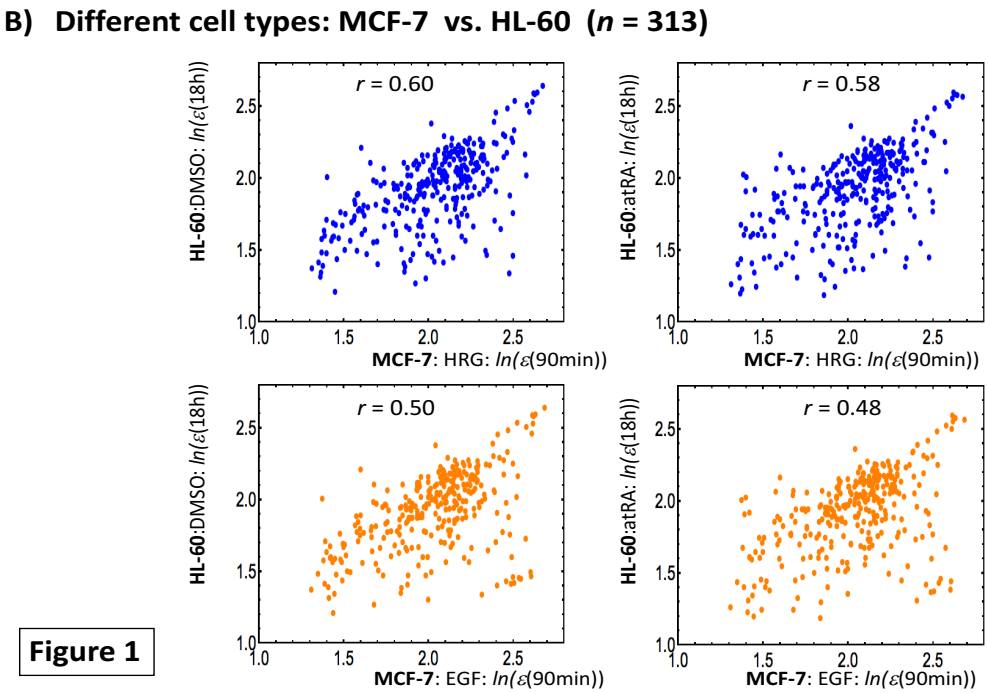


Figure 1

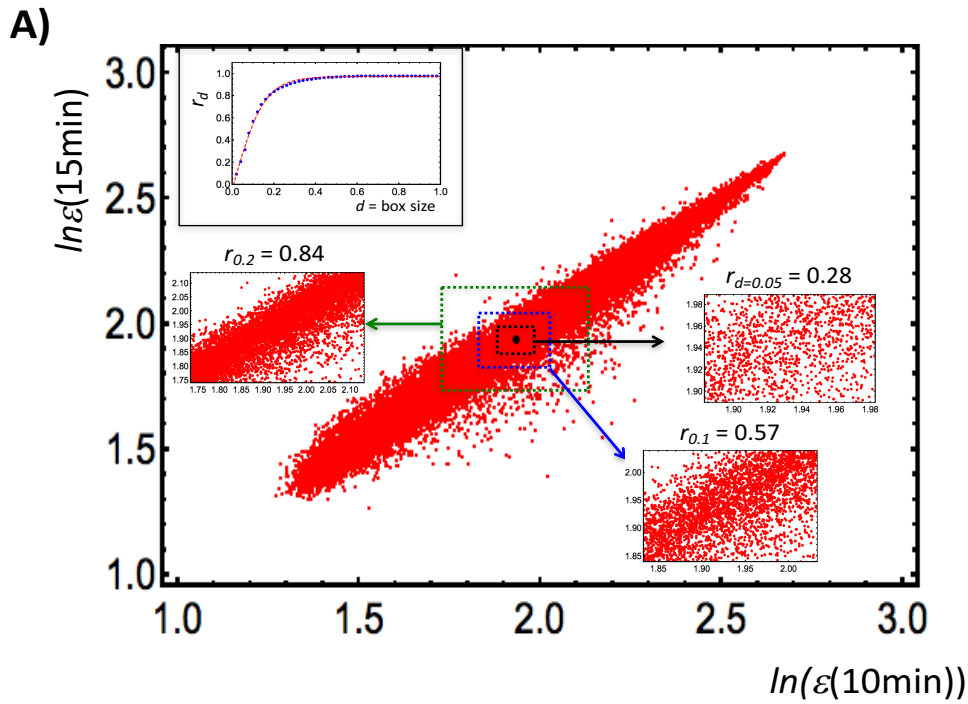
Figure 1: ***Correlation of gene expression profiles:***
A) ***The same cell types****: right panels: a near-unity Pearson correlation, r, in whole gene expression (N: total number of mRNAs) within the same cell type is shown for different types of molecular stimulation (first row: HRG- vs. EGF-simulated MCF- 7 cells; second row: DMSO- vs. atRA-simulated HL60 cells). Left panels: 313 (n) gene expressions, which have a common probe ID among four transcriptome microarray expression data (see Methods) also show a near-unity Pearson correlation within the same cell type.*
B) ***Different cell types****: The near-unity Pearson correlation between independent samples of the same cell type breaks down when the gene expression profiles come from different (HL-60 and MCF-7) cell types. ε(t) represents the ensemble of expression at time point t (N: the whole set; n: an ensemble set) and ln(ε(t)) represents its natural logarithm, where the natural log of an individual expression value is taken. Plots show t = 90min for MCF-7 stimulated cells and t = 18h for HL-60 stimulated cells.*

Furthermore, **Figure 2** suggests that the observed profile invariance is an emergent property that strictly depends on the considered gene expression range. If we change the box size from genes having very similar expression levels (low between genes expression variance) to the whole set, gene expression shifts from a stochastic to a genome-wide attractor profile (which causes a near-unity Pearson correlation). The development of this correlation demonstrates the presence of a transition that follows a tangent hyperbolic function (inset in **Figure 2**). This implies that, while myriad transcriptional regulation control circuits are active at the same time at local level (which gives a stochastic distribution; refer to **section IV**), at the global genome expression level, very efficient tissue-level self-organization accompanied with

"higher-order cooperativity" [12] emerges. Such self-organization involves the parallel regulation of more than twenty thousand of different and functionally heterogeneous genes. This in turn suggests that the ordination of gene ensembles according to their expression level could provide a potentially useful candidate to explore genome-wide regulation.

While by far the great majority of scientists have focused on the peculiarities of local gene expression control, in this work we approach the global level of gene expression regulation as an open-thermodynamic system (which is capable of exchanging small ions, water, and energy with the nuclear environment and beyond) by trying to answer some general questions:

- What is the underlying principle that regulates the whole genome expression through a global expression transition?
- Are there some differences among different biological systems regarding the global dynamics of genome expression?
- Is there a key player in the self-organization of expression?
- What is the mechanism of the self-organization determining the change in the cell fate?



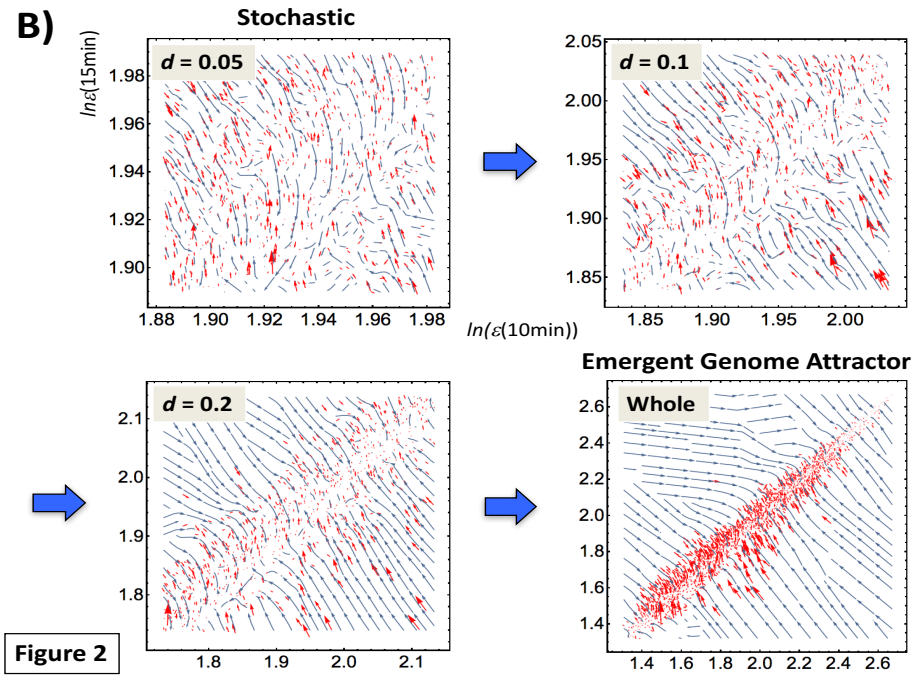


Figure 2: ***Transition of gene expression from a stochastic to a genome-wide attractor profile****:*
A) Plot shows the whole expression profiles at 10min (x-axis) and 15min (y-axis) for the HRG response in MCF-cells. A box is constructed from the center of mass (i.e., average of whole expression), (CM(10min), CM(15min)) (black dot), which contains gene expression within the range from CM(t_j) - d to CM(t_j) + d with a variable box size, d. To highlight the scaling of the Pearson correlation with box size d, r_d for d = 0.05, 0.1 and 0.2 are reported. The plot in the upper left corner shows that, between gene expression profiles, the Pearson correlation r_d follows a tangent hyperbolic function: $r_d = 0.97 \cdot tanh(6.79 \cdot d - 0.039)$ *($p<10^{-4}$), which reveals a critical transition in the correlation development, and B) Stream plots for the box sizes in Panel A. These plots are generated from the vector field values {Δx_j, Δy_j} given at expression points {x_i(10min), x_i(15min)}, where $\Delta x_i = x_i$(15min)- x_i(10min), $\Delta y_i = x_i$(20min)- x_i(15min), and $x_i(t_j)$ is the natural log of the ith expression: $x_i(t_j) = ln(\varepsilon_i(t_j))$ at $t = t_j$ (t_j = 10min or 15min; i = 1,2,.., N = 22,277). Blue lines represent streamlines and red arrows represent vectors at a specified expression point (every 2nd, 6th 10th and 20th point for d = 0.05, 0.1, 0.2, and the whole set, respectively). When we move from a small number of genes to the whole set, gene expression shifts from a stochastic to a genome-wide attractor profile.*

To address these important and still largely unsolved questions, we analyzed experimental transcriptome time-series of both microarray and RNA sequencing technology (RNA-Seq) types. The aim is to demonstrate the presence of critical transitions in different biological processes associated with changes in the cell fate. We considered i) early embryonic development in human and mouse, ii) the induction of terminal differentiation in human leukemia HL60 cells by dimethyl sulfoxide (DMSO) and all-trans-retinoic acid (atRA), iii) the activation of ErbB receptor ligands in human breast cancer MCF-7 cells by epidermal growth

factor (EGF) and heregulin (HRG), and iv) T helper 17 cell differenation induced by Interleukin-6 (IL-6) and transforming growth factor-β (TGF-β) (**Methods**).

Our approach is based on an analysis of dynamics of transcriptome data by means of the grouping (gene ensembles) of gene expression (averaging behaviors) built upon the results obtained in our recent papers [10,11] dealing with MCF-7 cell population (see more in **Methods**). These studies revealed self-organizing whole genome expression coexisting distinct response domains (critical states), where the self-organization exhibits criticality (critical behaviors) and self-similarity at a critical point (CP) – self-organized criticality control (SOC control) of overall expression.

To understand current analysis based on our previous studies, it is important to stress that

i) In each critical state, coherent (collective/coordinated) behavior emerges in ensembles of stochastic expression made by more than 50 elements [11]. Due to the coherent-stochastic behavior, *it is important to stress that characteristics of the self-organization through SOC become apparent only in collective behaviors of groups* with more than 50 genes in terms of their average value (mean-field approach),

ii) SOC control in overall expression explains the *self-organization* at a certain time developmental stage, which is *not the occurrence of (first- or second-order) phase transition* under thermal equilibrium condition. In a living cell, there are constantly incoming and outgoing fluxes (small ions, molecules, and thermodynamic energy) to develop a non-equilibrium dissipative molecular system in time. Thus, SOC in living cells emerges through the dynamics of gene expression along the time development generated by the characteristics of non-equilibrium systems [13].

iii) Furthermore, despite of many research/studies for several decades regarding SOC in science under genuine equilibrium conditions such as critical phenomena in first- or second-order phase transition [14], a general mathematical argument or formulation regarding the SOC hypothesis under non-equilibrium conditions still remains in a primitive, developing stage. This is due to the fact that there exists a myriad of scenarios of self-organization with critical behaviors under non-equilibrium conditions, where "a universal classification scheme is still missing for non-equilibrium phase transitions and the full spectrum of universality classes is unknown; it may be large or even infinite." [15]. Thus, up to the present, there exists no stereotypic view of SOC in non-equilibrium systems.

In the present study, we report the existence of a temporal interval (which differs for each analyzed system), where the change in transcriptome expression occurs via SOC at both single cell (based on RNA-Seq data) and population levels (Microarray data). Notably, the erasure of initial state SOC, i.e., the disappearance of a sandpile type critical point of the initial state (t = 0 or initial cell state) determines *when and how a crucial change in the genome state occurs* (**sections I and II**), which intriguingly coincides with real biological critical events determining the change in cell fate (**Discussion**).

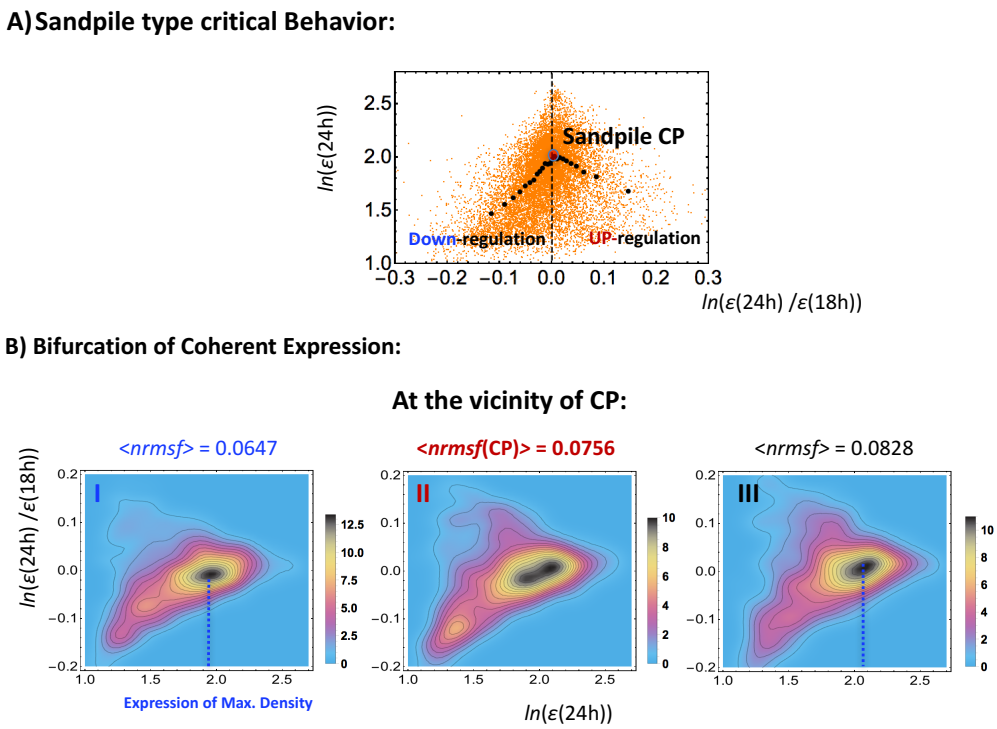
Furthermore, SOC control occurs in a model-specific manner, which reveals that the spatio-temporal profiles of self-organization in overall expression regulation are different among the different systems tested; distinct critical states coexist (**section III**). Furthermore, the emergent property of the coherent dynamics in critical states helps us to understand how the emergent sloppiness exhibits in the genome-wide expression dynamics (**section IV**).

In **sections V and VI**, we demonstrate that a molecular stressor such as HRG in MCF-7 cells and DMSO in HL-60 cells, which induce cell differentiation, dynamically perturbs the self-organization in SOC, and as a result, terminal cell fates occur at the end of a dissipative pulse-like global perturbation in self-organization. The perturbation of SOC occurs due to the exchange of expression flux among critical states through the cell nucleus environment as an open thermodynamic system. Quantitative evaluation of such flux flow reveals a mechanical picture of interactions of critical states ("genome engine"; **Discussion**) and their roles in self-organization; most notably, sub-critical states (ensembles of genes with low-variance expression) are the central players for the temporal development of global self-organization.

The elucidation of a statistical mechanism of cell-fate change revealed through the perturbation of SOC in open thermodynamic gene regulations could lead to new advances in our understanding of the dynamic aspects of epigenomics together with the material bases of biological regulation (**Discussion**).

Results

The coherent dynamics of an ensemble of stochastic expression can be represented as a hill-like function, which is defined as a 'coherent expression state (CES)', with regard to the probability density profile in the regulatory space (expression vs. fold change in the expression) [10,11]. The emergence of a CES at around the critical point (CP) that marks the transition allows us to describe critical transitions in distinct cell types. **Figure 3A** shows, as an example, that in DMSO-stimulated HL-60 cell differentiation, through the grouping of expression based on the fold change in expression (see why in **Methods**), a sandpile-type critical behavior is observed at 18-24h.



