bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Underlying Genomic Mechanism for Cell-Fate Change from Embryo to Cancer Development

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

[1] SEIKO Life Science Laboratory, SRI, Osaka 540-659, Japan
[2] Environment and Health Department, Istituto Superiore di Sanitá, 00161 Rome, Italy; alessandro.giuliani@iss.it
[3] Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe 610-0394, Japan; keyoshik@mail.doshisha.ac.jp

*Correspondence: tsuchiya.masa@gmail.com

Abstract

Through our studies on whole genome regulation, we have demonstrated the existence of self-organized critical control (SOC) of whole gene expression at both the cell population and single cell level. In this paper, we go further in depth into the elucidation of general genomic mechanism that underlies the cell-fate change from embryo to cancer development.

In both single-cell and cell-population genome expression, a systematic determination of critical point (CP) and associated critical states with dynamical pictures for between-state flux provide a potential universal mechanism of self-organization in terms of 'genome-engine'. An autonomous critical control system is developed by a highly coherent behavior of low-variance genes (sub-critical state) generating a dominant cyclic expression flux with high-variance genes (super-critical state) through the cell nuclear environment. To determine when and how cell-fate decision occurs under the SOC mechanism, the coherent dynamics on the genome-engine are working through the dynamic transition of higher-order structure of genomic DNA (corresponding to the CP), which causes either an activated (ON) or inactivated (OFF) state in a self-organized manner.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

I. Introduction

A mature mammalian somatic cell can reprogram its state (and consequently acquire a very different gene expression profile) through a few reprogramming stimuli [Takahashi, K., Yamanaka, S., 2016]. This state change (cell-fate change) involves on/off switching on thousands of functionally unique heterogeneous genes in a remarkably coordinated manner [MacArthur, B. D., et al., 2009]. There are fundamental physical difficulties in eliminating such large scale coordinated control from a gene-by-gene basis. These difficulties become more evident in a situation where there is a lack of sufficient number of molecules to reach a stable thermodynamic state and a consequent stochastic noise due to the low copy number of specific gene mRNAs, thereby inducing a substantial instability of genetic product concentrations falsifying any gene-by-gene feedback control hypothesis [Raser, J. M., O'Shea, E. K., 2005; Yoshikawa, K., 2002].

In our previous studies [Tsuchiya, M., et al., 2014-2017; Giuliani, A., et al., 2018], we have demonstrated the existence of self-organization in whole genome expression at both the population and single cell level that constitutes a 'physically motivated' alternative to the gene-specific regulation. The mechanism of self-organization eliminates massive changes in the expression profile through the genome reprogramming inside a small and highly packed cell nucleus.

The core of self-organization mechanism is the presence of massive system changes elicited by minor 'apparent' external causes: Per Bak and colleagues [Bak, P., et al., 1987] proposed self-organized criticality (SOC; the Bak-Tang-Wiesenfeld sandpile model) to address this problem. SOC is a general theory of complexity that describes self-organization and emergent order in non-equilibrium systems (thermodynamically open systems), where self-organization is considered to occur at the edge between order and chaos [Langton, C. G., 1990; Kauffman, S. A., 1993], often accompanied by the generation of exotic patterns (good description of SOC in [Jensen, H. J. 1998; Marković, D., Gros, C., 2014]; also see current review on criticality in [Muñoz, M. A., 2018]).

SOC builds upon the fact that the stochastic perturbations initially propagate locally (sub-critical state), but due to the particularity of the disturbance, the perturbation can spread over the entire system in a highly cooperative manner (super-critical state); as the system approaches its critical point where global behavior emerges in a self-organized manner.

The above-depicted classical concept of SOC, has been extended to propose a conceptual model of the cell-fate decision (critical-like self-organization or rapid SOC) through the extension of minimalistic models of cellular behavior. The cell-fate decision-making model considers gene regulatory networks to adopt an exploratory process, where diverse cell-fate

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

options are generated by the priming of various transcriptional programs, and then a cell-fate gene module is selectively amplified as the network system approaches a critical state [Halley, J. D., et al., 2009]. Such amplification corresponds to the emergence of long-range activation/deactivation of genes across the entire genome.

We investigated whole genome expression and its dynamics to address the following fundamental questions:

- Is there any underlying principle that self-regulates whole-genome expression?
- Does a universal mechanism exist to guide the self-organization so as to determine the change in the cell fate?

Our findings suggested that at a specific time point on cellular development, a transitional behavior of expression profile occurs in the ensemble of genes (e.g., unimodal-bimodal transition in MCF-7 cancer cells [Tsuchiya, M., et al., 2014]). Such a specific feature exhibits the characteristic of self-similarity around a critical transition point, which is evident when gene expression is sorted and grouped according to temporal variance of expression (normalized root mean square fluctuation: *nrmsf*: **Methods**). On the contrary, randomly shuffled gene expression exhibits Gaussian normal distribution across the entire genome, with no evidence of cooperative behavior (no emergence of coordinated motion). This is consistent with the existence of self-organization according to *nrmsf*, that is, *nrmsf* acting as an order parameter of self-organization. This suggests that the grouping of expressions averages out expression noises coming from biological and experimental processes allows to highlight the self-organizing behavior through distinct response expression domains (critical states). Therefore, we have stressed the importance to examine and analyze the global behaviors (mean-field approach) of group expression emerged in genome expression.

Our findings of self-organization with critical behavior (criticality) (refer to Fig.1 in [Giuliani, A., et al., 2018]) differ distinctly from the classical and extended SOC models in regard to the following issues:

i) Occurrence of cell-fate change through erasure of the initial-state 'sandpile criticality' (a sandpile, when reaching a critical height with respect to its base is sensitive to the addition of a single grain corresponding to a relatively minor stimulus).

ii) Coexistence of critical states (super-critical: high temporal-variance expression; near-critical: intermediate variance expression; sub-critical: low variance expression).

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

iii) The sub-critical state as the generator of autonomous SOC control (versus non-autonomous classical SOC [Halley, J. D., et al. 2009]), which guides the cell-fate change.

iv) The existence of a potential general mechanism of cell-fate change over different biological processes.

Two distinct critical behaviors emerged when gene expression values are sorted and grouped: (1) sandpile-type criticality and (2) scaling-divergent behavior (genome avalanche). Sandpile criticality is evident in terms of grouping according to expression fold-change between two different time points, whereas scaling-divergent behavior emerges according to grouping by *nrmsf*. Criticality allows for a perturbation of the self-organization (i.e., change in critical point) due to change in signaling by external or internal stimuli into a cell to induce a global impact on the entire genome expression system.

In this report, we update our previous findings and develop a unified model of cell-fate change as follows:

1) *Systematic determination of critical point (CP) and critical states (distinct response domains) for both single cell and cell population*.

Previously we had some technical difficulties in determining distinct response domains (critical states) for single-cell genome expression in RNA-Seq data, where there are lots of zero-value expression causing specific instability in bimodal transitional behaviors of expression profile. Our findings show that the CP relative to a specific group of genes corresponds to the center of mass of the whole genome. Therefore, a new correlation metrics (CM correlation) based upon grouping of expression (CM grouping) from the center of mass (CM) of genome expression is developed; this metrics reveals fixed point (regarding temporal expression variance, i.e., time) behavior of critical points (CPs) in a specific biological regulation covering from embryo development to cell differentiation. Furthermore, these fixed CP behaviors reveal systematically distinct response domains (critical states) for both single-cell and population cell genome.

2) *Singular behaviors of the CP are interpreted in terms of specific transition of the higher-order structure of genomic DNA, corresponding to the region when cell-fate change occurs, which suggests that the CP competes between the active (swelled or coil) and inactive (compact or globule) states.*

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For population of cells, temporal CM correlation shows the timing of activation of the CP, whereas for single cell, specific transition on the higher-order structure of DNA is generated around the point where initial-state criticality is erased, i.e., the timing of when sand-pile criticality between initial (or cell state) and different time points (or cell state) disappears. Molecular activation mechanism of the CP is expected to lead to a novel cell-fate control mechanism.