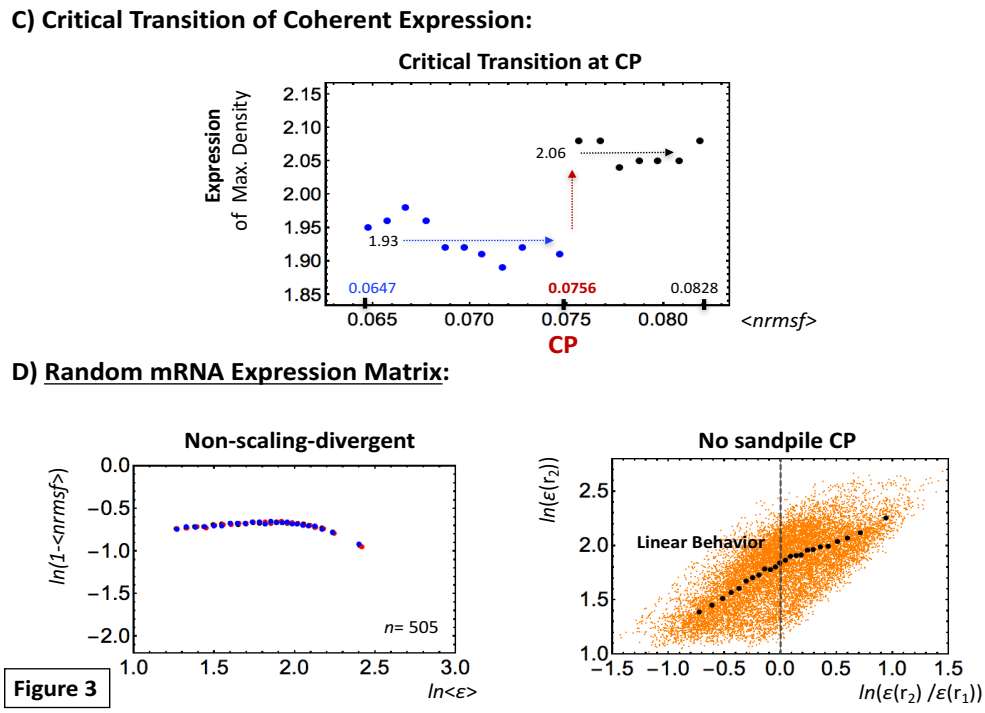


Figure 3*: Self-organized criticality (SOC) in the DMSO-stimulated HL-60 cell fate:*
A) The grouping of the whole expression at 18-24h generates 25 groups with an equal number of 505 elements (mRNAs) according to the fold change in expression. A plot of the average value for each group in log-log space (x: fold change at 18-24h vs. y: expression at 24h) reveals a sandpile-type critical behavior at the critical point (CP), where the CP in terms of the group average (< >) occurs at the near-zero-fold change (null expression change; x-axis) with nrmsf(CP) = 0.076. Orange dots represent single mRNA expression in the background.
*B) The probability density function for the ensemble of expression (coherent expression state: CES) is shown in the regulatory space (x: expression at 24h vs. y: fold change in expression at18h-24h). Plots show that around nrmsf(CP) = 0.076, a CES (highest density: x< 2.0) is annihilated and a new CES (highest density: x> 2.0) is bifurcated. The left to right panels show the sequence of the bifurcation-annihilation event: before (**I**: 0.060 <nrmsf< 0.070: <nrmsf> = 0.0647), onset (**II**: 0.071 <nrmsf< 0.081: <nrmsf> = 0.0756)) and after the event (**III**: 0.078 <nrmsf< 0.088: (<nrmsf> = 0.0828). Colored bars represent the probability density.*
*C) The corresponding plot (B) reveals that a step function-like critical transition occurs at nrmsf(CP) = 0.076 in the space (x: <nrmsf> of coherent state vs. y: its expression of the highest density at 18h). Blue and black arrows represent average values of the density trends before and after the transition, respectively. The plot shows at the vicinity of the CP, the occurrence of a self-similar bifurcation (symmetry breaking) in the expression profile to that of the overall expression (see **section III**).*
D) Random mRNA DMSO expression matrix reveals that, in this case, a CP does not exist. This is confirmed by anomalous features of the corresponding SOC (Methods): non-scaling-divergent (3 different time points are shown by colors) and non-sandpile critical behavior of random expression between two different time points (orange dots: single random expression). The random matrix is made by randomly selecting each matrix component (i,j) from the original DMSO expression matrix (12625 expression (i) times 13 time points (j)). We observed similar linear correlative behaviors for other cells in both microarray and RNA-Seq data.

Around the critical behavior in terms of *nrmsf* (normalized root mean square fluctuation, see **Methods**), symmetry breaking corresponds to the annihilation of a CES at a lower expression level and the bifurcation of another CES at a higher expression level through a flattened profile in the regulatory space (**Figure 3B**). The maximum density of the coherent state follows a step functional like critical transition at the CP (**Figure 3C**). Moreover, the critical transition of coherent expression at the vicinity of the CP shows the self-similar behavior to that of the overall expression (see **section III**) in DMSO-stimulated HL-60 cells, i.e., the sandpile type critical point exhibits a critical transition (called sandpile type critical transition). We also observed a sandpile type critical transition at a CP in HRG-stimulated cell differentiation [11].

Next, when the DMSO-induced expression matrix is randomly shuffled, no sandpile type CP is present (**Figure 3D**: right panel): the fold change scales almost linearly with the logarithm level of expression. The corresponding frequency distribution changes from non-Gaussian (**Figure 9B**) to Gaussian distribution (**Figure 9C)** according to *nrmsf*. This shows that i) gene expression becomes random gene expression by random shuffling of gene expression, and ii) the *randomized expression destroys the sandpile type critical behavior* seen on **Figure 3A**.

Hence, it is crucial to investigate the existence of a critical point for SOC control in overall expression.

I. Perturbation of SOC control and the Genome-State Change

We investigate the occurrence of critical transitions in distinct cell types. First, we examine whether the sandpile-type critical transition (see Methods) occurs around the zero-fold change (i.e., null change in expression) between different time points. The critical point (CP), at the top of the sandpile, corresponds to a group of genes that shows almost no (average) change in expression. Next, we assess whether or not the CP is a fixed point in time by evaluating if the average *nrmsf* value of the CP group changes over time. The basic hypothesis that justifies the choice of *nrmsf* as the metrics to evaluate gene expression dynamics is that the entity of gene expression group scales with the topology-associated chromatin domains (TAD) [16-20]. The degree of gene expression normalized fluctuation is the image in light of the entity of chromatin

remodeling, i.e. *nrmsf* is expected to concern with the physical plasticity of genomic DNA and the high-order chromatin structure [11]. In other words, we can expect that, as the flexibility of a given genome patch increases, so should the *nrmsf* of the corresponding genes. To confirm this further, the scaling-divergent behaviors (**Figure 4** and **Methods**) between *nrmsf* and average expression, another important feature of SOC, may reveal a quantitative relation of the aggregation state of chromatin. In **Supplementary S1 file**, we showed that a collective behavior of gene expression (as coherent-stochastic behavior: CSB) does exhibit in the power law scaling through interactions among genes.

The transcriptome analysis based on a mean-field (grouping) approach (**Methods**) reveals that sandpile transitions occur and the position of the CP exhibits time-dependence in terms of *nrmsf*, which reflects the temporal development of SOC, i.e., the CP is not a fixed point (**Figure 4**). Interestingly, the CP of initial state disappears in time, suggesting the occurrence of a crucial change in the genome state (**Figure 5)**. Regarding critical transitions around the CP (**Figure 6**), different types of dynamical bifurcation or annihilation of a characteristic coherent expression state occur around the CP in different cell models:

1) Unimodal-flattened-bimodal transition for MCF-7 cell- HRG and -EGF models, and
2) Unimodal-flattened-unimodal transition for the HL-60 response to atRA and DMSO; both models point to symmetry breaking in the expression profile.

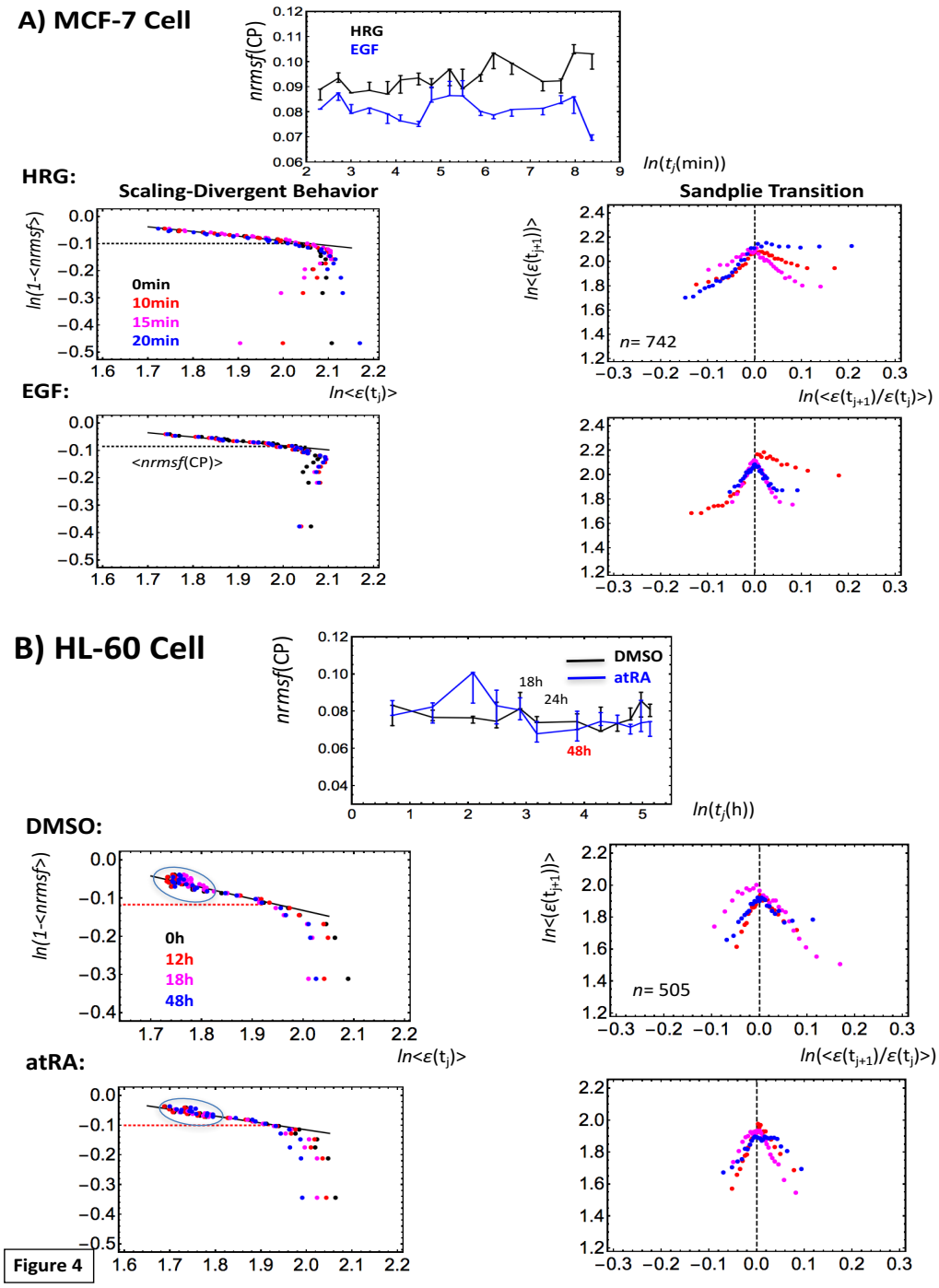


Figure 4: ***Time-development of the characteristic behaviors of SOC****:*
A) MCF-7 cells and B) HL-60 cells. At each experimental time point, t_j (A: 18 time points and B: 13 points: Methods), the average nrmsf value of the CP, nrmsf(CP(t_j)) is evaluated at the sandpile type critical point (top of the sandpile: right panels). In the top center panels, nrmsf(CP(t_j)) is plotted against the natural log of t_j (A: min and B: hr). Error bar at CP(t_j) represents how sensitive nrmsf(CP(t_j)) is around the CP, where the bar length corresponds to the change in nrmsf in the x-coordinate (fold-change in expression) from x(CP(t_j)) - d to x(CP(t_j)) + d for a given d (A: d = 0.005; B: d =0.01). Temporal averages of nrmsf(CP(t_j)) are A) $\langle nrmsf(CP)\rangle_{HRG}$= 0.094 and $\langle nrmsf(CP)\rangle_{EGF}$ = 0.081, and B) $\langle nrmsf(CP)\rangle_{DMSO,\ atRA}$ = 0.078.

A) MCF-7 cells: The temporal trends of nrmsf(CP(t_j)) are different for HRG and EGF. The onset of scaling divergence (left panels: second and third rows) occurs at around <nrmsf(CP)> (black dashed line), and reflect the onset of a 'genome avalanche' (Methods).
B) HL-60 cells: The trends of nrmsf(CP(t_j)) for the responses to both DMSO (black line) and atRA (blue) seem to be similar after 18h (i.e., global perturbation; see ***section VI****). The scaling-divergent behaviors for both DMSO and atRA reveals the collapse of autonomous bistable switch (ABS [11]) exhibited by the mass of groups in the scaling region (black solid cycles) for both DMSO and atRA. The onset of divergent behavior does not occur around the CP (<nrmsf(CP)>= 0.078), but is extended from the CP (see main text).*
The power law of scaling behavior in the form of 1- <nrmsf> = $\alpha<\varepsilon>^{-\beta}$ is:
A) α = 1.29 & β = 0.172 ($p< 10^{-10}$) for HRG, and α = 1.26 & β = 0.157 ($p<10^{-9}$) for EGF;
B) α = 1.60 & β = 0.301 ($p< 10^{-6}$) for DMSO, and α = 1.42 & β = 0.232 ($p< 10^{-6}$) for atRA. Each dot (different time points are shown by colors) represents an average value of A) n = 742 mRNAs for MCF-7 cells, and B) n = 505 mRNAs for HL-60 cells.

These results offer the following insights:

1. **SOC control in different cell models:** The self-similarity of symmetry breaking around the CP suggests the existence of critical states (distinct transcription response domains) in different cell models (see **section III**). Furthermore, the self-similarity in the overall expression suggests the occurrence of i) a unimodal-flattened-bimodal transition for MCF-7 cell fates, and ii) a unimodal-flattened-unimodal transition for HL-60 cell fates, which will be shown to be robust over time (**section III**), except for a transient change (see **Figure 6A**, an example in the HRG response: a bimodal-bimodal transition at 15-20 min due to the bifurcation in the unimodal profile; refer also to Figure 7B in [10]). The unimodal-unimodal transition in HL-60 cell fates stems from the fact that a pair of two coherent expression states (a bimodal expression profile as an autonomous bistable switch [11]) collapses to a single coherent state in the scaling region (low *nrmsf* region: **Figure 4B**).

2. **Perturbation of SOC control**: The temporal development of SOC control indicates dynamic changes in critical states in terms of both shuffling of gene expression between critical states, and changes in expression profile. This implies the perturbation of SOC control through the interaction between critical states (see **sections V and VI**). Interestingly, regarding HL-60 cell fates, at 12-18h, pulse-like global perturbations involving the regulation of critical states occur for the responses to both DMSO and atRA (**section VI**). After 18h, the temporal trends of the CP(t_j) of DMSO and atRA in

terms of *nrmsf* become similar (**Figure 4B**). Note that the bifurcation-annihilation events of CES around the CP for both DMSO and atRA at 24-48h become almost identical (data not shown). This shows that the dissipation of the stressor-specific perturbation in HL-60 cells drives the cell population toward the same attractor state.

3. **Genome-state change:** The erasure of the CP of initial state (**Figure 5**) indicates that the change in the genome state occurs at 3h in HRG stimulated MCF-7 cell, at 24h and 48h in DMSO and atRA stimulated HL-60 cells, respectively; these HL-60 genome-state changes further confirm that both DMSO and atRA stimulated HL-60 cells become the same at 48h. Intriguingly, the erasure of the initial-state critical behavior occurs among different cell types; divergent behavior in up-regulation (a partial erasure) disappear in HRG-stimulated MCF-7 cells, whereas in both DMSO- and atRA simulated HL-60 cells, the full erasure of divergent behaviors of up- and down-regulations occurs. As demonstrated in **section VI**, these genome-state changes occur after dissipative pulse-like global perturbations in SOC control (at 12-18h for HL-60 cells and at 15-20min for HRG stimulated MCF-7 cell). In contrast, the genome-state change does not occur in EGF stimulated MCF-7 cell (**Figure 5**; refer also to local perturbation; **section VI**), which is consistent with cell proliferation (no differentiation) in the EGF response [21,22]. Note: the developments of the genome state coincide with real biological critical events determining the change in cell fate (see Discussion).

To dig deeper into these findings, in the following sections, we investigate the transcriptome SOC control in embryonic development at a single-cell level based on next-generation RNA sequencing data, and address mechanism of the genome-state change in terminal cell fates revealed through the perturbation of genome-wide self-organization.

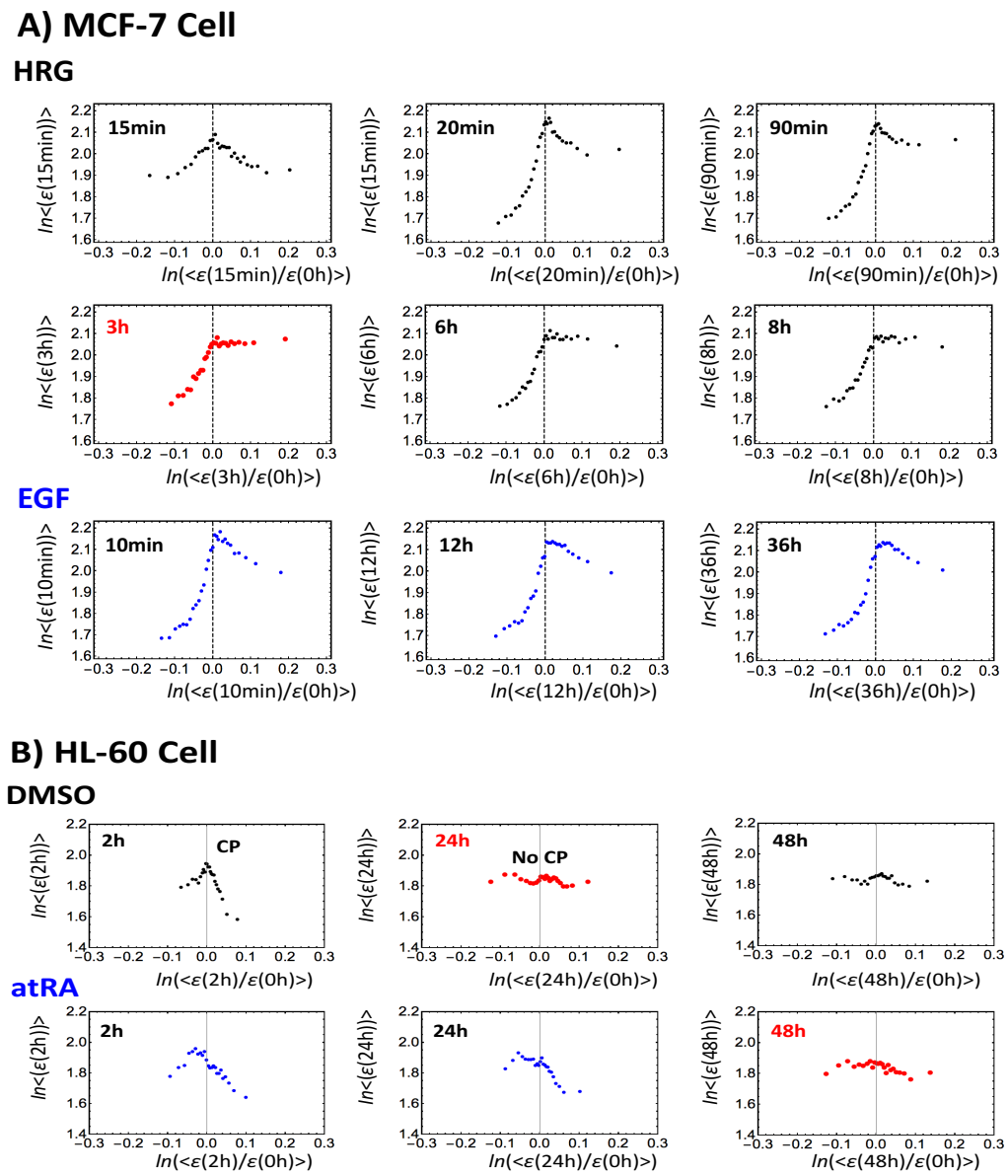


Figure 5

Figure 5: ***Genome-state change revealed through the erasure of initial-state critical point in overall expression (cell population level):*** *The grouping of overall expression at $t = t_j$ ($j \neq 0$) according to the fold change in expression from initial overall expression ($t= 0$) shows how the initial-state critical point erases in time, i.e., a sandplie profile is destroyed in overall expression at $t = t_j$ from $t= 0$. This event points to the time when the genome-state change occurs. In the x-axis, $ln(<\varepsilon(t)/\varepsilon(0h)>)$ (t: min or hr) represents the natural log of the ensemble average (< >) of the fold change in expression, $\varepsilon(t)/\varepsilon(t= 0h)$, and in the y-axis, $ln(<\varepsilon(t)>)$ represents the natural log of ensemble average of expression, $<\varepsilon(t)>$.*
A) MCF-7 cells: in HRG-stimulated cell (black), the divergent behavior in up-regulation no longer exits while remaining the same shape after 3h. This suggests that the genome-state change occurs at 3h (red)

through the erasure of the initial state up-regulation process (partial erasure). In contrast, in EGF-stimulated cell (blue), the almost same sandplie profile remains for up to 36h, which suggests that no genome-state change occurs (see the local perturbation in ***Figure 13****).*
*B) HL-60 cells: a CP erases at 24h (red) and 48h (red) in DMSO- (black) and atRA-stimulated (blue) cells through the disappearance of divergent behaviors in both up- and down-regulations of the initial state (full erasure), respectively. Furthermore, these plots suggest that the epigenetic state after 48h, in DMSO- and atRA-stimulated cells become the same. Note: Pearson correlations of overall expression between different time points are a near-unity (****Figure 1A****).*
Each dot represents an average value of A) n = 742 mRNAs for MCF-7 cells, and B) n = 505 mRNAs for HL-60 cells.

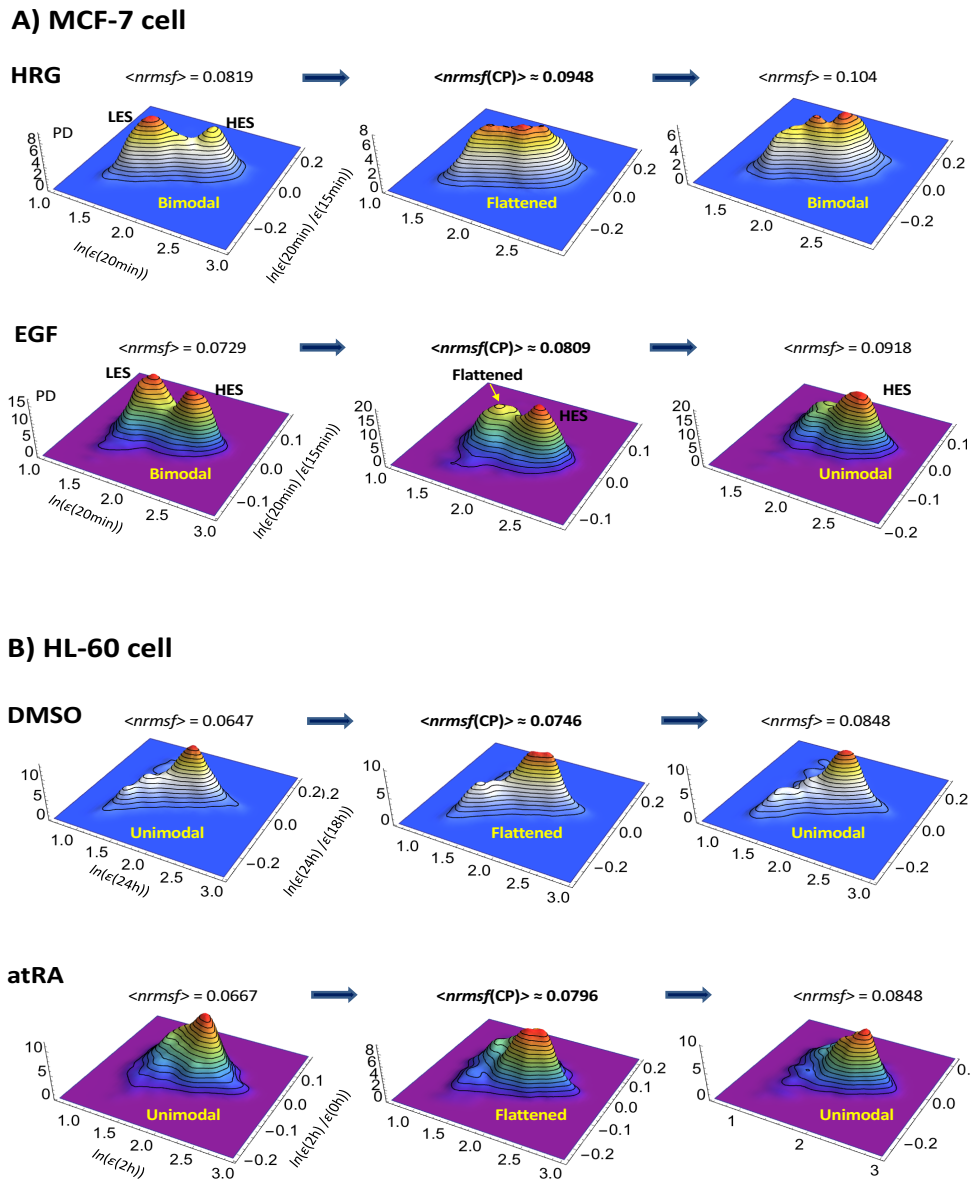


Figure 6

Figure 6: The self-similar dynamical bifurcation or annihilation of a characteristic coherent expression state at the vicinity of a critical point in different cell types. *A coherent expression state (CES), which contains approximately 1000 expression points, is bifurcated or annihilated around the CP through the nrmsf grouping. The average nrmsf of an expression group, <nrmsf> is evaluated for a variable range: $x + 0.01 \cdot (m-1) < nrmsf < x + 0.01 \cdot m$ (integer, $m \leq 10$; $x = 0.7$, 0.8 and 0.9). The probability density function (PDF) in the regulatory space around <nrmsf(CP)> shows:*
A) MCF-7 cells: in the HRG response at 15-20min, around <nrmsf(CP)> ($\langle nrmsf(CP)\rangle_{HRG} = 0.094$ and $\langle nrmsf(CP)\rangle_{EGF} = 0.081$; ***Figure 4A****), PDF exhibits a bimodal - flattened unimodal -bimodal transition, in which a bimodal profile points to the existence of two CESs: one represents low expression states (LESs) and another represents high expression state (HES); the valley defines the boundary between low and high expression [10]. In the EGF response (second row), above <nrmsf(CP)>, a low expression state (LES) is annihilated and only high expression state (HES) exists, i.e., the occurrence of a unimodal-bimodal transition. In the HRG response, at time period other than 15-20min, a unimodal-bimodal transition occurs [10]. The result shows the self-similar symmetry-breaking event to the overall expression for MCF-7 cells, even at a transient change in the HRG response: a bimodal-flattened-bimodal transition at 15-20 min.*
B) HL-60 cells: a unimodal-flattened-unimodal transition occurs at 18-24h (pseudo-3-dimensional PDF plots of ***Figure 3B****) for DMSO and at 0-2h for atRA, which reveals again the self-similarity to the overall expression for HL-60 cells (****section III****).*

II. Control of Single Zygotic Cell Embryonic Development and T helper 17 Cell Differentiation Through Self-Organized Criticality

RNA-Seq approach allowed us to check the tenability of SOC control at a single-cell level. Here we analyze RNA-Seq data related to immune cell differentiation, and human and mouse embryonic development.

Figure 7A shows Pearson correlations between gene expression profiles related to the zygote and early embryo single-cell states. We previously demonstrated (see **Figure 1**) that the between-profiles Pearson correlation follows a tangent hyperbolic function with an increasing number of genes. A similar result is observed in both human and mouse early embryo development, and indicates the presence of a critical transition. This transition takes place between the 4-cell and 8-cell states in human and between the middle and late 2-cell states in mouse.

Notably, the correlation transition corresponds to the onset of the breakdown of SOC control in early embryo development according to the number of cell states starting from the zygote single-cell state. In both human and mouse embryos, the maternal SOC zygote controls are broken down through cell-state development, and the overall expression transfers to a

stochastic-type. In human, the maternal SOC zygote control survives until the 8-cell state stage. After the morula state (**Figure 7B**), no sandpile-type critical point exists, and the overall gene expression profile is fully stochastic compared with the zygote overall expression (see **Supplementary Figure S1**), which indicates that the memory of the zygote genome expression is lost at the morula state. In contrast, in the mouse embryo (**Figure 7C**), the maternal SOC controls survive from the zygote to the 2-cell state (middle stage).

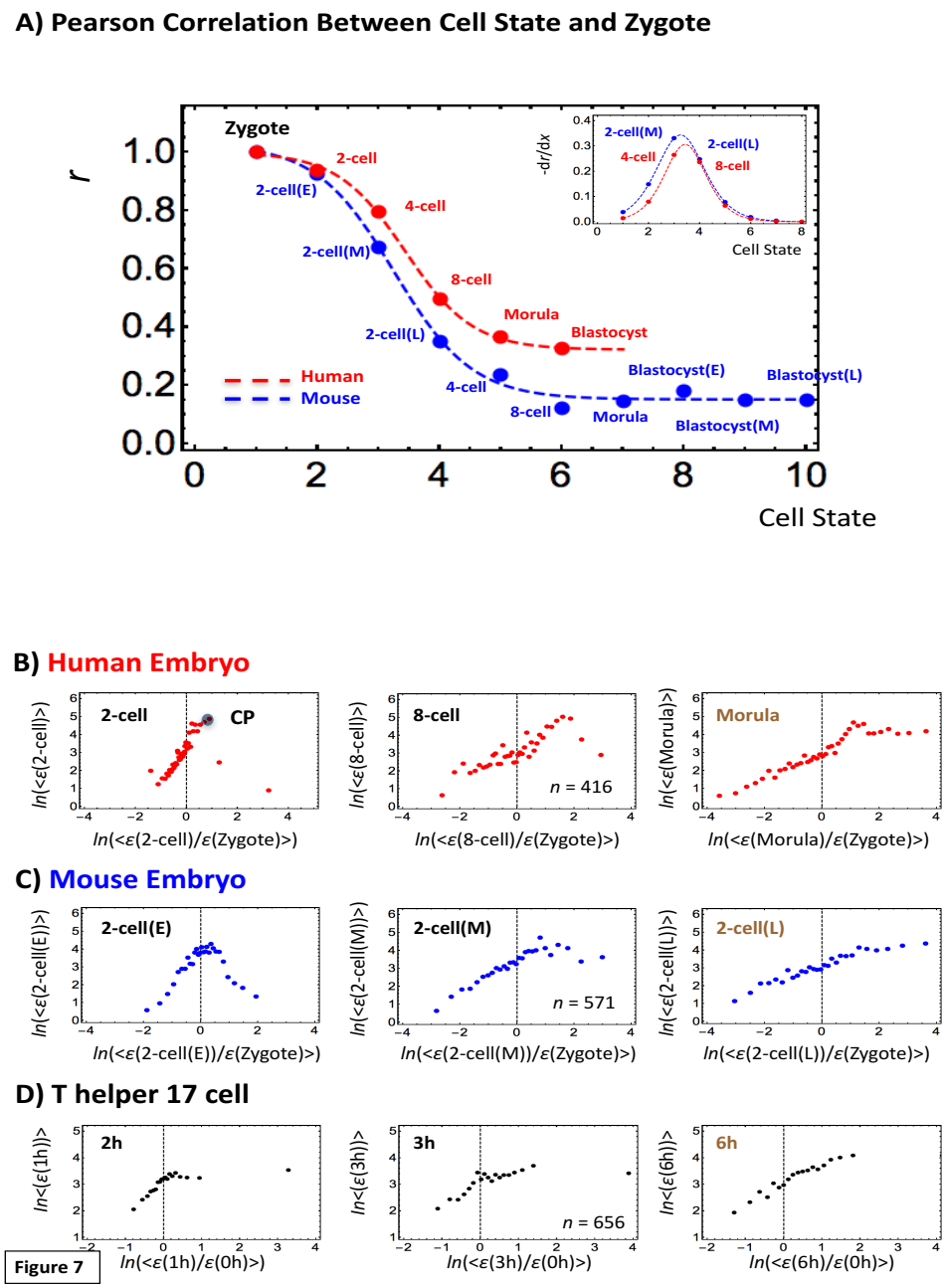


Figure 7

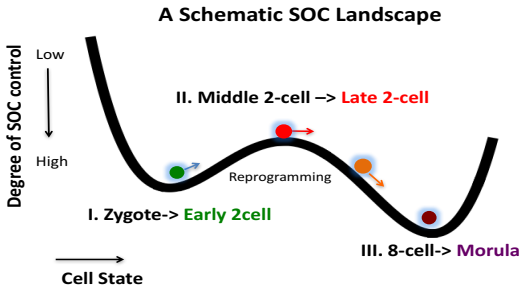
Figure 7: ***Genome-state change revealed through the erasure of initial-state critical point in overall expression (single cell level):***
Transcriptome RNA-Seq data (RPKM) analysis for A-C) human and mouse embryo development and D) T helper 17 cell differentiation.
A) In both human and mouse embryos (red: human; blue; mouse; refer to development stages in ***Methods****), a critical transition is seen in the development of the overall expression correlation between the cell state and the zygote single-cell stage, with a change from perfect to low (stochastic) correlation: the Pearson correlation for the cell state with the zygote follows a tangent hyperbolic function:* $a - b \cdot tanh(c \cdot x - d)$*, where x represents the cell state with a= 0.59, b= 0.44 c= 0.78 and d = 2.5 (p<* 10^{-3}*) for human (red dashed line), and a= 0.66, b= 0.34 c= 0.90 and d = 3.1 (p<* 10^{-2}*) for mouse (blue). The (negative) first derivative of the tangent hyperbolic function, -dr/dx, exhibits an inflection point (zero second derivative), indicating that there is a phase difference between the 4-cell and 8-cell states for human, and between the middle stage and late 2-cell states for mouse (inset); a phase transition occurs at the inflection point. Notably, the development of the SOC control in the development of the sandpile-type transitional behaviors from the zygote stage (30 groups; n: number of RNAs in a group:* ***Methods****) is consistent with this correlation transition:*
B) In human, a sandpile-type CP (at the top of the sandpile) disappears after the 8-cell, and thereafter there are no critical points. This tells that the zygote SOC control in overall expression (i.e., the zygote self-organization through criticality) is destroyed after the 8-cell state, which indicates that the memory of the zygote genome expression is lost through a stochastic pattern as in the linear correlation trend (refer to the random expression matrix in ***Supplementary Figure S1*** *and to* ***Figure 3D****). The result suggests that reprogramming of the genome occurs after the 8-cell state.*
C) In mouse, a sandpile-type CP disappears right after the middle stage of the 2-cell state and thereafter a stochastic linear pattern occurs, which suggests that reprogramming of the genome after the late stage of the 2-cell state destroys the SOC zygote control.
D) In Th17 cell differentiation, a sandpile-type CP disappears at 3-6h through a stochastic linear pattern. Therefore, the plot suggests that the genome state change occurs around 6h in a single Th17 cell.

The breakdown of early SOC zygote control in overall expression indicates that significant global perturbation (refer to **section VI**) occurs to destroy the SOC zygote control in early embryo development. The human scenario may be explained by the known fact that the genome of the human embryo is not expressed until the 4-8 cell stage, which suggests that no apparent significant perturbation as reprogramming occurs in human embryo early development (see more in Discussion).

Along similar lines of reasoning, T helper 17 (Th17) cell differentiation shows that initial SOC control (t =0) is destroyed at 3-6h or at 6h (**Figure 7D**), which reveals the Th17 genome-state changes around 6h after the induction. The embryo and immune cell results confirm the presence of specific SOC controls, not only in large cell populations, but also at a single-cell level.

Next, we examine the development of SOC control between sequential cell states. **Figure 8** reveals that from the zygote to morula state in mouse, there occurs the transition of

one sandpile type critical regulation to another through non-critical transition type (absence of critical behaviors: non-SOC control) of regulation at a middle-late 2-cell states. This confirms that the reprograming of the early mouse embryo cells from zygote destroys SOC control to initiate self-organization in the new embryonal genome at the late 2-cell state exhibiting a stochastic overall expression pattern (**Figure 7C**). Thereafter again, the SOC control takes over the embryo development. Non-SOC control exhibits almost a linear behavior, which represents a characteristic of randomized expression (see **Supplementary Figure S1** and refer to the similar linear behavior in **Figure 3D** in the different cell type). The transition of SOC control through the non-SOC control suggests that SOC-control landscape, i.e., a valley (SOC control) - ridge (non SOC control) - valley (SOC control) is seen in early mouse embryo development. Thus, the genome expression dynamics in the early embryo, through the development of SOC control, is consistent with the 'epigenetic landscape' frame, in the broad terms of global activation-deactivation dynamics of the genome as a whole, as in the case of DNA de-methylation-methylation landscape [23].



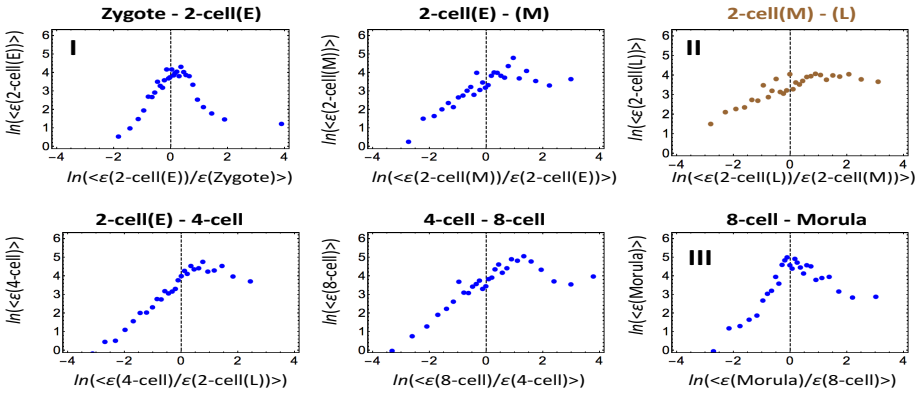


Figure 8

Figure 8: ***The SOC control landscape as revealed by a sandpile-to-sandpile transition in mouse embryo development.*** *The development of a sandpile transitional behavior between sequential cell states suggests the existence of the SOC-control landscape (first row: schematic picture of a valley-ridge-valley transition; x-axis: cell state; y-axis: degree of SOC control). The second and third rows show that a sandpile-to-sandpile transition occurs in mouse embryo development from the zygote single-cell stage to the morula cell state:* ***I****: a sandplie (i.e., critical transition) is well developed from the zygote single-cell stage to the early 2-cell state =>* ***II:*** *a sandplie is destroyed from the middle to the late 2-cell state, which exhibits stochastic expression (i.e., no critical transition; refer to random mouse expression matrix in* ***Supplementary S1****) =>* ***III:*** *a sandplie is again well-developed from the 8-cell to the morula state. The result shows that a significant perturbation (reprogramming) in self-organization occurs from the middle stage to the late 2-cell stage through a stochastic overall expression (refer to* ***Figure 7****). Note: the degree of SOC control qualitatively represents a high degree for a well-developed shape of the sandpile type transition (SOC control), intermediate for a weaken (broken) sandpile, and low degree for a non-SOC control. The latter is due to stochastic expression. The linear behavior (absence of critical point) in mouse embryo development is also exhibited as a low Pearson correlation (r ~ 0.21) (****Figures 7A, C****).*

III. Distinct Time-Averaged Critical States in Terminal Cell Fates

The self-similarity around the CP (**Figure 6**) to overall expression highlights three distinct distribution patterns of gene expression relative to different critical states:

i) A unimodal profile corresponding to high-variance expression for a super-critical state, which corresponds to a flexible genomic compartment for dominant molecular transcriptional activity.

ii) A flattened unimodal profile (intermediate-variance expression) for a near-critical state, corresponding to an equilibrated genomic compartment, where the critical transition emerges.

iii) Bimodal for HRG and EGF responses in MCF-7 cells or Unimodal profiles for the DMSO and atRA responses in HL-60 cells for a sub-critical state (low-variance expression). The sub-critical state is the compartment where the ensemble behavior of the genomic DNA structural phase transitions is expected to play a dominant role in the expression dynamics. In the HRG response, a subset of consecutive genes pertaining to the same critical state (called barcode genes) on chromosomes, spanning from kbp to Mbp, has been shown to be a suitable material basis for the coordination of phase transitional behaviors (refer to Figure 8A in [11]).