3) *Cell-fate change occurs through coherent perturbation on the dominant cyclic state-expression flux (genome-engine) such as through enhancement-suppression of the genome-engine.*

Coherent behavior (i.e., CM dynamics) emerges in stochastic expression (coherent-stochastic behavior) in each critical state. These coherent behaviors exhibit a *universal genome-engine mechanism* [Tsuchiya, M., et al., 2016, 2017] for SOC-control of genome expression; the sub-critical state (ensemble of low-expression variance genes) is a generator that sustains the SOC control forming a dominant expression cyclic flux between sub- and super-critical (high-expression variance genes) states through the cell nuclear environment.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

II. Results

A. Fixed Critical Point (CP): A Specific Group of Genes Corresponding to the Center of Mass of Whole Genome

To develop a unified view of self-organizing genome expression in distinct biological regulations, the existence of a critical point (CP) plays an essential role in determining distinct response domains (critical states) [Tsuchiya, M., et al., 2016], here we go in depth into specific features of the CP exhibiting sandpile type critical behaviors (*sandpile criticality*) (**Figures 1A**) in genome expression. The sandpile criticality emerges when whole gene expression is sorted and grouped according to fold-change in expression between two different time points (e.g., between t = 0 and t =10min). For the same groupings, *nrmsf* value of the CP in HRG-stimulated MCF-7 cancer cells (population level) is estimated (ln<$nrmsf$> ~ -2.5: **Figure 1B**).

Our study on HRG-stimulated MCF-7 cancer cells, demonstrated that the temporal group correlation (between-groups correlation) along the order parameter (*nrmsf*) reveals a focal point (FP) when we consider the center of mass (CM) of whole expression (changing in time) as a reference expression point (see Fig.5B in [Tsuchiya, M., et al., 2015]). The grouping (baseline as the CM) according to the degree of *nrmsf* is called ***CM grouping***, $c^k(t)$ (k^{th} group; k =1,.., K), where grouping from the CM distinguishes from that of non-reference, $g^k(t)$.

Notably, as shown in **Figure 1C**, the CP is zero-expression point in the CM grouping, which explains why the CP is a specific set of genes corresponding to the CM. This feature holds for both single-cell and population data (**Figures 1D-F**: refer to natural log of *nrmsf* value for different biological regulations). Therefore, we develop the correlation metrics based on CM grouping (called **CM correlation**: **Methods**) to grasp how the whole expression can be self-organized through critical/singular behavior of the CP.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

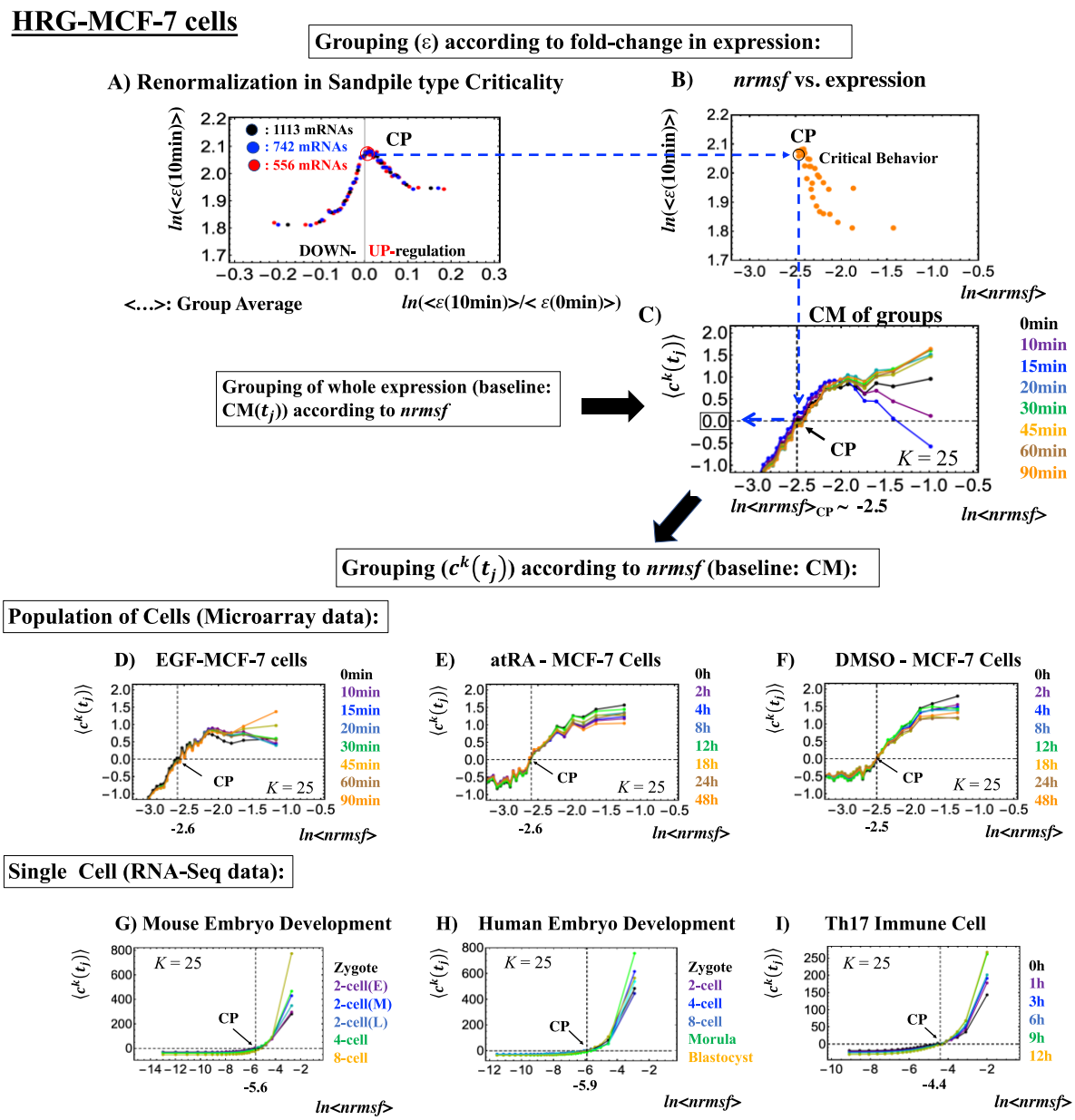


Figure 1: The fixed CP corresponding to the center of mass (CM) of genome expression:

A) Different number of groups (20: black, 30: blue, 40: red groups; one group respectively contains 1113, 742, and 556 mRNAs) in HRG stimulated MCF-7 cells follow the same sandpile-type critical behavior. This reveals the existence of scaling behaviors (i.e., renormalization).

B): *ln<nrmsf>* value of the CP: logarithm plot of average *nrmsf* value (**Methods**) of group, *ln<nrmsf>* vs. average expression value, *ln*<*ε*(10min)> shows that *ln<nrmsf>*$_{CP}$~ -2.5 in sandpile type criticality, where *ε*(10min) represents expression value of groups at t = 10min and <…> represents ensemble average of group expression. Grouping for **A)** and **B)** is ordered according to fold-change in expression at 0-10min.

C): Grouping $\left(c^k(t)\right)$ (CM grouping) of whole expression (baseline: CM(t_j)) according to *nrmsf* reveals that the CP is a fixed point and furthermore, corresponds to the CM of genome expression (whole expression). This fact is true for both the population (**C-F**) and single cell levels (**G-I**). K represents the number of groups with n number of elements (**C, D**: n = 891 mRNAs; **E, F**: n = 505; **G**: n = 685 RNAs; **H**: n = 666; **I**: K = 25, n = 525; coloring: **Methods**)

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B. Population of Cells: Activated or Inactivated State of Critical Point (CP)

We investigated the temporal and spatial development of correlation of CM groups (**Methods**) to grasp spatio-temporal response on whole expression for population of cells. Dynamics of the CM correlation reveals additional features of the CP:

i) **Temporal CM correlation**: Development of the CM correlation between the initial and other experimental time points: $\hat{\boldsymbol{c}}^k(t_0).\hat{\boldsymbol{c}}^k(t_j)$ (k = 1,2, .. ,K) over experimental point, t_j, where $\hat{\boldsymbol{c}}^k(t_j)$ is unit vector (unit length) of the kth group vector, $\boldsymbol{c}^k(t_j)$. The temporal CM correlation in HRG stimulated MCF-7 cells with cell differentiation reveals a divergent behavior at t_j = 15min (**Figure 2A**: left panel), whereas EDF-stimulated MCF-7 cells with non-differentiation (**Figure 2A:** right panel) does not show any divergent behavior (see biological issue of cell differentiation of MCF-7 cells in [Saeki, Y., et al., 2009]).

ii) **Spatial CM correlation**: Development of the CM correlation between the first group (highest *nrmsf* group) and other vectors (k) at $t = t_j$: $\hat{\boldsymbol{c}}^1(t_j).\hat{\boldsymbol{c}}^k(t_j)$; ($k$ = 1,2,3,..K). **Figure 2B** shows that for both HRG- and EGF-stimulated MCF-7 cells the spatial CM correlation exhibits a focal point (FP) at the CP, which tells that the CP has no correlation with the highest *nrmsf* group.

To understand the EGF response in terms of temporal CM correlation, we investigated if the erasure of initial-state sandpile criticality explains the behavior at the CP. The genome-state change (corresponding to cell-fate change) occurs in such a way that the initial-state SOC control of overall gene expression (i.e., initial-state global gene expression regulation mechanism) is destroyed through the erasure of an initial-state sandpile criticality [Tsuchiya, M., et al., 2016]. **Figure 2C** shows that HRG-stimulation induces the erasure of initial-state sandpile criticality at 2-3h, whereas EGF-stimulated MCF-7 cells does not erase initial-state criticality throughout the time course of the EGF stimulation and consequently, genome-state change does not occur. This suggests that **the CP possesses activated or inactivated state**, i.e., ON or OFF expression state for a set of genes (**critical gene set**) corresponding to the CP. In EGF-stimulated MCF-7 cells, the CP is in the inactivated (OFF mode) state, whereas in HRG-stimulated MCF-7 cells, the CP is ON at 10-15min and thereafter turns OFF. A direct evidence in terms of averaging behavior is shown in **Figure 2D**, where for HRG-stimulation, discrete transition of DNA (15min (ON): swelled coil state; 20min (OFF): compact globule

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

state) occurs at the CP, whereas in the case with the EGF-stimulation, such transition does not occur during the early time points.

Note: Fold change in ensemble (group) average of expression groups, $<\boldsymbol{c}^k(t_{j+1})>/<\boldsymbol{c}^k(t_j)>$ exhibits a clear transitional behavior with characteristics of first-order phase transition emerged in genome sized DNA molecules (see more in **Discussion**). Through our current studies, it has become evident that the transition occurs as coherent behavior (mega bp level) emerging from stochastic expression represented by CM (average) of group following law of large numbers (coherent-stochastic behavior [Tsuchiya, M., et al., 2017]). On the other hand, ensemble average of fold change in individual expressions between two temporal groups, $<\boldsymbol{c}^k(t_{j+1})/\boldsymbol{c}^k(t_j)>$ does not reveal such characteristics in the transition that is attributed to the stochastic behavior of expression (sensitive in fold change). Interestingly, ensemble average of time difference in the expression group, $<\boldsymbol{c}^k(t_{j+1})- \boldsymbol{c}^k(t_j)>$ supports the coherent scenario. This indicates that fluctuation (noise) on coherent dynamics is eliminated (see attached **Supplementary Figure S1**). Therefore, the CP accompanied by the transition of the higher-order structure of genomic DNA suggests that there exist coherent behaviors guiding the transition of the CP, where fluctuation occurs on the coherent dynamics.

atRA- and DMSO-stimulated HL-60 cells further support this condition of the CP (**Figure 3A**: left panel: atRA; right: DMSO). For the atRA stimulation, the CP is ON at 24-48h, which coincides with the timing of the erasure of initial-state sandpile criticality while for the DMSO stimulation the CP is ON at 12-18h, which occurs before the erasure of initial-state criticality (i.e., cell-fate change: refer to Discussion in [Tsuchiya, M., et al., 2016]). Singular behavior of the CP is attributable to the transition of DNA between swelled and compact states for the both stimulations: at 24-48h (compact: 24h; swell: 48h) for atRA, and at 18-24h (compact: 18h; swell: 24h) for DMSO. This coincides with the timing of the erasure of initial-state sandpile. As for the DMSO response, it is interesting to observe a multi-step process of erasure of initial-state sandpile criticality (**Figure 3C**): erasure at 8-12h; recovery of the criticality at 12h-18h; then erasure again at 18-24h. Multiple erasures suggest that the cell population response passes over two SOC landscapes [Tsuchiya, M et al., 2016] at 8-12h and 18-24h. In the atRA response, CP becomes ON at 24-48h coinciding with the occurrence of coil (ON) transition. Whereas in the DMSO response, CP becomes ON at 12-18h with the occurrence of folding transition into a compact state (OFF), which can be due to passing the first SOC landscape at 8-12h; notably, 18h (compact on the CP) is in the middle of transition from enhancement to suppression on the genome-engine (see **Section IIE**).

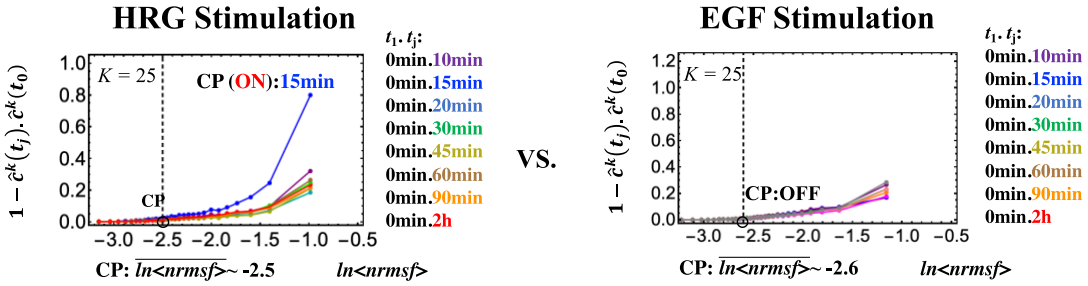
Therefore, results obtained on cancer cells suggest that **the activation of critical gene set (CP) plays an important role in cell-fate change.**

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MCF-7 cells (Cell Population)

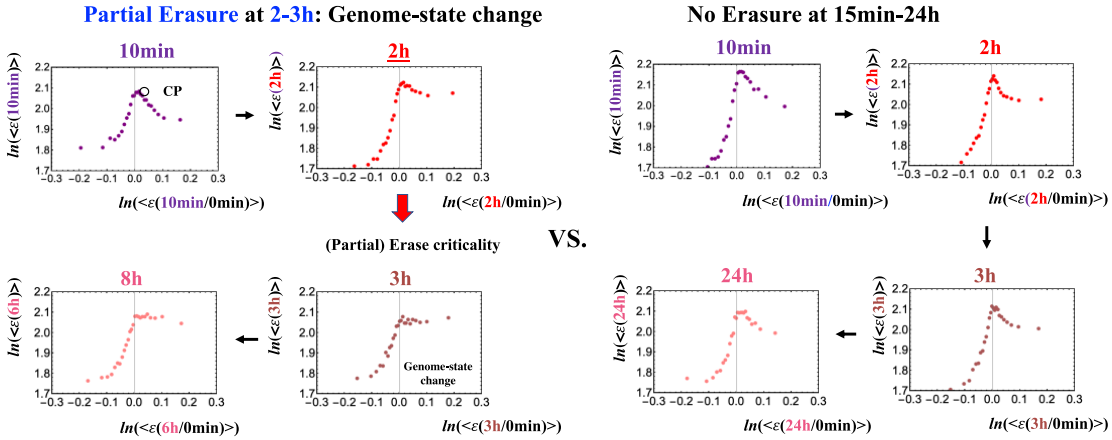
Grouping ($c^k(t_j)$) according to *nrmsf* (baseline: CM)

A) Temporal CM correlation:



B) Spatial CM correlation with $k = 1$:

C) Erasure of initial-state criticality: Grouping according to fold-change



bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D) Singular Behaviors at the CP