The presence of different distributions of a sub-critical state points to different forms of SOC in biological processes with a varying sub-critical state, as for the bimodal character of the corresponding expression profile.

In the next section, we will show the existence of a flux ("genetic energy flow") between critical states, which induces the temporal fluctuation of critical states as shown in **Figure 4**. For evidence of such a flow, we need to focus on the average critical state, and then show how perturbation from this average generates activation/deactivation fluxes across critical states.

The self-similarity of the symmetry break around the CP suggests that critical states have distinct profiles. The degree of *nrmsf* acts as the order parameter for distinct critical states in mRNA expression [10,11]. Thus, to develop a sensible mean-field approach, we estimate bimodality coefficients along *nrmsf* by the following steps (**Figure 9**):

i) Sort and group the whole mRNA expression according to the degree of *nrmsf*. The *nrmsf* grouping is made at a given sequence of discrete values of *nrmsf*, and

ii) Evaluate the corresponding temporal average of the bimodality coefficient over time to examine if the mean field (behavior of averages of groups) shows any transitional behavior to distinguish critical states.

Figure 9 clearly reveals that, in the overall expression, the unimodal-flattened-bimodal transition of MCF-7 cells and the unimodal-flattened-unimodal transition of HL-60 cells (i.e., HL-60 model). Notably, mean-field behaviors of bimodality coefficients follow tangent hyperbolic functions (**Figure 9A**) for MCF-7 cells and Heaviside-like step functions for HL-60 cells (**Figure 9B**). The distinct time-average behaviors between different cell types further support cell type-specific SOC control.

On the contrary, a randomly shuffled expression matrix (DMSO: HL-60 cells) does not exhibit any apparent transitional behavior (**Figure 9C**), which further confirms the existence of distinguished averaged critical states in both MCF-7 and HL-60 cells.

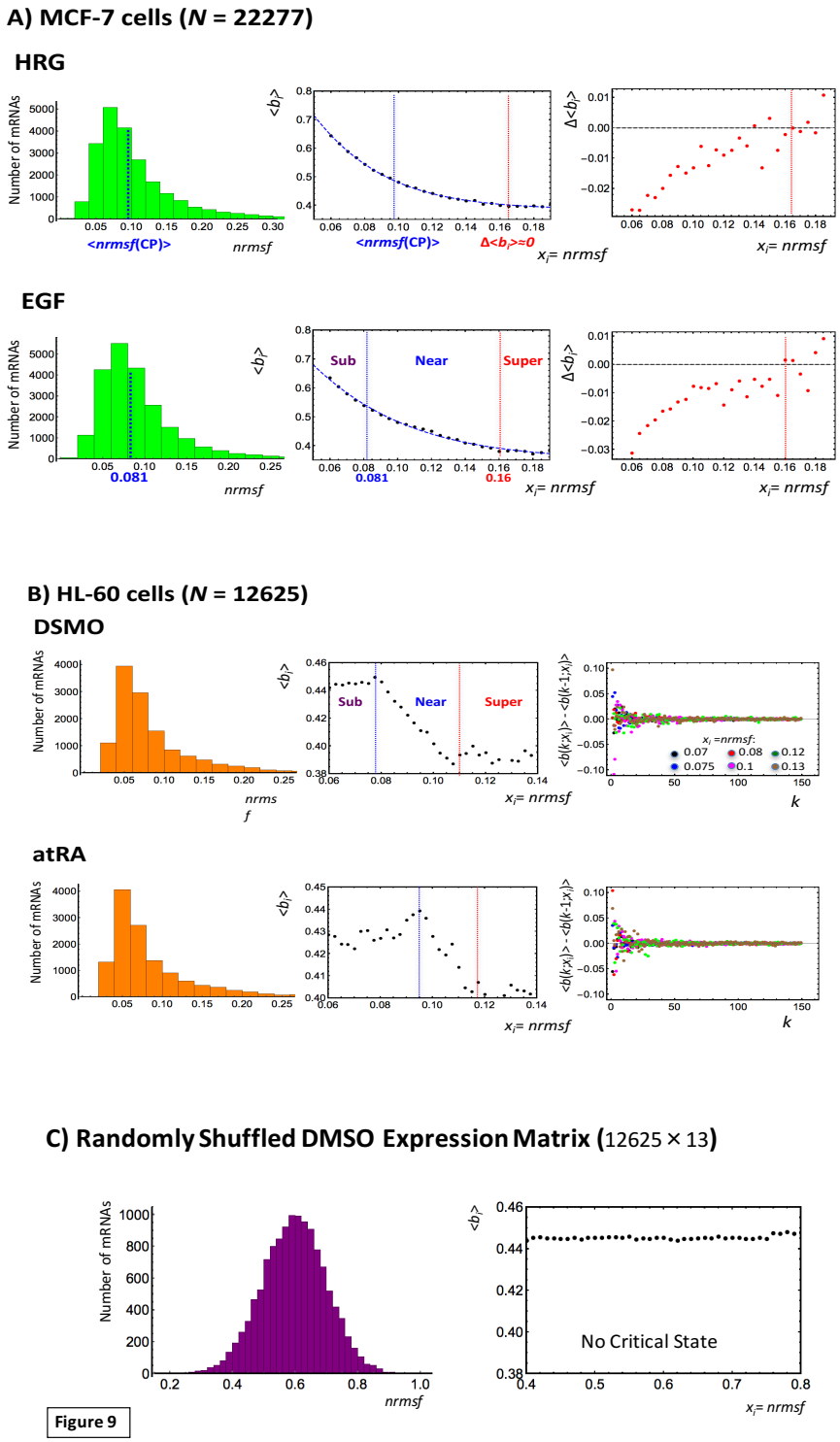


Figure 9: ***Critical states revealed through distinct functional behaviors of the bimodality coefficient:*** *A) MCF-7 cells, B) HL-60 cells and C) Random DMSO expression matrix (see* ***Figure 3D****). The left panels show the frequency distribution of expression according to the degree of nrmsf. The center panels show the temporal average of Sarle's bimodality coefficient, $<b_i>$, of the ith group over time; the value 5/9 represents the threshold between the unimodal (below 5/9) and bimodal or multimodal distributions (above). The grouping of expression is made at a specific sequence of discrete values of nrmsf (x_i: $nmrsf_i$*

= i/100; i: integers) with a fixed range $x_i - kdx < x_i < x_i + kdx$. The values of k and d are set to be k= 150 and d = 0.0001 for MCF-7 and HL-60, and k = 100 and d = 0.001 for a random matrix based on the convergence of the bimodality coefficient. The convergence of the difference of bimodality coefficients at x_i as k increased is shown between the next neighbors, $<b_i(k;x_i)> - <b_i(k-1;x_i)>$ for HL-60 cells in the right panels of B). The 6 colored dots represent the convergent behaviors of different nrmsf points.

The behavior of the time average of the bimodality coefficient reveals

A) Tangent hyperbolic functions, $b_i = a - tanh(b + c\langle nrmsf\rangle_i)$; a= 1.38 and 1.35; b = 0.123 and 0.301; c = 13.8 and 10.2 for HRG ($p< 10^{-4}$) and EGF ($p< 10^{-10}$), respectively,

B) Heaviside step function-like transitions for HL-60 cell fates, and

C) No transition for a random DMSO expression matrix, which importantly reveals that random noises through the formation of Gaussian distribution destroy a sandpile critical behavior.

*Based on these distinct behaviors, we can determine the boundaries of averaged critical states (**Table 1**; see the **section III**):*

*A) Critical states are defined by two points: the average CP, <nrmsf(CP)>, the onset of genome avalanche (**Figure 4A**), for the upper boundary of the sub-critical state, and the point where the change in the bimodality coefficient, $\Delta<b_i>$, reaches zero for the lower boundary of the super-critical state; the near-critical state is between them, and*

B) Step function-like transitions reveal the boundaries of averaged critical states, where the near-critical state corresponding to the transitional region separates the other states.

Two different behaviors in terms of the bimodality coefficient (see **Table 1**) are evident:

a) In MCF-7 cells, the averaged CP corresponds to the onset of scaling-divergent behavior and the unimodal-bimodal symmetry breaking of expression profile, so that the averaged sub-critical state (bimodal profile) is below the averaged CP (*nrmsf*< 0.094 for HRG and *nrmsf*< 0.081 for EGF; **Figure 4**). The super-critical state (unimodal profile) is above the point where the change in the bimodality coefficient reaches zero (*nrmsf*> 0.165 for HRG and *nrmsf*> 0.160 for EGF), and the near-critical state (flattened unimodal) is between them (**Figure 9A**).

b) In HL-60 cells, the transition of bimodality coefficients clearly distinguishes average critical states (**Figure 9B**). It is worth noting that the averaged CP does not correspond to the onset of scaling-divergent behavior for both responses (**Figure 4B**). In fact, the onset is extended: in the DMSO response, the scaling region exists up to the upper boundary of the near-critical state, while in the atRA response, the scaling region exits up to the upper boundary of the sub-critical state, where the averaged CP exists in the (averaged) sub-critical state. This is attributable to the collapse of autonomous bistable switch (ABS) of the sub-critical state as discussed in the previous section.

In summary, the mean-field behavior of bimodality coefficients reveals markedly different behaviors that distinguish averaged critical states.

Table 1: Averaged Critical States

Averaged Critical States	MCF-7 cells		HL-60 cells	
	N = 22277		N = 12625	
	HRG	EGF	DMSO	atRA
Super-critical	0.165<nrmsf	0.160<nrmsf	0.110<nrmsf	0.115 <nrmsf
	3051 mRNAs	1969 mRNAs	2582 mRNAs	2465 mRNAs
Near-critical	0.094<nrmsf<0.165	0.081<nrmsf<0.160	0.078<nrmsf<0.110	0.095<nrmsf<0.115
	6814 mRNAs	9119 mRNAs	2226 mRNAs	995 mRNAs
Sub-critical	nrmsf<0.094	nrmsf<0.081	nrmsf<0.078	nrmsf<0.095
	12412 mRNAs	11189 mRNAs	7817 mRNAs	9205 mRNAs

IV. Coherent-Stochastic Behavior (CSB) in Critical States

Critical states display coherent-stochastic behavior (CSB), where coherent behavior emerges in ensembles of stochastic expressions [11]. In **Figures 10 A,B**, random sampling of the averaged critical states for both MCF-7 and HL-60 cells clearly shows that

1) The near-zero Pearson correlation between different randomly selected gene ensembles in the critical states reveals stochastic expression, and

2) There is a sharp damping in variability (Euclidean distance of single time points from the center of mass, CM(t_j) of the critical states). This is a further confirmation that the CM(t_j) of the critical states represents their coherent dynamics.

The emergent CSB of critical states through SOC control of the entire expression shows how a population of cells can overcome the problem of stochastic fluctuation in local gene-by-gene regulation. Moreover, the fact that CM represents the coherent dynamics of CSB tells us how macroscopic control can be tuned by just a few hidden parameters through SOC. This

collective behavior emerges from intermingled processes involving the expression of more than twenty thousand of genes; this corresponds to the notion that, despite their apparent bewildering complexity, cell state/fate changes collapse to a few control parameters, this 'sloppiness' [9] derives from a low effective dimensionality in the control parameter-space emerging from the coherent behavior of microscopic-level elements.

Furthermore, mRNA expression in a microarray refers to populations of millions of cells, so that the ensemble of expression at $t = t_j$ represents a snapshot of time-dependent thermodynamic processes far from equilibrium, where the usual pillars of equilibrium thermodynamics such as 'time-reversibility' and 'detailed balance' break down, revealing the characteristics of a dissipative or far-from equilibrium system.

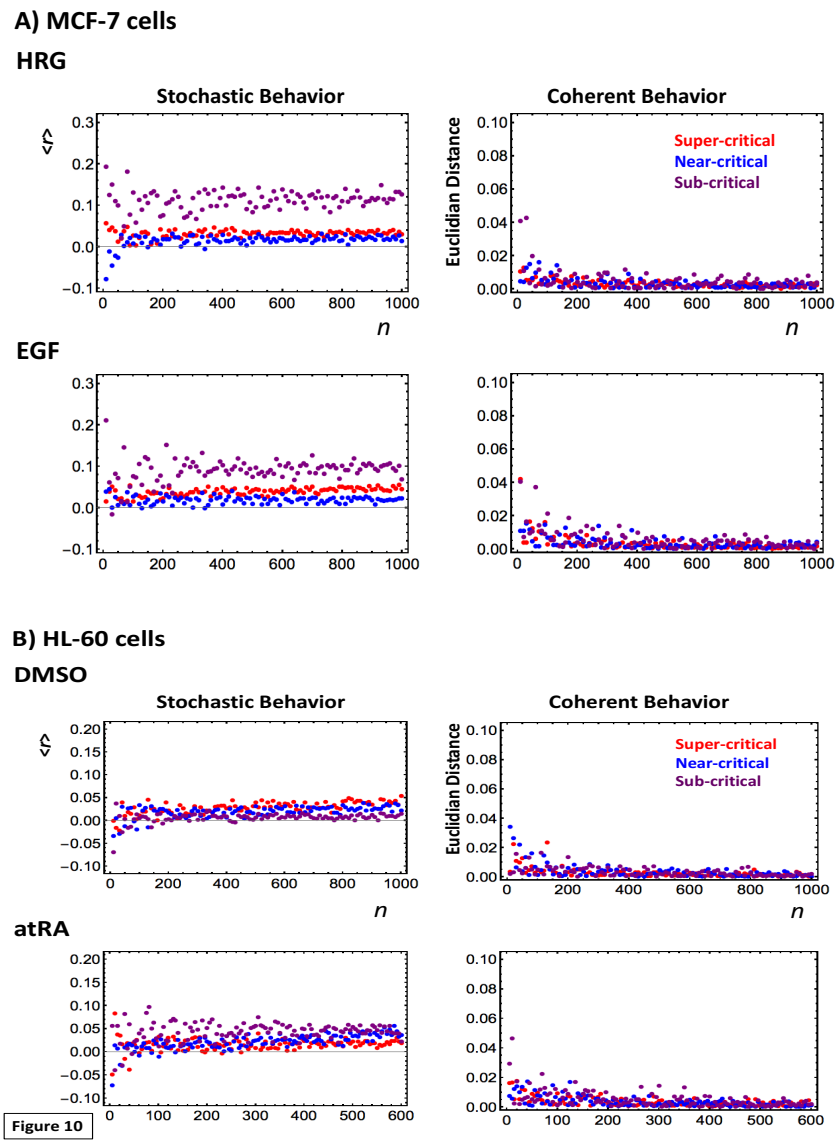


Figure 10

Figure 10: ***Coherent-stochastic behaviors in critical states:*** *A) MCF-7 cells and B) HL-60 cells. 200 random-number ensemble sets are created, where each set has n (variable) sorted numbers, which are randomly selected from an integer series {1,2,..,N} (N: total number of mRNAs in a critical state:* ***Table 2****). These random sets are used to create random gene ensembles from each critical state.*
*1) Left panels (A and B): For each n, Pearson correlations are evaluated between different random gene ensembles in the critical states by averaging over the 200 ensemble sets and experimental time points (****Methods****). This gives a near-zero Pearson correlation, consistent with the global stochastic character of microscopic transcriptional expression regulation in critical states.*
2) Right panels (A and B): For each random gene ensemble, the Euclidean distance of single time points from the center of mass $CM(t_j)$ of the critical states are evaluated by averaging over the 200 ensemble sets. The sharp damping of variability confirms that the emergent coherent dynamics of the critical states correspond to the dynamics of the $CM(t_j)$.

V. Sub-Critical State as a Generator of Self-Organizing Global Gene Regulation

The temporal development of critical transitions in overall expression suggests that molecular stressors on MCF-7 and HL-60 cells induce the perturbation of self-organization through interaction between critical states. The emergent coherent-stochastic behavior (CSB) corresponds to the dynamics of the center of mass (CM) of critical states. Hence, an understanding of the dynamics of the CM of critical states and their mutual interactions should provide insight into how the perturbation of self-organization in whole-mRNA expression evolves dynamically through perturbation.

Here, it would be useful to abstract the essence of the dynamics of critical states and their mutual interactions into a simple one-dimensional CM dynamical system: the CM of a critical state, $X(t_j)$, is a scalar point, and thus, the dynamics of $X(t_j)$ can be described in terms of the change in the one-dimensional effective force acting on the CM. From a thermodynamic point of view, this force produces work, and thus causes a change in the internal energy of critical states. Hence, we can investigate the genome as an "open-thermodynamic system" surrounded by the nuclear environment, which, through the exchange of 'genetic energy or activity', shows self-organized overall expression under environmental dynamic perturbations; the regulation of mRNA expression is managed through the mutual interaction between critical states and the external connection with the cell nucleus environment.

To quantitatively designate such flux flow, we set up the effective force acting on the CM, $f(X(t_j))$ at $t = t_j$, where each gene expression is assigned to have an equal constant mass (set to unity). The impulse, $F\Delta t$, corresponds to the change in momentum: ΔP is proportional to the change in average velocity: $v(t_{j+1})$ - $v(t_j)$. Since a consideration of the center of mass normalizes the number of genes being expressed in a critical state, we set the proportional constant, i.e., the mass of CM to be the unity. Thus, $f(X(t_j)) = -F = (v(t_j) - v(t_{j+1}))/\Delta t$, where $\Delta t = t_{j+1} - t_{j-1}$, and the force is given a negative sign, such that a linear term in the nonlinear dynamics of $X(t_j)$ corresponds to a force under a harmonic potential energy. Thus, $f(X(t_j))$ can be decomposed into IN flux: incoming flux from the past $t = t_{j-1}$ to the present $t = t_j$, and OUT flux: outgoing flux from the present $t = t_j$ to the future $t = t_{j+1}$:

$$f\left(X(t_j)\right) = \frac{1}{\Delta t}\left\{\frac{\left(X(t_j) - X(t_{j-1})\right)}{\Delta t_j} - \frac{\left(X(t_{j+1}) - X(t_j)\right)}{\Delta t_{j+1}}\right\}$$

$$= \text{IN flux} - \text{OUT flux}$$

We call the force, $f(X(t_j))$, the *net self-flux of a critical state*. The net self-flux: IN flux - OUT flux shows an incoming force (net IN self-flux) for a positive sign and an outgoing force (net OUT self-flux) for a negative sign.

When we adapt this concept of flux, it becomes straightforward to define the *interaction flux* of a critical state $X(t_j)$ with another critical state or the environment Y:

$$f\left(X(t_j); Y\right) = \frac{1}{\Delta t}\left\{\frac{\left(X(t_j) - Y(t_{j-1})\right)}{\Delta t_j} - \frac{\left(Y(t_{j+1}) - X(t_j)\right)}{\Delta t_{j+1}}\right\},$$

where, again, the first and second terms represent IN flux and OUT flux, respectively, and the net, 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. As noted, the interaction flux between critical states can be defined as when the number of expressions in a critical state is normalized, i.e., when we consider the CM. Due to the law of force, the net self-flux of a critical state is the summation of the interaction fluxes with other critical states and the environment (see Methods).

Next, we check how the net IN (OUT) flux of a critical state, the effective force acting on the CM of a critical state corresponds to the dynamics of its CM. **Figure 11** clearly shows that the trend of the dynamics of the CM of a critical state follows its net self-flux dynamics, in that the CM is up- (down-) regulated for net IN (OUT) flux, where the CM is measured from its

temporal average value. This implies that the respective temporal average values are the baselines for both flux and CM; this is further confirmed by the existence of an average flux balance in critical states, where the net average fluxes coming in and going out at each critical state are balanced (near-zero) (**Methods**).