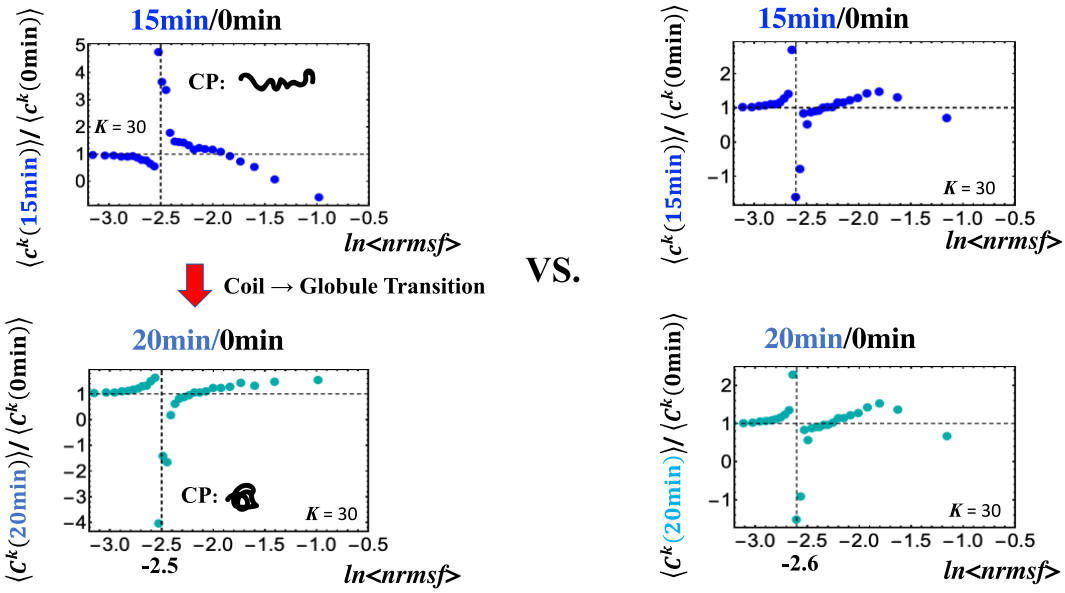


Figure 2: HRG- (left) and EGF-stimulated (right) MCF-7 cell population responses. HRG-stimulation induces MCF-7 cell differentiation, whereas EGF induces only cell proliferation with no differentiation [Saeki, Y., et al., 2009]. The CM correlation analyses (**Methods**) reported in the figure are relative to gene groups ordered along their relative *nrmsf*: **A) temporal CM correlation**: $1-\hat{c}^{k}(t_j).\hat{c}^{k}(t_0)$ and B) **spatial CM correlation,** $1-\hat{c}^{k+1}(t_j).\hat{c}^{k=1}(t_j)$, where $\hat{c}^{k}(t_j)$ represents unit vector of the kth group expression vector at $t = t_j$ (expression vector divided by its length) and dot (.) corresponds to inner product of the kth group between initial and other experimental points for (**A**) and for (**B),** to the inner product of group vectors at $t = t_j$ between initial (k = 1) and other groups.

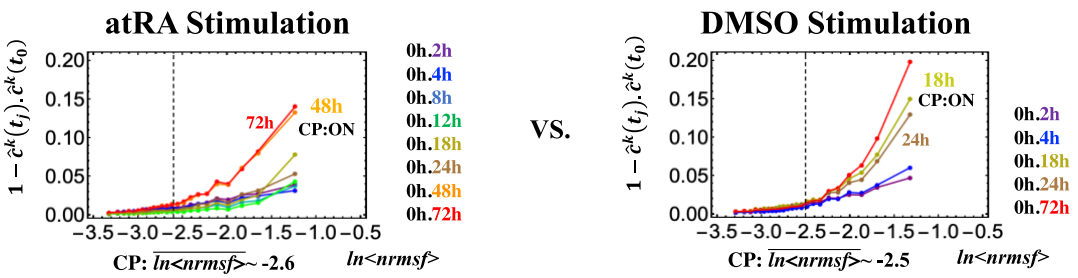
Spatial CM correlation reveals that the CP corresponds to zero correlation with the initial group (k =1) in addition to the CM of whole gene expression (for experimental times refer to **Methods**). Whereas temporal CM correlation reveals that the CP is activated: between 10 and 15 min for HRG and inactivated for EGF. Activation or inactivation of the CP marks the occurrence of cell-differentiation for HRG and only proliferation (no differentiation) for EGF.

Panel **C)** depicts the (partial) erasure of the initial-state criticality at 2-3hr for HRG and no erasure for EGF, and **D)** demonstrates the occurrence of the transition of the higher-order structure of the CP for 15-20min as shown in fold change from initial time point: swelled coil state at 15min and compact globule state at 20min respectively corresponds to ON state and OFF state of the CP (K = 30 groups with each group containing n = 742 mRNAs).

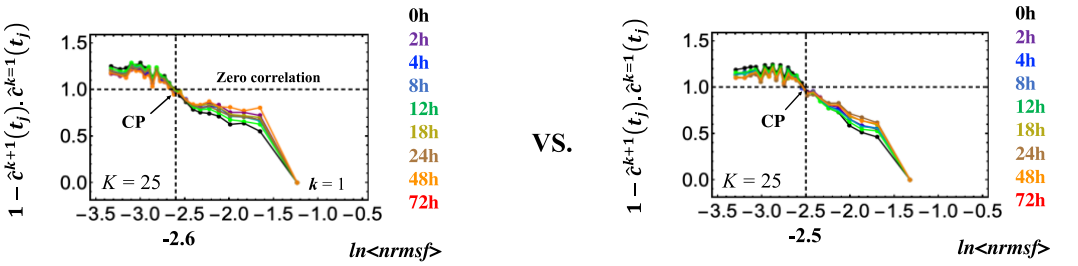
bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HL-60 cells (Cell Population)

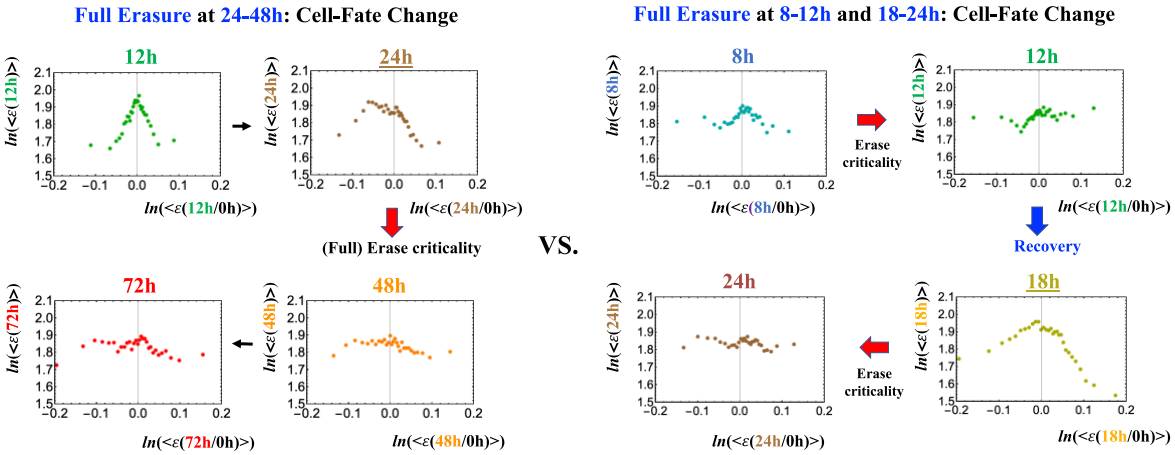
A) Temporal CM correlation:



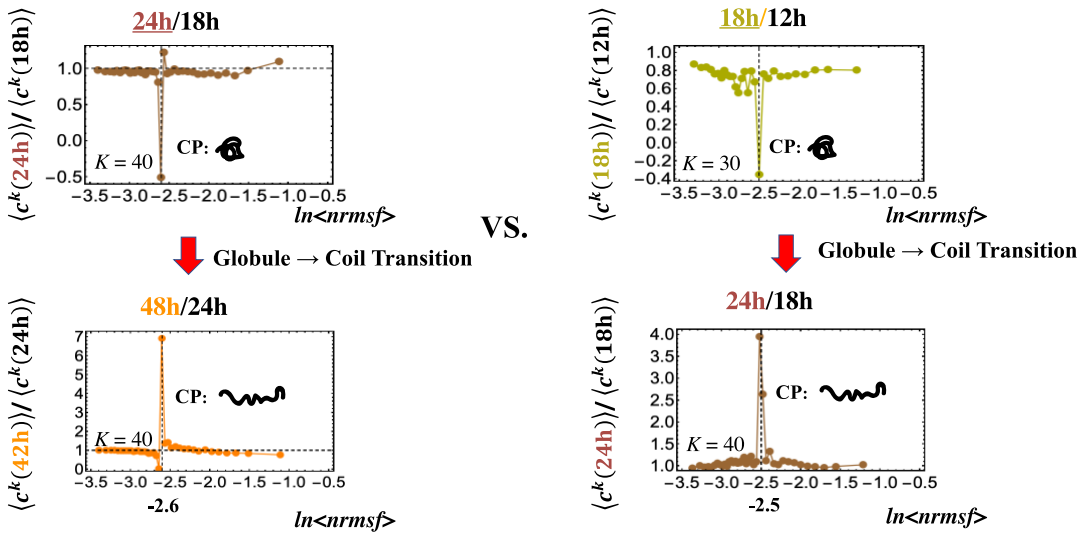
B) Spatial CM correlation with k =1:



C) Erasure of initial-state criticality: Grouping according to fold-change



D) Singular Behaviors at the CP



bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3: atRA- (left) and DMSO stimulated (right) HL-60 cell population. Activation of the CP occurs **A)** between 24 and 48h for atRA (left panel) and between 12 and 18h for DMSO (right).
B) The CP occurs at zero spatial correlation with k = 1 group, which is the same for the MCF-7 cells. The timing of the cell-fate change [Tsuchiya, M., et al., 2016] is indicated by the timing of erasure of initial sandpile criticality.
C) Erasure happens 24-48h for atRA, and at 18-24h for DMSO, additional full erasure of the initial-state criticality occurs at 12-18h; thereafter recovering the initial-state criticality at 18h indicates the existence of two SOC landscapes at 8-12h and 18-12h, respectively.
D) Transition of the higher-order structure at the CP is suggested to occur at 24-48h for atRA stimulation and at 18-24h for DMSO stimulation, when the initial-state criticality is erased (**C**). **Note:** Singular behaviors at the CP, however, shows sensitivity of the higher-order structural transition in terms of noisy (zigzagging) behaviors around the CP (see **Figure 1C-D**) due to change in the number of groups (plots for K = 40 groups: n = 420 mRNAs; K = 40 groups: n = 315 mRNAs). Underlined numbers represent the timing just before the cell-fate change.

B. Single Cell: Reprogramming of Embryo Development and Cell Differentiation Through Inversion of Singular Behaviors of the CP

In the case of embryo development, the behavior of the CP is different from cell population (microarray) data. In the temporal CM correlation, the CP is a point with no differential (**Figure 4A**), while it appears as a divergent point in cell populations. This feature reveals distinct response domains (critical states) in the single-cell genome expression (see below); in the spatial CM correlation (**Figure 4B**), the CP is a focal point of spatial CM correlations having no correlation between neighboring groups (vs. focal point with initial group in population).

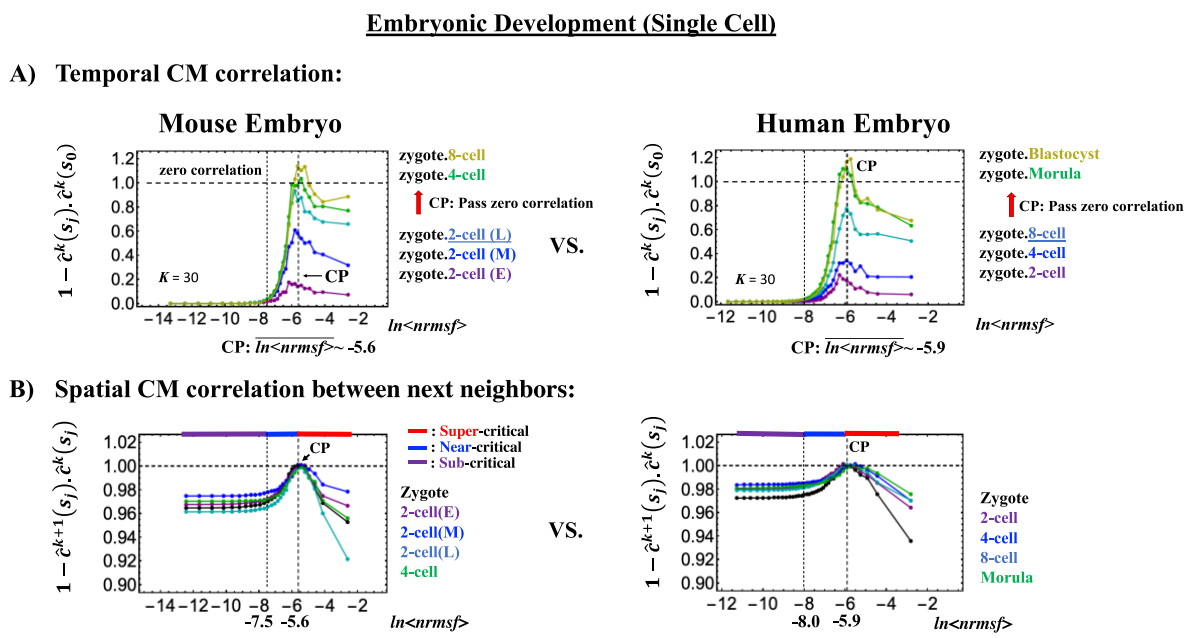
In the reprogramming event, the temporal CM correlation for the CP traverses zero value (corresponding to random-like behavior) after late 2-cell and 8-cell state for mouse and for human, respectively (**Figure 4A**), which coincides with the erasure of the sandpile-type critical point (CP) (**Figure 4C**). In biological terms, this corresponds to the erasure of the initial stage of embryogenesis (driven by maternal heredity); Thus, an additional condition for the embryonic reprogramming is suggested whether or not the temporal CM correlation passes the zero point. It is important to note that groups of low-*nrmsf* presenting flattened CM correlations in time ($ln\langle nrmsf\rangle < -8.0$: **Figure 4A**) do not point to a no-response situation; on the contrary, they behave in a highly coherent manner to generate the autonomous SOC mechanism (see e.g., Fig.6 in [Tsuchiya, M, et al., 2015]).

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Notably, the transition of the CP (**Figure 4D**) supports the timing of reprogramming (late 2-cell state for mouse and 8-cell state for human embryo) in that inversion of singular behaviors at the CP occurs before and after the reprogramming. Before reprogramming, the folding transition onto a globule state (OFF) undergoes a coil like transition (ON) after reprogramming, and vice versa (see more in **Discussion** regarding intra-chain segregation in the transition).

On the other hand, in Th17 immune single cell (**Figure 5**), the CP does not pass over zero temporal correlation with the initial state (t = 0h), which indicates that cell differentiation induces a partial-scale (specific set) change in the whole expression (vs. whole-scale change in embryonic reprogramming). The timing of genome change, as in other biological regulations, is determined by the erasure of the initial-state sandpile criticality. This appears at 6h where an inversion of singular behaviors of the CP takes place: before cell differentiation, the folding transition onto a globule state (OFF) occurs at the CP followed by a switch to a coil (ON) state after cell differentiation (**Figure 5D**).

Furthermore, for both single cell and cell population levels as demonstrated in **section IIE**, the timing of the inverse transition at the CP through the cell-fate change coincides with the timing of inverse coherent perturbation on the genome engine (cyclic flux flow). This suggests that the change in singular transition at the CP through the cell-fate change provokes a global impact on the genome (see also [Tsuchiya, M., et al., 2016] in regard to expression flux dynamics).



bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

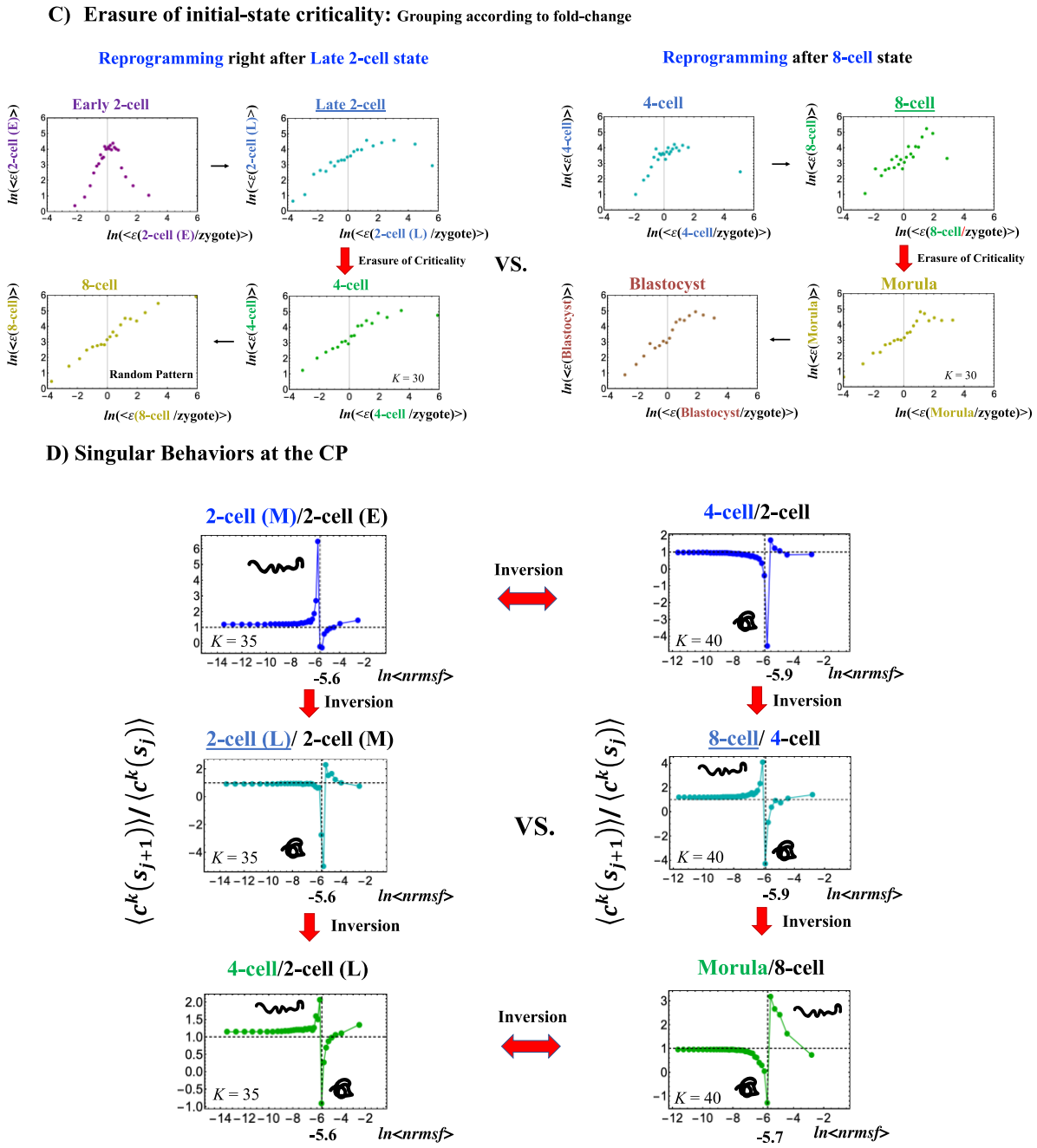


Figure 4: Mouse (left panel) and human (right) embryonic development (single cell). The CP corresponds to **A)** the peak of temporal correlation and **B)** zero spatial correlation between neighboring groups (see the difference from cell population in **Figures 3B**, **4B**).