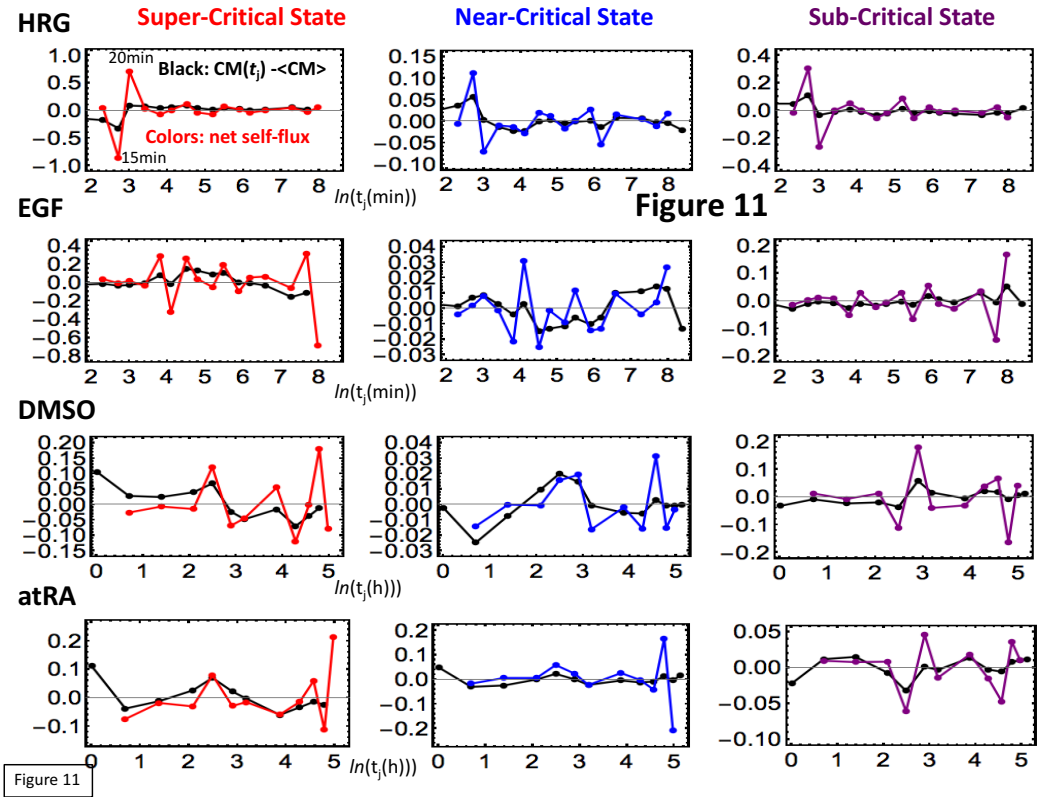


Figure 11: ***Dynamics of the center of mass (CM) of critical states compared with the net self-flux dynamics****: Colored lines (red: super-critical; blue: near-critical; purple: sub-critical state) represent net self-fluxes of critical states from their temporal averages (effective force acting on the CM:* ***Methods****). Black lines represent the dynamics of the CM of critical states from their temporal averages, which are increased three fold compared to the corresponding net self-fluxes. The plots show that the net self-flux dynamics follow up- (down-) regulated CM dynamics, such that the sign of the net self-flux (i.e., IN and OUT) corresponds to activation (up-regulated flux) for positive responses and deactivation (down-regulated flux) for negative responses. The natural log of experimental time points (MCF-7: minutes and HL-60: hours) is shown.*

Thus, we consider both temporal averaged expression flux among critical states and the fluctuation of expression flux from the average (*flux dynamics*), so that *averaged expression flux shows a temporal average energy flow among critical states through the environment, i.e., the characteristics of an open-thermodynamic genomic system* ("genome engine" in **Discussion**), *and flux dynamics represents fluctuation of the expression flux flow that is markedly different*

from the basic properties in equilibrium Brownian behavior under a detailed balance. Furthermore, the sign of the net self-flux (i.e., IN or OUT) corresponds to the activation (up-regulation) of flux for positive responses and deactivation (down-regulation) for negative responses.

We examined the average flux network for the processes of MCF-7 and HL-60 cells. **Figure 12A** and **Table 2** intriguingly reveal four distinct processes that share a common feature in their flux networks:

i). **A dominant cyclic flux between super- and sub-critical states:** Average net IN and OUT flux flows reveal how the internal interaction of critical states and external interaction with the cell nucleus environment interact: notably, the formation of robust average cyclic state-flux between super- and sub-critical states through the environment forms a dominant flux flow in the genomic system. This formation of the cyclic flux is a cause of strong coupling between the super- and sub-critical states revealed through the correlation analysis [11].

ii). **Sub-critical state as a source of internal fluxes through a dominant cyclic flux**: The direction of the average flux in super- and sub-critical states reveals that the sub-critical state is the source of average flux, which distributes flux from the nucleus environment to the rest of the critical states, where the super-critical state is the sink to receive fluxes from the near- and sub-critical states. Importantly, this clearly shows that the sub-critical state, an ensemble of low-variance expression act as a generator of perturbation in the genome-wide phase transition.

In summary, the results regarding average flux reveal the roles of critical states in SOC control of overall expression, and provide a mechanistic picture of how global self-organization emerges through the interaction among critical states and the cell nucleus microenvironment.

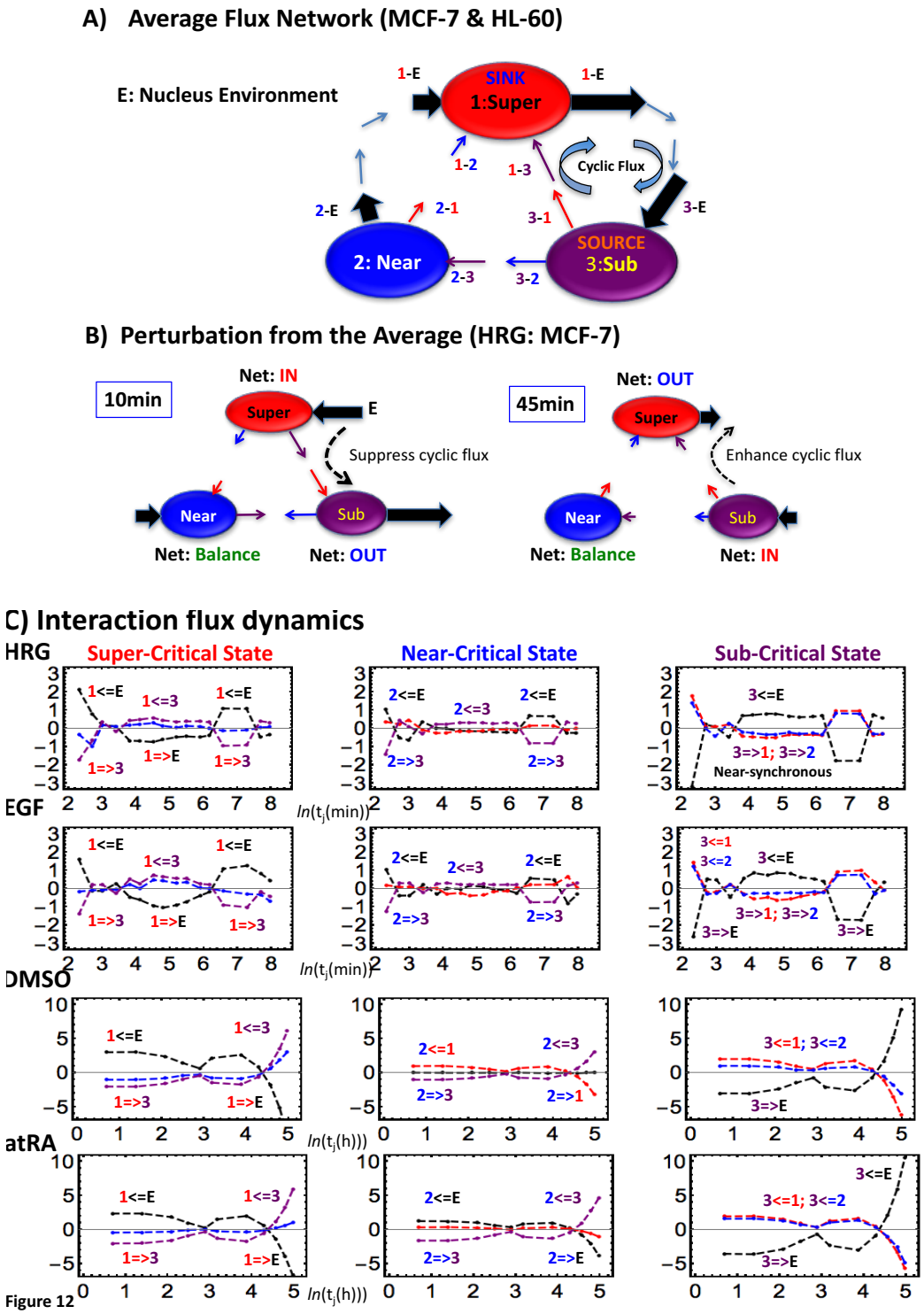


Figure 12: ***Genomic expression dynamics revealed through a flux analysis including crosstalk with the environment:***

A) Average values of interaction flux (colored arrows) for MCF-7 and HL-60 cells show that a sub-critical state acts as an internal 'source', where IN flux from the environment 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 (**Methods**). Furthermore, the formation of a dominant cyclic state-flux is revealed between super- and sub-critical states through the environment. The average interaction flux is represented as i-j: interaction flux of the ith critical state with the jth critical state (i, j= 1: super-; 2: near-; 3: sub-critical), and a colored arrow (color based on the jth critical state) points in the direction of interaction with a relative amount of flux (see details in **Table 2**). E represents the internal nucleus environment.*

B) *An early flux dynamics event in the HRG response resulted from the HRG interaction flux dynamics (see C) is shown. Interaction flux dynamics i<=j (i=>j; color based on j) represents the interaction flux coming from the jth critical state to the ith critical state or vice versa (the flux direction changes at y =0). The interaction fluxes (see HRG in C) align to suppress the cyclic state flux at 10min (the first point), where the interaction flux shows 1<=E, 1=>2, and 1=>3 at the super-critical state, 2<=E, 2<=1, 2=>3 at the near-critical state, and 3=>E, 3<=1, 3=>2 at the sub-critical state. Then, they align to enhance the cyclic state flux at 45 min (5th point). This change in the dynamic flux structure is due to the global perturbation at 15-20min (see **section VI**). At each node, the net flux (**Figure 11**; 10min: first point; 45min: 5th point) is indicated - IN for net incoming flux (y >0); OUT for outgoing (y <0); Balance (y~0). Note: the average flux balance at each node is maintained, but not at individual time points.*

C) *Notably, for MCF-7 cell fates (HRG and EGF), near-synchronous interaction flux dynamics at sub-critical states are seen. The plots further show that the overall patterns are similar between the same cell types, which again supports cell-type-specific SOC control.*

Table 2: ***Average Interaction Fluxes:*** *arrow shows the direction of flux.*

Super-Critical SINK	**MCF-7 HRG**	**MCF-7 EGF**	**HL-60 DMSO**	**HL-60 atRA**
Super **1-2** ← Near	0.3	0.2	1.3	0.5
Super **1-3** ← Sub	1.9	1.6	2.6	2.5
Super **1-E** → EN	-2.2: 1.3(Near) -3.5(Sub)	-1.8: 1.2 -3.0	-3.9: 0.0 3.9	-3.0: 1.5 -4.5
Near-Critical				
Near **2-1** → Super	-0.3	-0.2	-1.3	-0.5
Near **2-3** ← Sub	1.6	1.4	1.3	2.0
Near **2-E** → EN	-1.3	-1.2	0.0	-1.5
Sub-Critical SOURCE				
Sub **3-1** → Super	-1.9	-1.6	-2.6	-2.5
Sub **3-2** → Near	-1.6	-1.4	-1.3	-2.0
Sub **3-E** ←← EN	3.5	3.0	3.9	4.5

VI. Dynamic Mechanism of SOC Control in the Overall Expression

The flux dynamics at critical states reveal the dynamic mechanism of SOC control in overall expression to elucidate how dynamic interaction among critical states and the environment perturb the average flux flow in the genomic system, as follows:

1) Flux dynamics reveal early nucleus activities:

In **Figures 12B,C**, the flux dynamics of CMs in HRG stimulated MCF-7 cell show how the average flux network (**Figure 12A**) is perturbed substantially at early time points, which reflects the early occurrence of significant genetic energy dissipation. This can be seen from the net self-flux dynamics (HRG response; **Figure 11**). At 15-20min, *the global perturbation* involving a large change in net self-flux at more than one critical state occurs: the net self-flux of the super-critical state shows a pulse-like change from net OUT (negative value) to IN self-flux (positive), i.e., an increase in internal energy, which explains how the significant net flux into the super-critical state is used to activate the expression of genes in the super-critical state from 15 to 20min. In contrast, for the sub- and near-critical states, their net self-fluxes are significantly changed from the net IN to net OUT self-fluxes (anti-phase with respect to the dynamics of the super-critical state), i.e., they show a loss of internal energy.

The global perturbation at 15-20 min stems from the occurrence of the genetic energy flow in the genomic system, which shows a change from the strong suppression of cyclic flux at 10min before perturbation to enhancement at 45min after perturbation (**Figure 12B**): at 10min, the flow of the interaction fluxes between the super- and sub-critical states aligns against the average cyclic flux to suppress the cyclic flux (the strongest inhibition over time), and then, at 45 min, the interaction fluxes change, and align to enhance the cyclic flux.

These results suggest the presence of early nuclear activities in HRG stimulated MCF-cells. At 15min, genetic information, through signaling activities [21,22] in the cytoplasm from the cell membrane induced by HRG, reaches the nucleus, and at 20min it activates the high-variance genes of the super-critical state. In contrast, in the near- and sub-critical states (intermediate- and low-variance genes, respectively), their expressions are suppressed, so that their internal genetic energy flow into the environment should induce a change in the physical

plasticity of chromatin structures of genes in these states, i.e., less pliable structures at the ensemble scale (not at an individual scale).

2) Synchronization of interaction flux dynamics occurs in sub-critical states:

Intriguingly, synchronizing interaction flux dynamics in sub-critical states are observed in the responses of MCF-7 cells (**Figures 12C**), which is also confirmed by the (near) bisecting average interaction fluxes from sub-critical states to super- and near-critical states (**Table 2**). This (near) synchronization of interaction flux at sub-critical state results in the fact that the net self-flux of super-critical states, i.e., the genetic energy activities for the gene ensemble with the greatest variation, stems from their interaction with near-critical states, where a critical transition occurs. This tells us that the collective behavior in the super-critical state has a direct impact on the critical transition and vice versa. The synchronization of interaction in sub-critical states makes the net self-flux dynamics of super-critical states well-approximated by the interaction flux of the super-critical state with the near-critical state (**Supplementary Figure S2**). Interaction dynamic fluxes are almost balanced between the sub-critical state and the environment (see **Methods**).

3) Global and local perturbations exist in the SOC control:

So far, we have applied the expression flux concept to the effective force acting on the CM of a critical state. This concept can be also extended to define kinetic energy self-flux on the CM of a critical state. The kinetic energy of the CM with unit mass at t = t_j is defined as $1/2 \cdot v(t_j)^2$, such that the net kinetic energy self-flux can be defined as

$$K\left(X(t_j)\right) = \frac{1}{2}\left\{\left(\frac{\left(X(t_j) - X(t_{j-1})\right)}{\Delta t_j}\right)^2 - \left(\frac{\left(X(t_{j+1}) - X(t_j)\right)}{\Delta t_{j+1}}\right)^2\right\}$$

$$= \text{IN kinetic engery flux} - \text{OUT kinetic engery flux}$$

We investigate when a significant kinetic energy self-flux occurs to better understand global perturbation of the SOC control. **Figure 13** shows that global perturbations are more evident in the net kinetic energy self-flux dynamics than in the effective force (**Figure 11**):

i) MCF-7: Global perturbation involving a change in net kinetic flux in critical states occurs only at 10-30min, where a pulse-like transition occurs from outgoing kinetic

energy flux to incoming flux at 15-20min. The occurrence of such transition is confirmed by a pulse like maximal response in Pearson autocorrelation ($P(t_j;t_{j+1})$ at 15-20min: see Figure 4A in [11]). Dissipation of the kinetic energy is evident in the HRG response, which quickly ends at 30min. In contrast, in the EGF response, only vivid flux activity in the super-critical state is apparent until t = 36h, which exhibits *local perturbation* in the EGF response. Thus, global and local perturbations differentiate the cell fate between the responses to HRG and EGF: the dissipative pulse-like global perturbation in the HRG response at 15-20min derives the genome state change at 3h (**Figure 5A**), whereas the local perturbation in the EGF response does not induce the state change (see **section I**). Furthermore, this global perturbation (see also paragraph 1 above) suggests *the existence of a novel primary signaling transduction mechanism or a biophysical mechanism which can induce the deactivation of gene expression in the near- and sub-critical states (a majority of expression: mostly low-variance gene expression) within a very short time.* This global deactivation mechanism that is associated with more than ten thousands of low-invariance mRNAs in a coordinated manner should be causally related to the corresponding higher-order chromatin structural changes (as discussed in the literature from a theoretical perspective [24-26] and a biological perspective [27,28]).

ii) HL-60: In the response to DMSO, a clear pulse-like global perturbation occurs at 12-18h. In the response to atRA, the first significant global perturbation occurs at an early time (2-4h) and a second smaller global perturbation occurs at 12-18h, which can be confirmed by a change in the effective force (**Figure 11**). They both show distinct dissipative oscillatory behaviors of kinetic energy flux dynamics. Again, these global perturbations occur before the genome-state changes (24h for DMSO and 48h for atRA; **Figure 5B**).

4) Long-term global mRNA oscillation underlies SOC control: The sub-critical state is the source of internal genetic energy flow in SOC control, and therefore, the oscillatory net self-flux of sub-critical states generates a long-term global mRNA oscillation to guide the perturbation of SOC control.

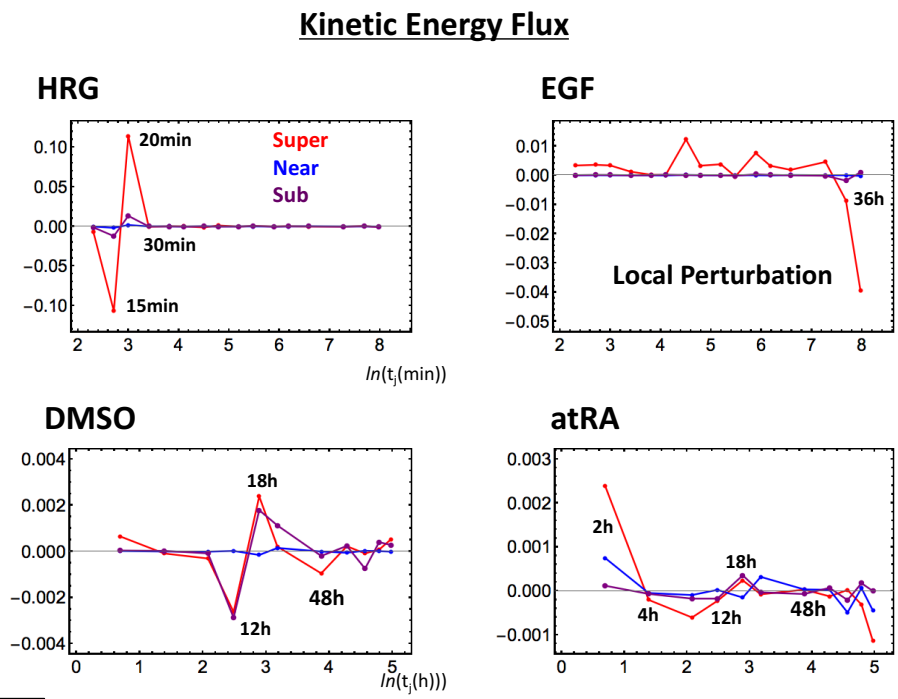


Figure 13

Figure 13: ***Local and global perturbations in self-organizing genome-wide expression:*** *The kinetic energy self-flux dynamics (y-axis) for the CM of a critical state (see* ***section VI****) reveals clear energy-dissipative behavior. Notably, the results show the occurrence of global and local perturbations of self-organization. Pulse-like global perturbations show a transition from IN to OUT net energy flux or vice versa (IN-OUT switching) in more than one critical state: at 15-20min, HRG; at 12-18h, DMSO; and at 2-4h (significant) and 12-18h, atRA. In contrast, local perturbation is revealed in the EGF response for up to 36h: there is only a marked response in the super-critical state, and almost no response in the other states (i.e., the dynamics of CM of critical states are localized around their average:* ***Figure 11****). The results suggest that the global and local perturbations differentiate MCF-7 cell fates, whereas global perturbations drive the state change in HL-60 cells (see* ***section I****).*

Discussion

We investigated the dynamics of collective gene behavior in several biological processes for changes in the cell fate:

1) Early embryo development in human and mouse,

2) Helper T 17 cell differentiation, induction of terminal differentiation in human leukemia HL60 cells by DMSO and atRA,
3) Activation of ErbB receptors in human breast cancer MCF-7 cells by epidermal growth factor (EGF) and heregulin (HRG).

Our approach builds upon the analysis of transcriptional expression of the gene ensembles ordered by the normalized amount of change in time (*nrmsf*) and by the fold change in expression.

In all of the models analyzed, despite temporally different experimental intervals, a self-organized critical transition (SOC) in whole-genome expression plays an essential role in the change of the genome state at both population and single cell levels. The results suggest that the full or partial erasure of the initial-state sandpile-type critical behavior can be an indicator of the genome-state change (see the summary in **Table 3**). Notably, regarding to the embryo development, the event of reprograming initiation, i.e., whether or not a single cell successfully achieves reprogramming can be determined by the erasure of sandpile-type critical point (CP) stemming from the initial stage of embryogenesis. Thus, it is highly probable that genes around the sandpile-type CP of the initial state play an essential role in reprogramming.

Table 3: ***Genome-State Change Through Distinct Erasures of Initial-State Critical Behavior:*** *Averaging behaviors according to the fold change in expression (t_0 and t_j or zygote state and early embryo single-cell state).*

EGF-stimulated MCF-7 cells	HRG-stimulated MCF-7 cells	atRA- and DMSO-stimulated HL-60 cells	Mouse and Human Embryos & Th17 cell
Sandpile Type Critical Behavior: Divergent UP- and DOWN-Regulations	Absence of Divergent Behavior in UP-Regulation	No Critical Behaviors: Non-Stochastic Type	No Critical Behaviors: Stochastic Type
NO State Change Sandpile type CP DOWN-Regulation UP-Regulation Figure 5A	State Change at 3h NO Divergent Behavior Figure 5A	State Change at 24h (DMSO) and 48h(atRA) No CP NO Divergent Behaviors Figure 5B	Reprogramming: Late 2-cell (Mouse) After 8-cell (Human) 6h (Th17) No CP Figure 7
Cell Population	Cell Population	Cell Population	Single cell

The whole-genome expression self-organizes into distinct critical states (expression compartments) through a critical transition. Their coherent dynamics emerge from an ensemble of stochastic expression (coherent-stochastic behaviors) corresponding to the one-dimensional dynamics of the center of mass (CM) of expression, which revealed the characteristics of open thermodynamic system and the dynamic perturbation of SOC control for the genome-state change:

i) Average expression flux of critical states (**Figure 12A**) clearly showed that for MCF-7 and HL-60 cells, the collective behavior of low-variance gene ensembles acted as a driving force to transmit their potentiality, or energy of coherent transcription fluctuations, to high-variance genes. This can be interpreted by the metaphor of a 'genome engine':

Sub-critical states, which are an ensemble of low-variance mRNAs, act as a large piston that moves only slightly. This work propagates through cyclic flux (like a camshaft) to make the super-critical state, an ensemble of high-variance mRNAs, activated as a small piston that moves greatly, while remaining anti-phase to the dynamics of the sub-critical state. The genome engine emerges only when thousands of molecular gene regulations are integrated through the SOC control.

ii) Flux dynamics (expression flux dynamics from their averages) revealed that SOC control (i.e., the genome engine) is perturbed locally or globally for the determination of the genome-state change. In cell differentiation (HRG: MCF-7 cells, and DMSO & atRA: HL-60 cells), the genome-state changes occur at the end of the dissipation of global pulse-like perturbations, which involves the overall expression through the relevant responses of other critical states in addition to the super-critical state. In contrast, in EGF-stimulated MCF-7 cells (cell proliferation [20]), no genome-state change can occur due to the local perturbation: only super-critical state is affected, while other states show only very weak responses (i.e., staying around their average expression fluxes).

Thus, the global and local perturbations of SOC control differentiate MCF-7 cell fates, whereas global perturbations underlie the state change in HL-60 cells.

These results give us a mechanism of the genome-state change in terminal cell fates: the (partial or full) erasure of an initial-state critical behavior leads to a critical change in the genome state at the end of a dissipative pulse-like global perturbation in self-organization. Regarding

early embryo development, we need further studies to confirm this cell fate regulating mechanism.

Notably, an analysis of the literature strikingly confirms the coincidence of the time intervals of the observed thermodynamic changes revealed by a flux approach to transcriptome data, with real biological critical events determining the change in cell fate. The embryo model is particularly intriguing for our purpose because, in contrast to both HL60 and MCF-7, it is not based on the average behavior of a huge population, but rather on the behavior of a very few cells at a time.

Some consequences can be envisaged from the above findings:

i) Biological Interpretations for Global and Local Perturbations of SOC Control in MCF-7 Cells

Regarding our present finding of the global perturbation in the SOC control in MCF-7-HRG dynamics, the study of early gene response performed by [21] established that EGF and HRG induced a transient and sustained dose-dependent phosphorylation of ErbB receptors, respectively, followed by similar, transient and sustained activation kinetics of Akt and ERK.

Following ERK and Akt activation from 5-10 min, the ligand-oriented biphasic induction of early transcription key proteins of AP-1 complex (c-FOS, c-MYC, c-JUN, and FRA-1), high for HRG and negligible for ERG, took place. The proteins of AP-1 complex are the non-specific stress responders supported by activating phosphorylation of ERK in a positive feedback loop. In addition, the key reprogramming transcription factor c-MYC (whose protein peaking at 60 min was confirmed in JE laboratory) has the ability to amplify the transcription of thousands of active and initiated genes [29] or direct-indirect targets [30] and modify chromatin by recruiting histone acetyltransferase [31]. Further Saeki et al [32] revealed that the early sustained by ERK activation of AP-1, c-FOS in particular, induces after HRG the sequential activation of late transcription factors EGR4, FOSL-1, FHL2 and DIPA, the last peaking at 3h (DIPA is activating fatty acids and lipid metabolism genes). In turn, this step began to down-regulate ERK proliferative pathway by a negative feedback loop.

Thus, so far, the continuity of biological relay of the HRG-induced early (pre-committing) and late transcription activities leads to commitment of differentiation from 3h (cell

fate change), which is necessarily coupled to suppression of proliferation and stops the genome boost by ERK pathway. Both effects are not achieved in the ERG case.

In the corresponding expression data, we observed after HRG the powerful genome engine causing a pulse like global perturbation (15-20min) as the pre-committing and the erasure of the initial state critical transition inducing the genome-state change at 3h, which was absent after treatment with EGF, where only local perturbation (i.e., the only vivid activity of the super-critical state) was observed.

It is important to note that the sub-critical state (low variance expression: a majority of mRNA expressions; **Table 1**) is the generator of the global perturbation in the HRG response (**section V**). As discussed above, the dynamic control of gene expression in the sub-critical state is expected from cooperative ensemble behaviors of genomic DNA structural phase transitions through interaction with environmental molecules, which has been neglected from biological studies. Thus, a true biological picture for MCF-7 cells may come from deciphering biological functions of genes in the sub-critical state in a coordinated (not individual) manner.

ii) Global Perturbations of SOC Control in HL-60 Cells Committed to Differentiation

The general mechanisms of the commitment to differentiation are not yet well understood. Developmental biologists usually discriminate the two phases into: (1) reversible, with the capability of autonomous differentiation; and (2) essentially irreversible [33]. A study in an HL60-DMSO cell model [34] found that a minimum induction time of 12h was needed for cells to commit to differentiation. In turn, Tsiftsoglou et al. [35] found that exposure to differentiation inducers for only 8 to 18h, which is much shorter than the duration of a single generation, is needed to provide commitment for autonomous terminal differentiation.

Consistent with these findings, we revealed that, at 12-18h, global perturbation involving critical states is observed for the responses to both DMSO and atRA. The induction of differentiation for both inducers is different in the sense that, in contrast to DMSO, which induces development of macrophages/monocytes, RA treatment leads to segmented neutrophils [35]. Interestingly, our analysis also shows that the achievement of cell fate determination at 24h for DMSO and at 48h for atRA (**Figure 5B**) occurs in these two models in different ways, although they converge at the same final equilibrium state at 48h.

In particular, we observed the early global genome perturbation in the response to atRA (at 0-4h), absent in DMSO model. It might be induced by the difference in Ca^{2+} influx. Calcium release from the ER combined with capacitative calcium influx from the extracellular space leads to markedly increased cytosolic calcium levels and is involved in cell activation [36] to control key cell fate processes, which includes fertilization, early embryogenesis [37], and differentiation [38]. The Ca^{2+} response amplitude and duration are decoded by downstream effectors to specify cell fates [39]. Yen et al. [40] and others have shown that the pre-committed HL-60 cells display the early cytosolic Ca2+ influx. Moreover, intracellular calcium pump expression is modulated during myeloid differentiation of HL-60 in a lineage-specific manner, with higher actual flux in the atRA response [41].

iii) SOC Control in Human and Mouse Related to the Developmental Oocyte-to-Embryo Transition

Fertilized mature oocytes change their state to become developing embryos. This process implies a global restructuring of gene expression. The transition period is dependent on the switch from the use of maternally prepared stable gene transcripts to the initiation of proper embryonic genome transcription activity. It has been firmly established that, in mice, the major embryo genome activation occurs at the two-cell stage (precisely between mid and late 2-cell stage), while in humans this change occurs between the 4- and 8-cell stages [42]. We detected these time intervals, which differ for mouse and human, in the development of SOC control: from sandpile-type critical transitions to stochastic expression distributions (**Figures 7A,B**). Reprogramming of the genome destroys the SOC control of the initial stage of embryogenesis. Developmental studies by Wennekamp et al. in Hiiragi's group [43] revealed the onset of symmetry breaking between cells in the early embryo and consequently the specification of distinct cell lineages strictly consistent with our model. In addition, what about the physical state of chromatin during these developmental steps?

The erasure of paternal imprinting by DNA 5-methylcytosine de-methylation and hydroxymethylation as part of epigenetic reprogramming occurs in the embryo, which allows significant decompaction of the repressive heterochromatin and an increase in the flexibility of

transcribing part of the chromatin [44]. Detailed studies of the DNA methylation landscape by Guo and colleagues in human embryo [23] revealed a large decrease in the level of methylation of gene promoter regions from the zygote to the 2-cell stage, which would erase oocyte imprinting. The strength of this correlation increases gradually until it becomes particularly strong after human embryonic genome activation (full reprogramming) occurring at the 8-cell stage [45]. DNA de-methylation unpacks repressive heterochromatin, which manifests in dispersal and spatio-temporal reorganization of pericentric heterochromatin found as an important step of embryonic genome reprogramming [46]. This synchronization of the methylome to confer maximum physical decompaction and flexibility to chromatin and the reprogramming of transcriptome activity for totipotency, support the feasibility of SOC, which was determined here by an independent method.

In addition, transposable elements, which are usually nested and epigenetically silenced in the regions of hypermethylated constitutive heterochromatin, also become temporarily activated during the oocyte-to embryo transition in early embryogenesis [47]. In turn, in human, a peak in SINE activation coincides with the 8-cell stage [23]. The significance of transient retrotransposon activation in embryogenesis, as has been suggested [44], may be that the thousands of endogenous retro-elements in the mouse genome provide potential scope for large-scale coordinated epigenetic fluctuations (further harnessed by piRNA along with *de novo* DNA methylation). In other words, this transient burst should create a necessary critical level of transcriptional noise as a thermodynamic prerequisite for the non-linear genome expression transition using SOC.

iv) 'Genome Engine' Mechanism in Self-Organization and Genome Computing

In the genome engine, the role of the sub-critical state as a generator leads to a new hypothesis that the genomic compartment, spanning from kbp to Mbp, which produces low-variance gene expression can be the mechanical material basis for the generator of global perturbation, where the coordinated ensemble behavior of genomic DNA structural phase transitions through interaction with environmental molecules plays a dominant role in the expression dynamics [48,49]. In HRG-stimulated MCF-7 cell differentiation, a subset of

consecutive genes pertaining to the same critical state (called barcode genes from kbp to Mbp; see Figure 8A in [11]) on chromosomes has been shown to be a suitable material basis for the coordination of phase-transitional behaviors. The critical transition of barcode genes was shown to follow the sandpile model as well as genome avalanche behavior. This indicated that there is a non-trivial similarity through SOC between the coherent-stochastic network of genomic DNA transitions and the on-off nerve firing in neuronal networks [50]. Thus, a potential function of the genome engine may be asynchronous parallel computing, and coherent-stochastic networks based on the on/off switching of sub-critical barcode genes may act as rewritable self-organized memory in genome computing.

Here we highlighted how the genome engine, i.e., global response is driven by the oscillatory behavior of sub-critical state, i.e., of genes whose change in expression is quantitatively minor but strongly coherent. The coherence appears when taking into consideration ensembles of genes with n (number of genes) greater than 50 [11]; this behavior is consistent with the particular organization of chromatin into the topology-associated domains [16-20]. The relative flexibility of each domain probably goes hand-in-hand with the changes in expression.

The demonstration of a 'genetic energy flux dynamics' across different critical states tells us that chromatin is traversed by coherent waves of condensation/de-condensation, analogous to the allosteric signals in protein molecules. The possibility of controlling such signal transmission through the control of the higher-order structure of genomic DNA raises the possibility of very intriguing applications, along the same lines as in the much more mature case of allosteric drugs [51,52].

Methods

Biological Data Sets

We analyzed mammalian transcriptome experimental data for seven distinct cell fates in different tissues:

i) Cell population: Microarray data of the activation of ErbB receptor ligands in human breast cancer MCF-7 cells by EGF and HRG; Gene Expression Omnibus (GEO) ID: GSE13009 (N = 22277 mRNAs; experimental details in [32]), which have 18 time points: $t_0 = 0$, $t_1 = 10$, 15, 20, 30, 45, 60, 90min, 2, 3, 4, 6, 8, 12, 24, 36, 48, $t_{T=17} = 72$h,

ii) Cell population: Microarray data of the induction of terminal differentiation in human leukemia HL60 cells by DMSO and atRA; GEO ID: GSE14500 (N = 12625 mRNAs; details in [53]), which have 13 time points: $t_0 = 0$, $t_1 = 2$, 4, 8, 12, 18, 24, 48, 72, 96, 120, 144, $t_{T=12} = 168$h,

iii) Single cell: RNA-Seq data of T helper 17 cell differentiation from mouse naive CD4+ T cells in RPKM values (Reads Per Kilobase Mapped), where Th17 cells are cultured with anti-IL4, anti-IFNγ, IL-6 and TGF-β (details in [54]); GEO ID: GSE40918 (mouse: N = 22281 RNAs), which has 9 time points: $t_0 = 0$, $t_1 = 1,3,6,9,12,16,24$, $t_{T=8} = 48$h,

iv) Single cell: RNA-Seq data of early embryonic development in human and mouse developmental stages in RPKM values; GEO ID: GSE36552 (human: N = 20286 RNAs) and GSE45719 (mouse: N = 22957 RNAs) (experimental details in [55]) and [56], respectively).

We analyzed 7 human and 10 mouse embryonic developmental stages:

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.

As for microarray data, the Robust Multichip Average (RMA) was used to normalize expression data for further background adjustment and to reduce false positives [57-59], whereas for RNA-Seq data, RNAs that have zero RPKM values over all of the time points were excluded. Random

real numbers in the interval [0-1] generated from a uniform distribution are added to all expression values for the natural logarithm. This procedure avoids the divergence from zero of RPKM values. The robust mean-field behavior through the grouping of expression (see **section II** or **Figures 7B-D, 8**) was checked by multiplying the random number by a positive constant, a (a< 10), and thus, we set a = 1. Note: the addition of large random noise (a>>10) destroys the sandpile CP.