Distinct response domains (critical states) as manifested in both temporal and spatial CM correlations are shown (red: super-critical; blue: near-critical; purple: sub-critical), where location of the CP lies at the boundary between near- and super-critical states. Notably, the timing of reprogramming from the late 2 cell to 4 cell states for mouse embryo and from the 8-cell state to morula state for human embryo is demonstrated by the timing of **A) zero temporal CM correlation**, **C) erasure of the initial-state (zygote) criticality**, and **D) inverse singular transition of the CP**.

D) Mouse (left panel) and human embryo (right) developments have an infinite (albeit numerically finite) type discontinuity at the CP (i.e., opposite singular behaviors at before and after the CP in terms of *nrmsf*); this suggests the occurrence of intra-chain segregation in the transition for genome sized

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA molecules (see more in **Discussion**). When passing through reprogramming (late 2-cell state in mouse and 8-cell state in human), inversion of singular behaviors at the CP occurs and furthermore, inversions between mouse and human embryo developments are also observed (K= 35 groups with n = 489 RNAs for mouse and K= 40 groups with n = 416 RNAs for human). Different types of singular behavior occur between reprogramming of embryo development and cell differentiation (see **Figures 3D, 5D**). **Note:** The location of the CP lies at boundary between near- and super-critical states.

Th17 Immune Cell (Single Cell)

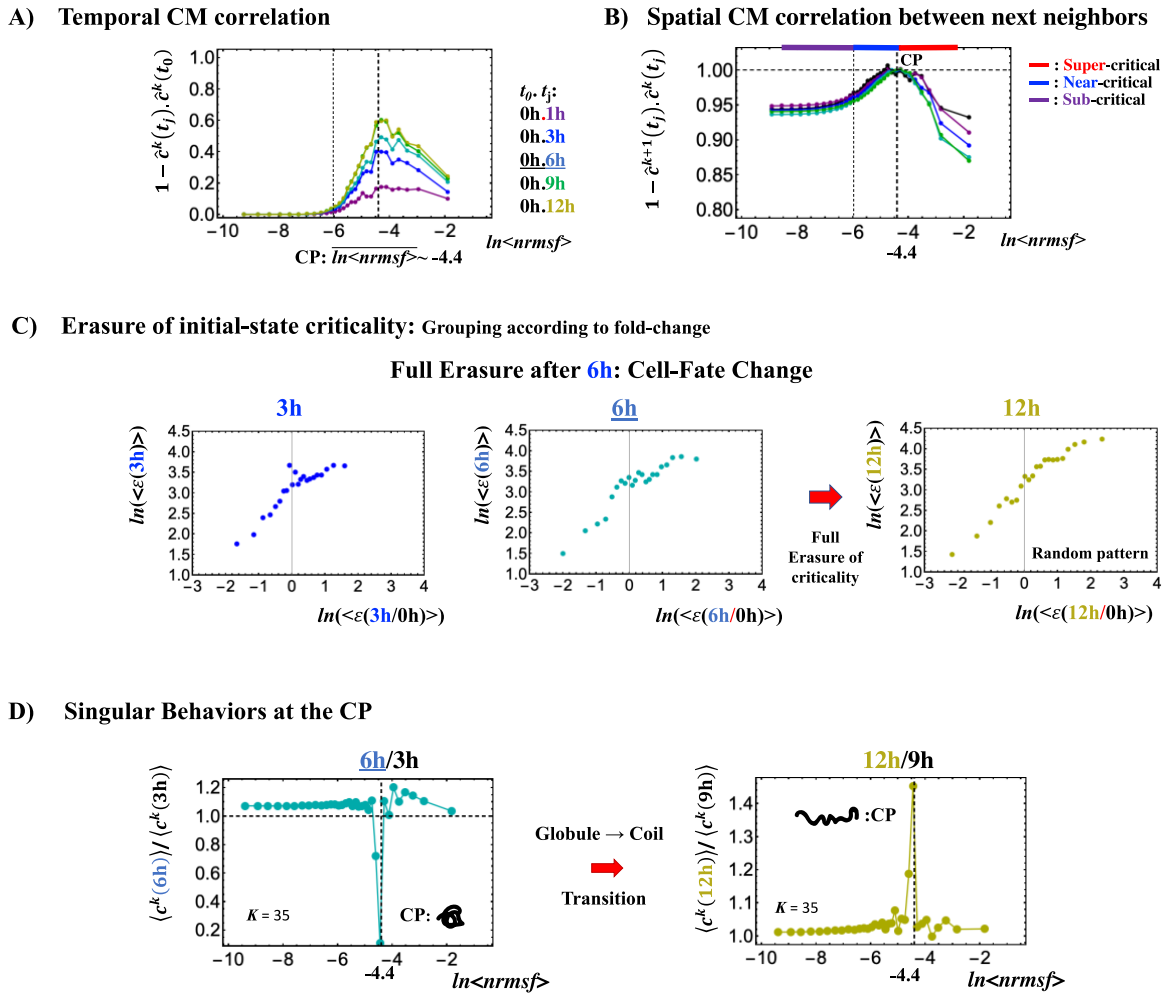


Figure 5: Th17 immune cell response (single cell). Regarding the features of the CP, **A)** and **B)** are the same as the embryo cases, having only one difference in that the CP never passes the zero-temporal correlation. The timing of cell-fate change after 6h is demonstrated by **C)** erasure of the initial-state (t = 0) criticality. This timing is further supported by the same timing occurrence as seen in the switch from suppression to enhancement on the genome engine (**Figures 9E,F**). **D)** Singular behavior shown in fold change in neighboring time points is sharply peaked (as in cell-population for cell differentiation: **Figure 3D**) at the CP, implying the occurrence of micro-segregations on the higher-order structure between swelled coil and compact globule near the CP (K = 35 groups with n = 375 RNAs). The sensitivity according to the number of groups is observed as occurring in the cell population (**Figure 3D**).

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

C. Systematic Determination of Critical States Exhibiting Coherent-Stochastic Behaviors

We demonstrate that the CP is a fixed point relative to a given biological regulation. Next, based on this fact, we show that critical states in genome expression can be determined systematically for both single cell and cell population genome expression:

1) **Single cell level**: both temporal and spatial CM correlations (**Figures 4A,B, 5A,B**) manifest distinct response domains according to *nrmsf*: low-variance expression (sub-critical state) for region of flattened correlation (region of perfect correlation), intermediate-variance for near-critical state from the edge of flattened correlation to the CP, and high-variance expression for super-critical state above the CP (summary: **Table 1**).
2) **Cell population level**: As shown in **Figure 5**, for both MCF-7 and HL-60 cancer population cells, the Euclidian distance (from the highest *nrmsf* group) between the temporal responses of CM grouping (**Figure 1**: initial state response: t = 0 vs. other experimental time points) reveals critical states (summary: **Table 2**), where the CP exists at the boundary between near-and sub-critical states (vs. between super-and sub-critical states in single cell). In our previous studies on microarray data (cell population level), critical states were determined by the transition of expression profile by means of Sarle's bimodality coefficient putting in evidence of distinct response domains (super-, near- and sub-critical domains according to temporal variance of expression) were evident in genome expression [Tsuchiya, M., et al., 2016].

These critical states possess coherent-stochastic behaviors - coherent behavior emerged from an ensemble of stochastic expression [Tsuchiya, M., et al., 2016, 2017]. To capture this behavior, we applied a bootstrap simulation approach to examine if i) the CM represents critical state through the convergence of the CM of randomly selected gene ensembles (hundreds of repetitions) to that of critical state as the number of elements (*n*) is increased and if ii) mixed states between critical states does not show the convergence (see Figs. 3B,C in [Tsuchiya, M., et al., 2017]). Mixing state does not converge to the CM of the critical state due to distinguished coherent dynamics of critical states. The bootstrap approach can even reveal stochastic behavior of gene expression within a critical state with low correlation convergence between randomly selected gene ensembles as the number of elements (*n*) is increased. Here, the existence of a threshold *n* at around 50 randomly picked genes [Censi, F., et al., 2011; Tsuchiya,

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

M., et al., 2015], which allows to reproduce the properties of the critical state with a random choice of N genes with $N > n$, is a further proof of the reliability of coherent-stochastic behaviors.

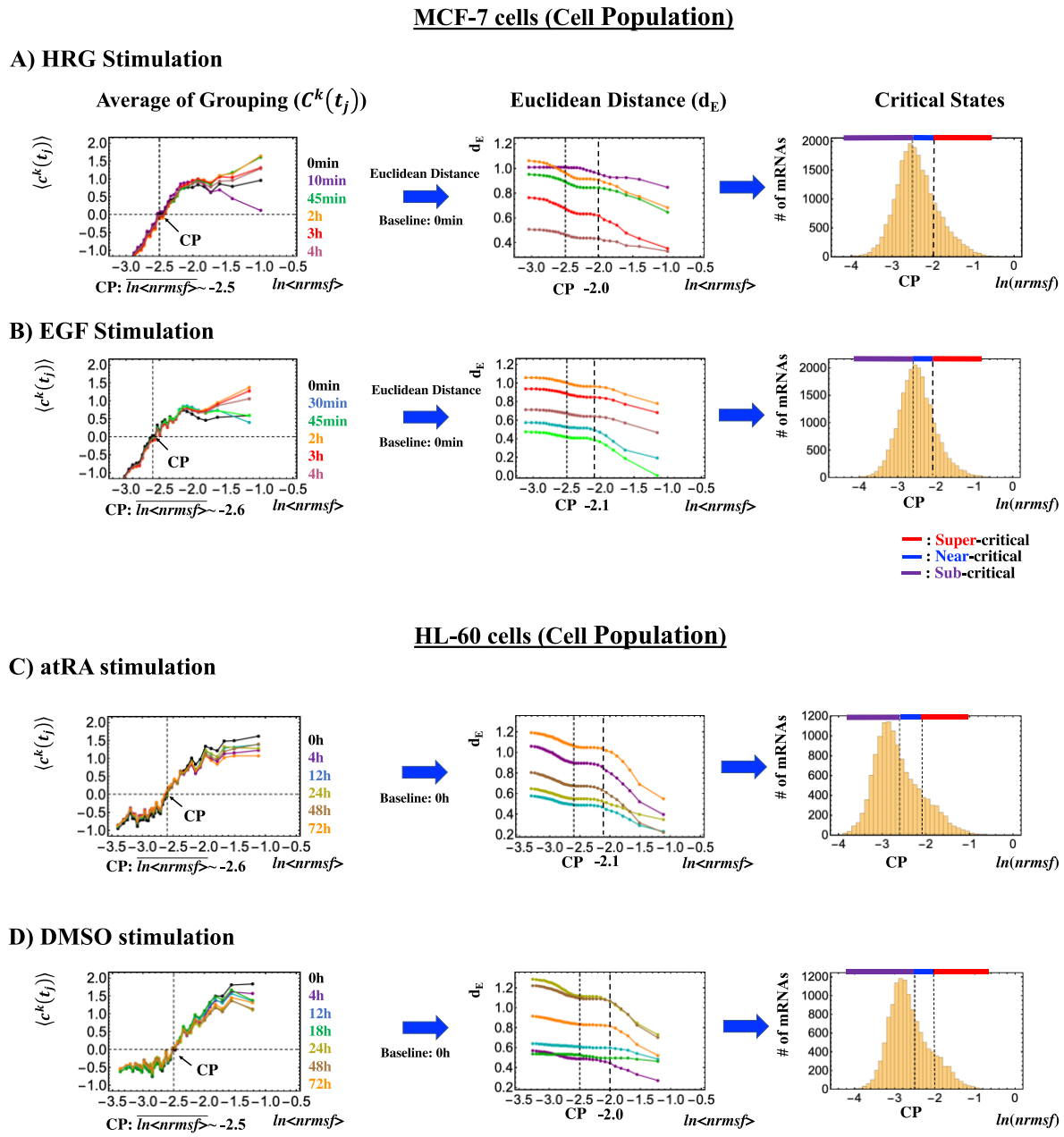


Figure 6: Systemic determination of critical states for cell-population (microarray data). MCF-7 cells: A) HRG-stimulated and **B)** EGF-stimulated MCF-7 cells; **C)** atRA-stimulated and **D)** DMSO-stimulated HL-60 cells; first column: the temporal response of the CM group (refer to **Figure 1**), second column: the Euclidian distance between two ensemble averages, $<\boldsymbol{c}^k(t=0)>$ and $<\boldsymbol{c}^k(t_j)>$ (k = 1, 2,..,K = 25) (see **Figure 1**) from higher *nrmsf* revealed three distinct response behaviors (critical states) with a boundary indicated by dashed vertical lines. Third column: corresponding region of critical states (red: super-critical; blue: near-critical; purple: sub-critical) in a histogram of gene expression according

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to *ln*(*nrmsf*). **Note**: The location of the CP lies at the boundary between near- and sub-critical states, different from the case with single cell (**Figures 4, 5**).

D. A Universal Genome-Engine Mechanism for SOC-Control of Genome Expression

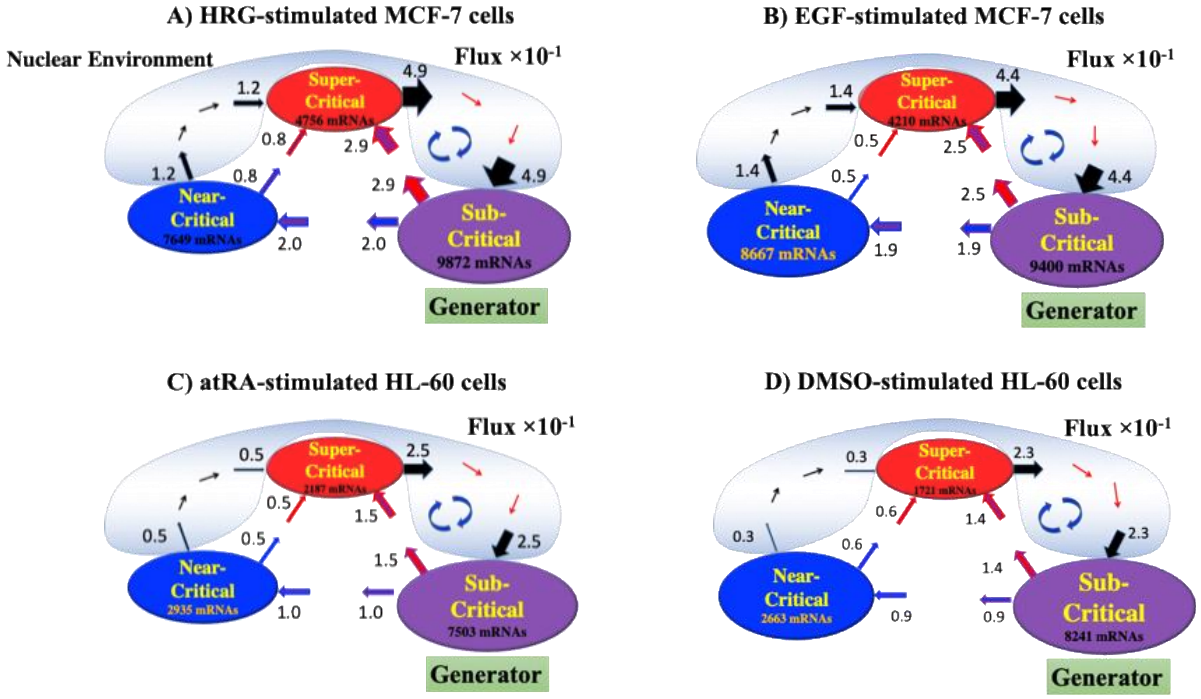
The existence of distinct critical states with fixed critical point suggests that self-organizing principle of genome expression is the SOC control of overall expression for both population and single cell levels; this points to the existence of a universal mechanism - genome-engine mechanism for autonomous SOC control of genome expression from embryo to cancer development.