Emergent Properties of SOC in Mean-Field Approach

In this report, we examine whether characteristic behavior at a critical point (CP) is presented in the overall expression for the examination of the occurrence of SOC in various cell fates. We briefly summarize below how SOC in overall gene expression was elucidated in our previous studies [10,11].

i) **Global and local genetic responses through mean-field approaches**: Our approach was based on an analysis of transcriptome data by means of the grouping (gene ensembles) of gene expression (mean-field approach) characterized by the amount of change in time to reveal the coexistence of local and global gene regulations in overall gene expression [60,61]. Global expression response emerges in collective behavior of low and intermediate variance gene expression, which is cut off by an artificial threshold from the whole gene expression in many expression studies, whereas a local response represents genetic activity of high variance gene expression elucidated in molecular biology.

ii) **Self-organized criticality as an organizing principle of genome expression**: To understand a fundamental mechanism/principle for the robust coexistence of global and local gene regulations, and further, the role of global gene expression response, we elucidated self-organized criticality (SOC) principle of genome expression able to conduct global gene regulations [11]. In self-organization, temporal variance of expression, *nrmsf* (normalized root mean square root) plays an order parameter to self-organize the whole gene expression into distinct expression domains (distinct expression profiles) defined as *critical states*, where *nrmsf* is defined by dividing *rmsf* by the maximum of overall $\{rmsf_i\}$:

$$rmsf_i = \sqrt{\frac{1}{T+1}\sum_{j=0}^{T}\left(\varepsilon_i(t_j) - \langle\varepsilon_i\rangle\right)^2},$$

where $rmsf_i$ is the *rmsf* value of the ith expression (mRNA or RNA), $\varepsilon_i(t_j)$, which is expressed as $\varepsilon_i(t_j)$ at $t = t_j$ and $<\varepsilon_i>$ is its average expression value.

iii) **Coherent-stochastic behaviors in critical states**: Coherent expression in each critical state emerges when the number of stochastic expressions is more than 50 expressions [12] (called coherent-stochastic behavior: CSB). The bifurcation - annihilation event of ensemble of coherent gene expression determine the boundary of critical states [10].

iv) **Biophysical reason for specific groupings**: Coherent dynamics such as coherent oscillation is emerged in the expression change (e.g., fold change) (refer to Figures 7A, B in [10]), but not in expression. Furthermore, the amplification in the expression change in a critical state is exhibited, but not in expression. This indicates a highly correlative behavior for expression and stochastic resonance effect for the expression change [11]. Thus, regarding the self-organization, we investigated averaging behaviors in *nrmsf* (order parameter) and change in expression.

v) **Characteristic properties of SOC**: Distinguished critical behaviors at a critical point, are emerged in averaging behaviors of two observables, the fold change in expression and *nrmsf*. Importantly, different critical behaviors (i.e., criticality) occur at the same CP originated from different averaging behaviors (MCF-7 cells), which confirms that the mean-fields are not an artifact. We call them *sandpile type transitional behavior and scaling-divergent behaviors at a critical point* (CP) summarized as follows:

a. **Sandpile type transitional behavior** based on the grouping of expression into k groups with an equal number of n elements according to the fold change in expression. A sandplie-type transition is the most common in SOC [62]. As n is increased, average value of a group (a mean-field) converges, and an ensemble of averages emerges a sandpile profile. Good convergence in the group is obtained above $n = 50$ (refer to Figure 2 in [11]), which stems from CBS. The top of the sandpile is a critical point (CP), where the CP usually exists at around a zero-fold change (null change in expression), which up- and down-regulated expression is balanced between different time points. This indicates that the critical behavior occurs through a 'flattened expression energy profile'. As noted, a critical point can

exist away from a zero-fold change. A sandpile type critical behavior shows that as the distance from the CP (summit of sandplie) is increased, two different regulatory behaviors, one representing up-regulation and another, down-regulation, are diverging. Furthermore, at the vicinity of the CP according to *nrmsf*, in terms of coherent expression, self-similar bifurcation event of the overall expression occurs to show the transitional behavior. Thus, as a critical behavior and critical transition occurs at the CP, we can characterize it as a *sandpile type transition*.

b. **Scaling-divergent behavior (genomic avalanche)** based on the grouping of expression according to *nrmsf*: A nonlinear correlation trend between the ensemble averages of *nrmsf* and mRNA expression at a fixed time point. Originally, we called this the DEAB (*Dynamic Emergent Averaging Behaviors*) of expression [10], which has linear (scaling) and divergent domains in a log-log plot. In the scaling domain, the quantitative relationship between the ensemble averages of *nrmsf* and mRNA expression, $\langle nrmsf \rangle$ and $\langle \varepsilon \rangle$, is shown in terms of power law scaling behavior, where higher $\langle nrmsf \rangle$ corresponds to higher $\langle \varepsilon \rangle$:

$$1 - \langle nrmsf \rangle = a\langle \varepsilon \rangle^{-b}.$$

Such scaling is lost at the CP in the MCF-7 cell fates. This shows that order (scaling) and disorder (divergence) are balanced at the CP, which presents genomic avalanches. In **Supplementary S1 file**, we address i) the genuineness of the power law scaling and ii) the existence of collective behavior of gene expression in the power-law scaling. Chromosomes exhibit a fractal organization; thus, the power law behavior may reveal a quantitative relation between the aggregation state of chromatin through *nrmsf* and the average expression of ensemble of genes. The entity of gene expression likely scales with the topology-associated chromatin domains (TAD) [16-20], such that the degree of *nrmsf* should be related to the physical plasticity of genomic DNA and the high order chromatin structure.

Dynamic Flux Analysis for an Open-Thermodynamic Genomic System

Sloppiness in SOC control emerges to reveal that genome expression is self-organized into a few critical states through a critical transition. The emergent coherent-stochastic behavior (CSB) in a critical state corresponds to the scalar dynamics of its center of mass (CM), $X(t_j)$, where $X \in \{\text{Super}, \text{Near}, \text{Sub}\}$: Super, Near and Sub represent the corresponding critical states. Thus, the dynamics of $X(t_j)$ are determined by the change in the one-dimensional effective force acting on the CM. We consider the effective force as a net flux of incoming flux from the past to the present and outgoing flux from the present to the future. Based on this concept of flux, it becomes possible to evaluate the dynamical change in the genetic system in terms of flux among critical states through the environment.

1. **Self-flux**: The effective force is deduced as the decomposition of IN flux from the past (t_{j-1}) to the current time (t_j) and OUT flux from the current time (t_j) to a future time (t_{j+1}), so that the effective force as a net self-flux, $f\left(X(t_j)\right)$ can be written as

$$f\left(X(t_j)\right) = \frac{\Delta P}{\Delta t} = \frac{1}{\Delta t}\left\{\frac{\left(X(t_j)-X(t_{j-1})\right)}{\Delta t_j} - \frac{\left(X(t_{j+1})-X(t_j)\right)}{\Delta t_{j+1}}\right\} \qquad (1)$$

$$= \text{IN flux} - \text{OUT flux},$$

where ΔP is the change in momentum with a unit mass (i.e., the impulse: $F\Delta t = \Delta P$) for a time difference: $\Delta t = t_{j+1} - t_{j-1}$; t_j is the natural log of the jth experimental time point (refer to biological data sets); $\Delta t_j = t_j - t_{j-1}$, the CM of a critical state is $X(t_j) = \frac{1}{N}\sum_{i=1}^{N} ln\left(\varepsilon_i(t_j)\right)$ with the natural log of the ith expression $\varepsilon_i(t_j)$, $ln\left(\varepsilon_i(t_j)\right)$ at $t = t_j$; $\Delta X(t_j) = X(t_j) - \langle X \rangle$ is the fluctuation of $X(t_j)$ from the temporal average $\langle X \rangle$, and N is the number of expressions in a critical state (Table 1).

As noted, the negative force, $f\left(X(t_j)\right)$, is taken such that a linear term in the nonlinear dynamics of $X(t_j)$ corresponds to a force under a harmonic potential energy. The sign of the net self-flux shows the net incoming force to $X(t_j)$ (X<=): net IN flux for $f(\Delta X(t_j)) > 0$, and net outgoing force from $X(t_j)$ (X=>): net OUT flux for $f(\Delta X(t_j)) < 0$, where the net IN (OUT) flux corresponds to activation (deactivation) flux (force), respectively.

2. **Interaction flux**: A degree of interaction can be evaluated as the exchange of effective force, so that the interaction force can again be decomposed into an IN-coming interaction flux of a critical state, $X(t_j)$ from another critical state or the environment, Y, from the past (t_{j-1}) to the current time (t_j), and an OUT-going interaction flux of $X(t_j)$ to Y from the current time (t_j) to the future (t_{j+1}), and the net interaction flux is defined as

$$f\left(\Delta X(t_j); \Delta Y\right) = \frac{1}{\Delta t}\left\{\frac{\left(\Delta X(t_j) - \Delta Y(t_{j-1})\right)}{\Delta t_j} - \frac{\left(\Delta Y(t_{j+1}) - \Delta X(t_j)\right)}{\Delta t_{j+1}}\right\}, \qquad (2)$$

where the first and second terms represent IN and OUT flux, respectively and $Y \in \{\text{Super}, \text{Near}, \text{Sub}, \text{E}\}$ with $Y \neq X$: E represents the environment. The sign of the net interaction flux (equation (2)) corresponds to the direction of interaction: net IN interaction flux (Y=>X) for positive and net OUT interaction flux (X=>Y) for negative. The interaction flux between critical states can be defined when the number of expressions in a critical state is normalized.

3. **The flux network**: Due to the law of force, the net self-flux of a critical state, $X(t_j)$, is the summation of interaction fluxes (equation (2)) with the other critical states and the environment given by

$$f\left(\Delta X(t_j)\right) = \sum_{i=1}^{M=2} f\left(\Delta X(t_j); \Delta A_i\right) + f\left(\Delta X(t_j); \Delta E\right), \qquad (3)$$

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). Equation (3) tells us that the sign of the difference between the net self-flux and the overall contribution from internal critical states, $f\left(\Delta X(t_j)\right) - \sum_{i=1}^{M=2} f\left(\Delta X(t_j); \Delta A_i\right)$, reveals incoming flux (positive) from the environment to a critical state or outgoing flux (negative) from a critical state to the environment; when the difference in all critical states becomes zero, the genome system itself is *closed* thermodynamically (no flux flow from the environment).

4. **The average flux balance**: If we take a temporal average of equation (1) or (3) and equation (2) for the data for both MCF-7 and HL-60 cells, we obtain the average flux balances of a critical state (Table 2), X, and its interaction with other states or the environment, respectively:

$$\langle f\left(\Delta X(t_j)\right)\rangle \approx 0, \text{and } \langle f\left(\Delta X(t_j); \Delta Y\right)\rangle + \langle f\left(\Delta Y(t_j); \Delta X\right)\rangle \approx 0. \qquad (4)$$

5. **Near-synchronous interaction flux dynamics at the sub-critical state**: The interaction fluxes of the sub-critical state with the near- and super-critical states are near-synchronous for MCF-7 cell fates (**Supplementary Figure S2**):

$$f\left(\Delta\mathrm{Sub}\left(t_j\right); \Delta\mathrm{Super}\right) \approx f\left(\Delta\mathrm{Sub}\left(t_j\right); \Delta\mathrm{Near}\right), \tag{5}$$

which suggests the existence of coherent molecular transcriptional events from the near- and super-critical states to the sub-critical state, and vice versa. From equation (3), this net interaction flux synchronization (Equation (5)) gives the balance dynamics of flux at the sub-critical state between the super-critical state and the environment, and the net self-flux of the sub-critical state is well-approximated by its interaction flux with the near-critical state:

$$f\left(\Delta\mathrm{Sub}\left(t_j\right); \Delta\mathrm{Super}\right) + f\left(\Delta\mathrm{Sub}\left(t_j\right); \Delta\mathrm{E}\right) \approx 0, \tag{6}$$

and

$$f\left(\Delta\mathrm{Super}\left(t_j\right)\right) \approx f\left(\Delta\mathrm{Super}\left(t_j\right); \Delta\mathrm{Near}\right). \tag{7}$$

Contributions

KY and MT initiated the project; MT, AG and KY designed the study; MT analyzed data; KY provided theoretical support; MT and AG wrote the manuscript; JE contributed research and a literature analysis of the biological interpretation of results; and MH performed data treatment to develop a systematic method.

Acknowledgments

MT sincerely thanks the Institute for Advanced Biosciences, Keio University, Tsuruoka City, and the Yamagata prefectural government, Japan, and his family for allowing him to complete this research project at Keio University. The authors are thankful to Dr. Vyacheslav Lyashenko for a fruitful discussion.

Reference:

1. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 131: 861–872.
2. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, et al. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1, 55–70.
3. Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, et al. (2008) Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2, 437–447.
4. Young RA (2011) Control of the embryonic stem cell state. Cell 144: 940–945.
5. Buganim Y, Dina AF, Rudolf J (2013) Mechanisms and models of somatic cell reprogramming, Nat Rev Gent, 14, 427–439.
6. Raser JM, O'Shea EK (2005) Noise in gene expression: Origins, consequences, and control. Science 309: 2010–2013.
7. Yoshikawa K (2002) Field hypothesis on the self-regulation of gene expression. J Biol Phys 28: 701–712.
8. Yamanaka S (2009) Elite and stochastic models for induced pluripotent stem cell generation. Nature 460: 49–52.

9. Transtrum MK, Machta B, Brown K, Daniels BC, Myers CR, Sethna JP (2015) Perspective: Sloppiness and emergent theories in physics, biology, and beyond. J Chem Phys 143, 010901-010913.
10. Tsuchiya M, Hashimoto M, Takenaka Y, Motoike IN, Yoshikawa K (2014) Global genetic response in a cancer cell: Self-organized coherent expression dynamics. PLOS One 9: e97411.
11. Tsuchiya M, Giuliani A, Hashimoto M., Erenpreisa J, Yoshikawa K (2015) Emergent Self-Organized Criticality in gene expression dynamics: Temporal development of global phase transition revealed in a cancer cell line. PLoS One 11, e0128565.
12. Estrada J, DePace, A, Gunawardena J (2016) Information Integration and Energy Expenditure in Gene Regulation. Cell 166: 234-244.
13. Prigogine, I, Nicolis, G (1977) Self-Organization in Non-Equilibrium Systems. Wiley.
14. Ricard V. Solé, (2011) Phase Transitions. Princeton University Press.
15. Marković D, Gros C (2014) Power laws and self-organized criticality in theory and nature. Physics Reports 536: 41-74.
16. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326: 289–293.
17. Sexton T, Bantignies F, Cavalli G (2009) Genomic interactions: Chromatin loops and gene meeting points in transcriptional regulation. Semin Cell Dev Biol 20: 849–855.
18. Gibcus JH, Dekker J (2013) Connecting the genome: Dynamics and stochasticity in a new hierarchy for chromosome conformation. Mol Cell 49: 773–782.
19. Nagano T, Lubling Y, Stevens TJ, Shoenfelder S, Yaffe E, Dean W, et al. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502: 59–64.
20. Jost D, Carrivain P, Cavalli G, Vaillant C (2014) Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res 42: 9553-9561.
21. Nagashima T, Shimodaira H, Ide K, Nakakuki T, Tani Y, Takahashi K, et al. (2007) Quantitative transcriptional control of ErbB receptor signaling undergoes graded to biphasic response for cell differentiation. J Biol Chem 282: 4045–4056.

22. Nakakuki T, Birtwistle MR, Saeki Y, Yumoto N, Ide K, Nagashima T, et al. (2010) Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics. Cell 141: 884–896.
23. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, et al. (2014) The DNA methylation landscape of early human embryos. Nature 511: 606-610.
24. Schiessel H (2003) Topical Review: The physics of chromatin. J Phys Condens Matter 15: R699-R774.
25. Kornyshev AA (2010) Physics of DNA: unravelling hidden abilities encoded in the structure of 'the most important molecule'. Phys Chem Chem Phys 12:12352-78.
26. Le Treut G, Képès F, Orland H (2016) Phase behavior of DNA in the presence of DNA-binding proteins. Biophys J 110: 51-62.
27. Tremethick DJ (2007) Higher-order structures of chromatin: The elusive 30nm fiber. Cell 128: 651-654.
28. Li G, Reinberg D (2011) Chromatin higher-order structures and gene regulation. Curr Opin in Genet Dev 21: 175-186.
29. Lin CY, Lovén J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA (2012) Transcriptional amplification in tumor cells with elevated c-Myc. Cell 151: 56-67.
30. Sabò A, Kress TR, Pelizzola M, de Pretis S, Gorski MM, Tesi A, et al. (2014) Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. Nature 511: 488-492.
31. Frank SR, Schroeder M, Fernandez P, Taubert S, Amati B (2001) Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. Genes Dev 15: 2069-2082.
32. Saeki Y, Endo T, Ide K, Nagashima T, Yumoto N, Toyoda T, et al. (2009) Ligand-specific sequential regulation of transcription factors for differentiation of MCF-7 cells. BMC Genomics 10: 545.
33. Gilbert SF (2013) Developmental Biology 10th ed. Sinauer Associates, US.
34. Tarella C, Ferrero D, Gallo E, Pagliardi GL, Ruscetti FW (1982) Induction of differentiation of HL-60 cells by dimethyl sulfoxide: evidence for a stochastic model not linked to the cell division cycle. Cancer Res 42: 445-449.

35. Tsiftsoglou AS, Wong W, Hyman R, Minden M, Robinson SH (1985) Analysis of commitment of human leukemia HL-60 cells to terminal granulocytic maturation. Cancer Res 45: 2334-2339.
36. Papp B, Brouland JP, Arbabian A, Gélébart P, Kovács T, Bobe R, et al. (2012) Endoplasmic reticulum calcium pumps and cancer cell differentiation. Biomolecules 2: 165-186.
37. Whitaker M (2006) Calcium at fertilization and in early development. Physiol Rev 86:25-88.
38. Schaefer A, Magócsi M, Stöcker U, Kósa F, Marquardt H (1994) Early transient suppression of c-myb mRNA levels and induction of differentiation in Friend erythroleukemia cells by the [Ca2+]i-increasing agents cyclopiazonic acid and thapsigargin. J Biol Chem 269: 8786-8791.
39. Dolmetsch RE, Lewis RS, Goodnow CC, Healy JI (1997) Differential activation of transcription factors induced by Ca2+ response amplitude and duration. Nature 386 :855-858.
40. Yen A, Forbes M, DeGala G, Fishbaugh J (1987) Control of HL-60 cell differentiation lineage specificity, a late event occurring after precommitment. Cancer Res 47: 129-134.
41. Launay S, Giannì M, Kovàcs T, Bredoux R, Bruel A, Gélébart P, et al. (1999) Lineage-specific modulation of calcium pump expression during myeloid differentiation. Blood 93: 4395-4405.
42. Telford NA, Watson AJ, Schultz GA (1990) Transition from maternal to embryonic control in early mammalian development: a comparison of several species. Mol Reprod Dev 26:90-100.
43. Wennekamp S, Mesecke S, Nédélec F, Hiiragi T (2013) A self-organization framework for symmetry breaking in the mammalian embryo. Nat Rev Mol Cell Biol. 14(7):452-459.
44. Choy JS, Wei S, Lee JY, Tan S, Chu S, Lee TH (2010) DNA methylation increases nucleosome compaction and rigidity. J Am Chem Soc 132:1782-1783.
45. Niakan KK, Han J, Pedersen RA, Simon C, Pera RA (2012) Human pre-implantation embryo development. Development 139:829-841.

46. Yang CX, Liu Z, Fleurot R, Adenot P, Duranthon V, Vignon X, Zhou Q, Renard JP, Beaujean N (2103) Heterochromatin reprogramming in rabbit embryos after fertilization, intra-, and inter-species SCNT correlates with preimplantation development. Reproduction. 145:149-159.
47. Peaston AE, Knowles BB, Hutchison KW (2007) Genome plasticity in the mouse oocyte and early embryo. Biochem Soc Trans 35:618-622.
48. Takenaka Y, Nagahara H, Kitahata H, Yoshikawa K (2008) Large-scale on-off switching of genetic activity mediated by the folding-unfolding transition in a giant DNA molecule: An hypothesis. Phys Rev E 77: 031905.
49. Nagahara H, Yoshikawa K (2010) Large system in a small cell: A hypothetical pathway from a microscopic stochastic process towards robust genetic regulation. Chem Phys Lett 494: 88–94.
50. Plenz D, Niebur E (2014) Criticality in Neural Systems. Wiley-VCH, Germany.
51. DeDecker BS (2000) Allosteric drugs: thinking outside the active-site box. Chem Biol *7*: R103-R107.
52. Dror RO, Green HF, Valant C, Borhani DW, Valcourt JR, Pan AC, et al. (2013) Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. Nature 503: 295-299.
53. Huang S, Eichier G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys Rev Lett 94: 128701–128705.
54. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, et al. (2012) A validated regulatory network for Th17 cell specification. Cell 151: 289-303.
55. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol 20:1131-1139.
56. Deng Q, Ramsköld D, Reinius B, Sandberg R (2014) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343:193–196.

57. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185–193.
58. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.
59. McClintick JN, Edenberg HJ (2006) Effects of filtering by present call on analysis of microarray experiments. BMC Bioinformatics 7: 49.
60. Tsuchiya M, Selvarajoo K, Piras V, Tomita M, Giuliani A (2009) Local and global responses in complex gene regulation networks. Physica A 388:1738–1746.
61. Tsuchiya M, Piras V, Choi S, Akira S, Tomita M, Giuliani A, et al. (2009) Emergent genome-wide control in wildtype and genetically mutated lipopolysaccharide-stimulated macrophages. PLOS One 4: e4905.
62. Bak P, Tang C, Wiesenfeld K (1987) Self-organized criticality: An explanation of 1/f noise. Phys Rev Lett 59(4): 381–384.

Supplementary File S1

Here, we further demonstrate that i) the power law scaling in the scaling divergent behavior (e.g., **Figure 4A**) is not a statistical artifact and ii) the dynamics of collective behavior of gene expression exists through interactions among genes in the power-law scaling as follows.

We showed that a scaling region in scaling-divergent behavior (**Figure 4A**) does actually exist in the sub-critical state (bimodal frequency distribution in MCF-7 cells [11]).

Figure X1 below (for MCF-7 cells) shows that the bimodality coefficient of random samplings of gene expression well converges to that of the sub-critical state, which reveals that, in the power-law scaling (**Figure 4A**: MCF-7 cells), a self-similar frequency distribution exists among random samplings of gene expression. The result provides additional supporting evidence to further confirm the existence of scaling behavior in the sub-critical state.

Next, we demonstrated in [11] that the coherent behavior emerges in ensemble of stochastic gene expression (coherent-stochastic behavior: CSB). The emergent coherent dynamics follows the dynamics of the center of mass (CM):

i) The oscillatory coherent dynamics (ABS: **Figure X2**) is shown to have a good correlation to dynamics of its center of mass (see Algebraic correlation of CBS in [M. Tsuchiya, et al, 2015], and

ii) The coherent dynamics in random samplings shown in **Figure X1** converges to CM of the sub-critical state (Figure 10A in section IV). Thus, scaling behavior in the sub-critical state stems from CSB through interactions among genes. In **Figure X2**, the dynamics of probability density profiles of gene expression in the sub-critical state represents the dynamics of CSB.

Therefore, the power law scaling is not a statistical artifact and collective behavior of expression as CBS through interactions among genes does exhibit in the power law scaling.

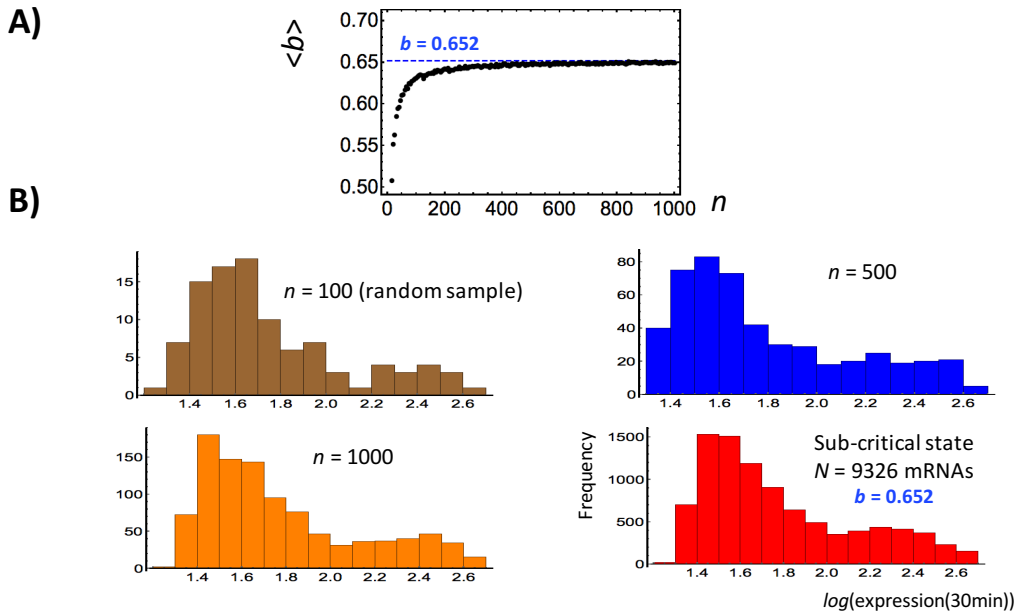


Figure X1: ***Evidence of scaling behavior in self-similarity of random sampling of gene expression:***
A) *The bimodality coefficient of random samplings (n: the number of randomly selected gene expressions) of gene expression well converges to that of the sub-critical state (b = 0.652 at t =30min as an example). b is Sarle's bimodality coefficient for a finite sample, when b> 5/9 (~0.556) indicates a bimodal or multimodal distribution. Average bimodality coefficient, <b>, over 200 repeats is estimated for each random sampling.*
B) *The frequency distribution of ensemble of randomly selected gene expression (n = 100: Brown, 500: Blue, 1000 mRNAs: Orange) from the sub-critical state has a self-similar bimodal distribution to that of the sub-critical state (Red histogram).*

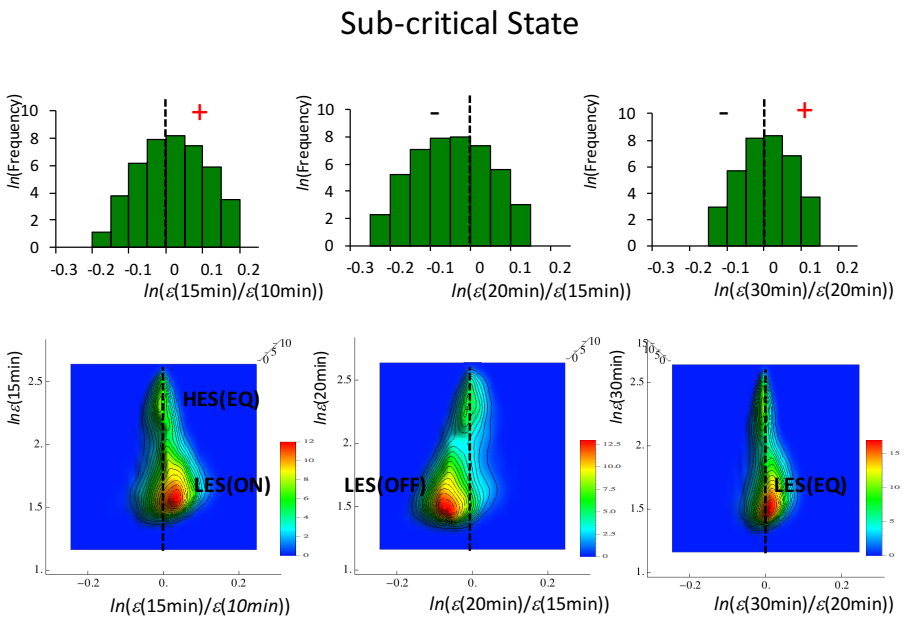
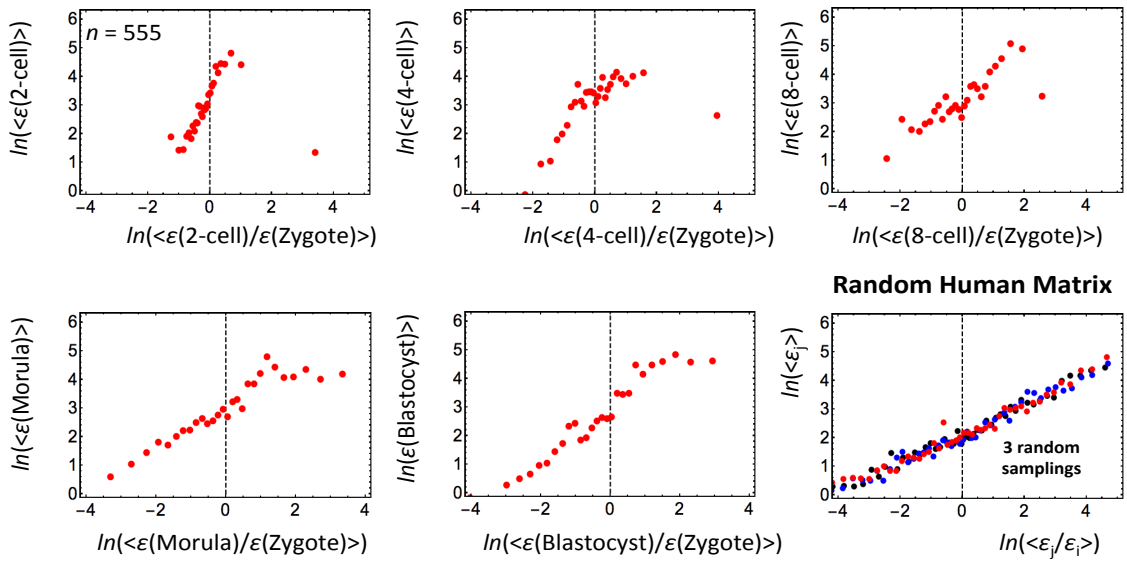


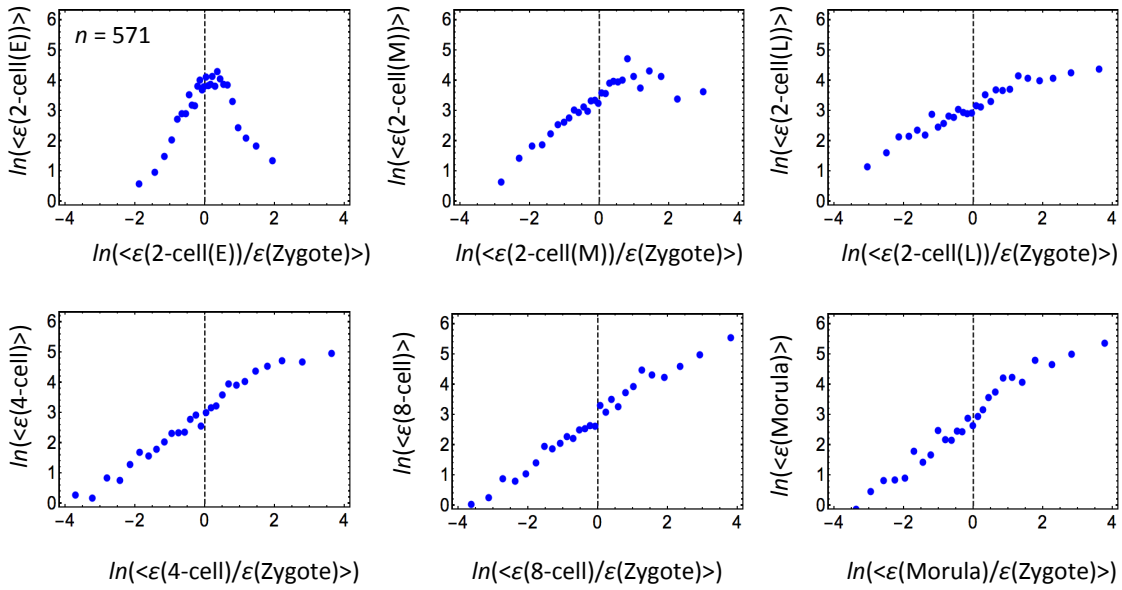
Figure X2: ***Dynamics of CSB: collective Behavior in the sub-critical state (HRG stimulated MCF-7 cells).*** *Pseudo-3-D probability density profiles on the regulatory space show that two CESs form a pair to develop a pendulum oscillatory system, low-expression state (LES) swinging around high-expression state (HES) (Figure cited from Figure 7 in [10]; ON: up-regulation; OFF: down-regulation; EQ: zero change). This pendulum-like oscillatory system acting as a pair of CESs is defined as an autonomous bistable switch (ABS). Furthermore, the oscillatory dynamics of ABS is shown to have a good correlation to dynamics of the center of mass of ABS. Please refer to detail analysis of ABS dynamics in [11].* $\varepsilon(t)$ *represents overall expression at time t. Note: Shu and colleagues [Performance assessment of kernel density clustering for gene expression profile data. Comp Funct Genomics 4: 287–299, 2003] demonstrated, by means of density analysis of noisy gene-expression profiles, the robustness of gene expression clustering.*

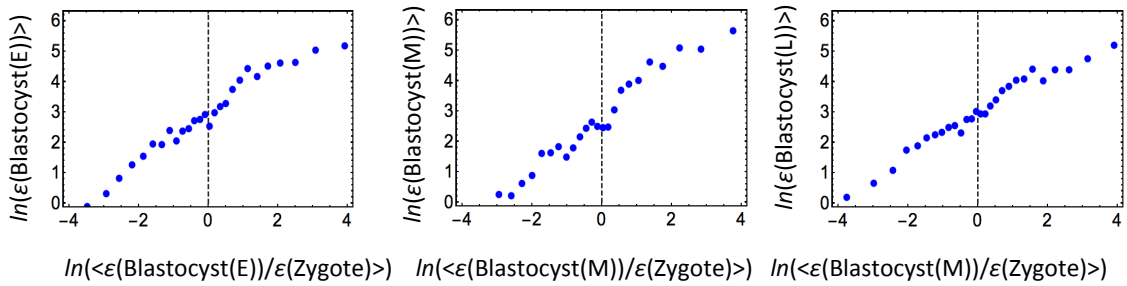
Supplementary Figure S1

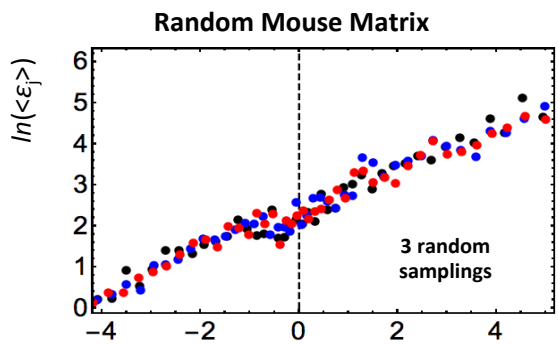
A) Human Embryo



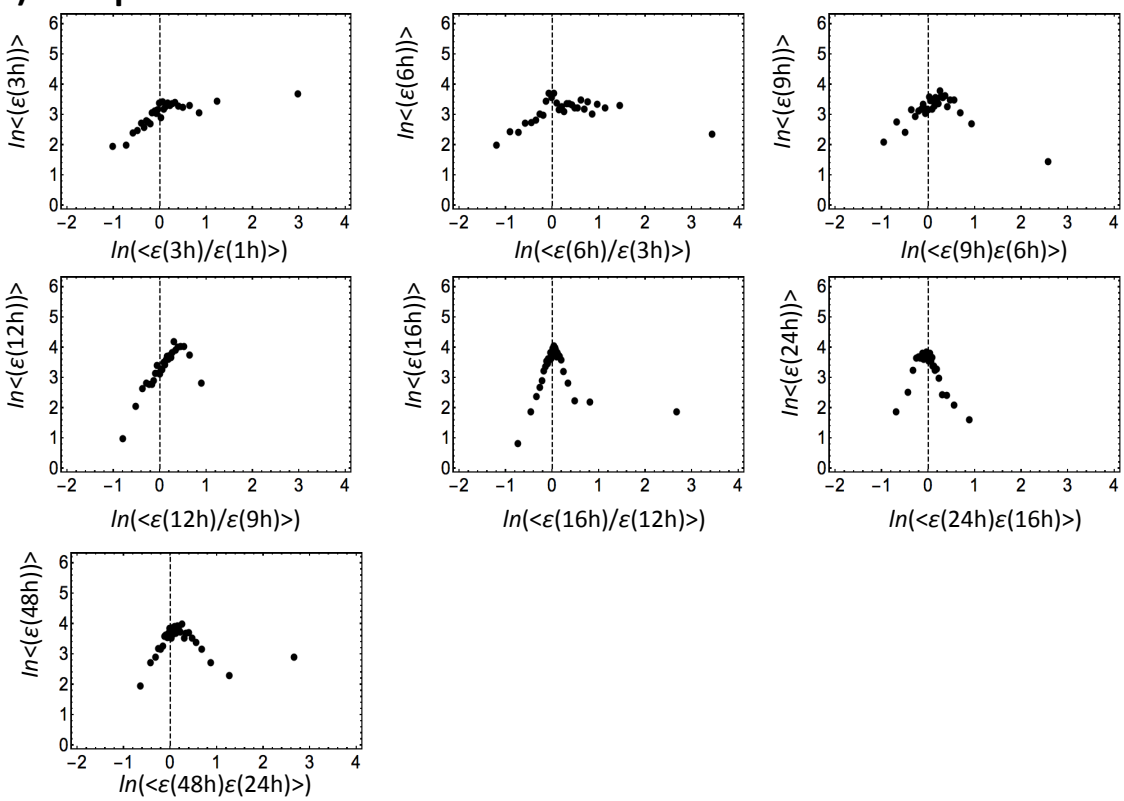
B) Mouse Embryo







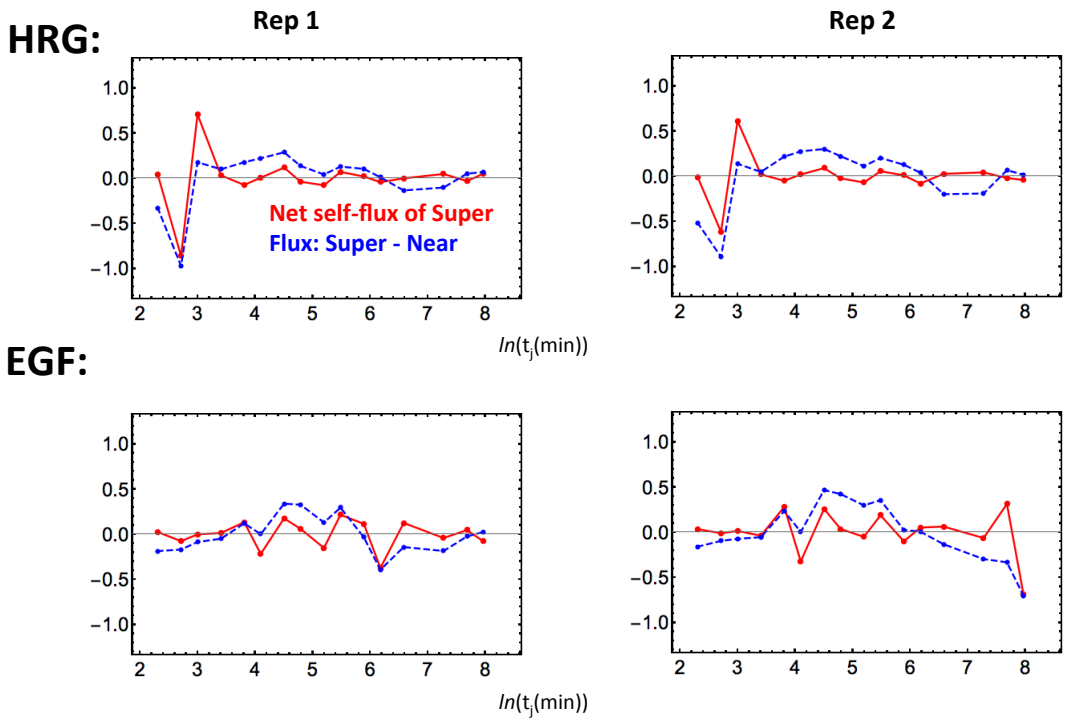
C) T helper 17 cell



Supplementary Figure S1:

Complete sandpile transitional analyses on RNA-Seq data (RPKM) for A) human and B) mouse embryo development from the zygote stage, and C) T helper 17 cell differentiation from naïve CD4+ T cells. Random human and mouse expression matrixes (A and B: last panels), (i,j) (i: number of time points; j: number of RNAs; right panels in the second row) reflect stochastic overall expression (i.e., zero correlation between any time points); random matrixes are generated by random shuffling of the corresponding original expression matrixes to show linear correlation behavior (3 random samplings are shown in different colors) similar to that of the DMSO random expression matrix (microarray data: ***Figure 3D****). The embryo results suggest that reprogramming of the genome destroys the SOC zygote control in early embryo development. In T helper 17 cell (Th17) differentiation, the development of a sandpile-type critical transition between sequential cell states is shown.*

Supplementary Figure S2



Supplementary Figure S2:

*Near-synchronous interaction fluxes between near- and super-critical states at sub-critical states (**Figure 12C**) for MCF-7 cell fates. This shows that the net self-flux of the sub-critical state is well-approximated by its interaction flux with the near-critical state (Methods), i.e., the genetic energy activities for the gene ensemble with the greatest variation stems from their interaction with near-critical states, where a critical transition occurs. This tells us that the collective behavior in the super-critical state has a direct impact on the critical transition, or vice versa. Two replicas (rep 1 and rep 2) of the same experiment (Methods) were considered.*