Distinct coherent dynamics in critical states emerge from stochastic expression and the CM of critical state represents its dynamics (Fig. 10 in [Tsuchiya, M., et al., 2016]). Thus, dynamic expression flux analysis [Tsuchiya, M., et al., 2016, 2018; **Methods**] to the CM of critical states can apply to reveal the genome-engine mechanism for describing how autonomous SOC control of genome expression occurs. **Figure 7** shows that the sub-critical state acts as internal source of expression flux and the super-critical state as a sink. Sub-Super cyclic flux forms a dominant flux flow that generates a strong coupling between the super- and sub-critical states accompanied by their anti-phase expression dynamics. This results in making its change in oscillatory feedback, and thus sustains autonomous SOC control of overall gene expression. The formation of the dominant cyclic flux provides a universal *genome-engine metaphor* of SOC control mechanisms pointing to a universal mechanism in gene-expression regulation of mammalian cells for both population and single cell levels. Global perturbation, which enhances or suppresses the genome engine (see **Section IIE**), induces the cell-fate change. Thus, molecular-based elucidation of when and how global perturbation induces cell fate is expected to provide know-how of cell-fate control in a precise manner.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Average Between-State Flux

Population Cell



Single Cell

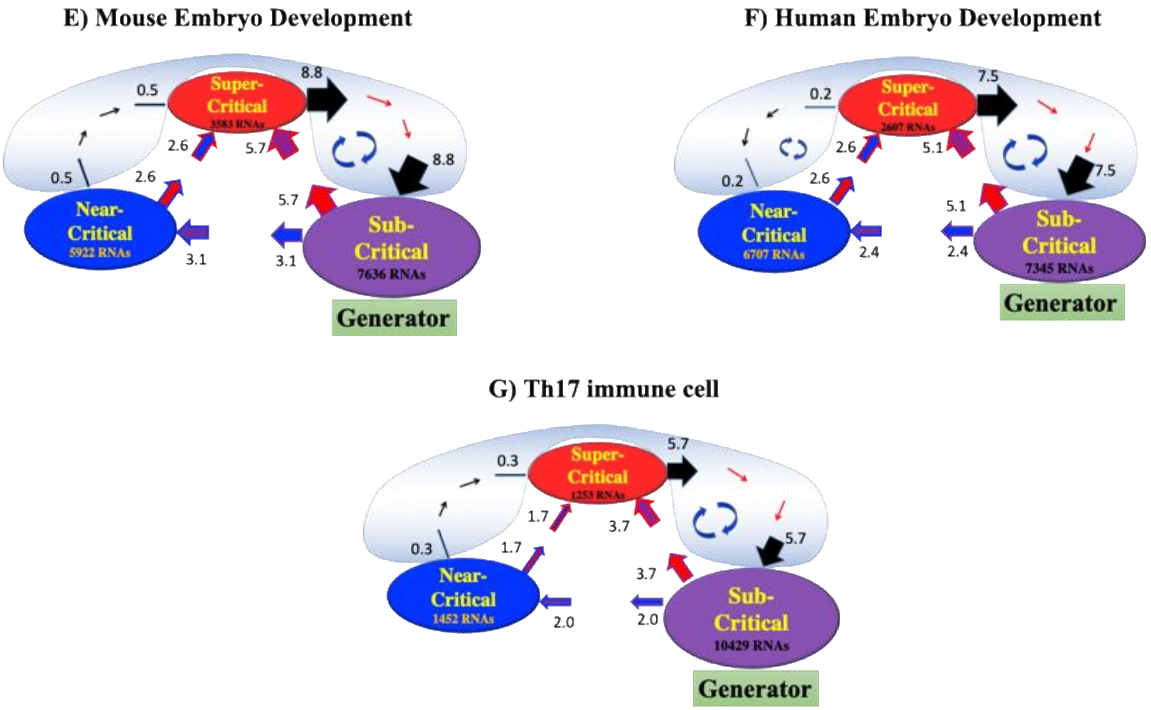


Figure 7: Genome engine mechanism revealed through average between-state expression flux. **Cell-Population: A)** HRG-stimulated and **B)** EGF-stimulated MCF-7 cancer cells; **C)** atRA-stimulated and **D)** DMSO-stimulated HL-60 cancer cells. **Single-Cell: E)** Mouse- and **F)** human-embryo development; **G):** Th17 immune cell. **Genome engine mechanism** [Tsuchiya, M., et al., 2016]: Sub-Super cyclic flux forms a dominant flux flow to establish the genome engine mechanism. The sub-

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

critical state acting as a "large piston" for short moves (low-variance expression) and the super-critical state as a "small piston" for large moves (high-variance expression) with an "ignition switch" (near-critical state with a critical point) are connected through a dominant cyclic state flux as a "camshaft", resulting in the anti-phase dynamics of two pistons. Numerical values represent average between-state expression flux, whereas in the cell population, the values are based on 10^{-1}.

E. Cell-Fate Change Through Global Perturbation on the Genome Engine

Here, to describe genome engine mechanism on genome expression, the key fact is that dynamics of distinct coherent behavior emerged from stochastic expression in critical states (coherent-stochastic behavior) follows the dynamics of the CM of a critical state. Based on this fact, the expression flux approach (**Methods**) was developed to reveal dynamic interaction flux between critical states [Tsuchiya, M., et al., 2016, 2017], especially how the SOC control of the whole gene expression evolves dynamically through perturbation to reveal genome engine mechanism for cell-fate change. Interaction flux between-states serves as the underlying basic mechanism of epigenetic self-regulation through incorporating a rich variety of transcriptional factors and non-coding RNA regulation to determine coherent oscillatory behaviors of critical states.

The flux dynamics approach is further developed to analyze quantitative evaluation of the degree of non-harmonicity and time reversal symmetry breaking in nonlinear coupled oscillator systems [Tsuchiya, M., et al., 2018].

Intersection of interaction fluxes (**Figure 8** for cell population and **Figure 9** for single cell) occurs just before cell-fate change. This shows that around the cell-fate change, a global perturbation induces enhancement or suppression on the genome engine, where there is a dominant cyclic flux flow between the super- and sub-critical states (**Table 3**: single cell; **Table 4**; cell population). In HL-60 cells (cell population), the genome-engine is enhanced before the cell-fate change and suppressed (enhancement-suppression) thereafter. On the contrary, a reverse process of suppression- enhancement takes place in the MCF-7 cancer cells (see Fig. 12 in [Tsuchiya, M., et. al., 2016]). In the single cell cases with embryo development and Th17 immune cell, a suppression-enhancement on the genome engine occurs. The different sequences of perturbation on the genome-engine may stem from different stages of the suppressive pressure on cell-differentiation against cell-proliferation.

Regarding *the activation of the CP for single cell*, it clearly shows that a global perturbation (**Figures 10A, B**) occurs on the genome-engine (at late 2-cell, 8-cell and 6h for mouse, human embryo, and Th17 cell, respectively), where the timing coincides with that of

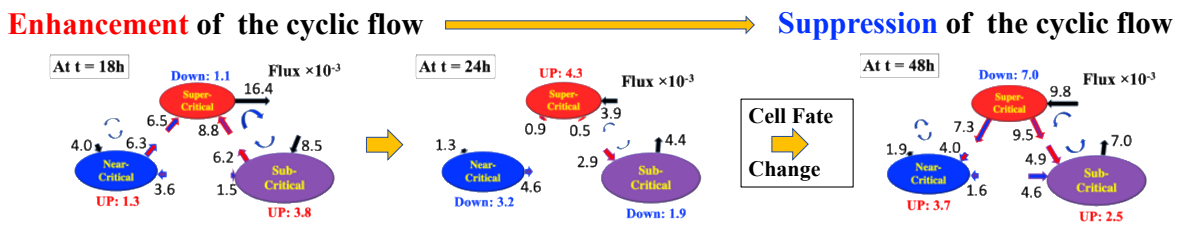
bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the cell-fate change (i.e., erasure of the initial-state CP memory). Hence, this suggests that in regard to the activation of the CP for single cell, the activation of the CP points to when the initial-state sandpile criticality erases: before and after the cell-fate change, inversion of singular behaviors at the CP occurs for embryonic reprogramming, and the transition from compact globule to swelled coil at the CP for cell-differentiation. Furthermore, as for embryo reprogramming, temporal CM correlation of the CP from initial state (zygote) passes zero-correlation (i.e., erasure of the initial-state CP memory). The activation of the CP (the edge of the criticality) provokes a global impact on the entire genome expression through the genome-engine perturbation (refer to Fig.13 and Discussion in [Tsuchiya, M., et al., 2016]). This global impact provoked by the CP is further supported by the fact that i) EGF-stimulated MCF-7 cells, where the CP is OFF with no cell-differentiation, induce only local perturbation (**Figure 10C**) and ii) HL-60 cells coincides with the timing of global perturbation, notably when the CP is ON at 12h-18h for DMSO and 24h-48h for atRA cells (**Figure 10D**).

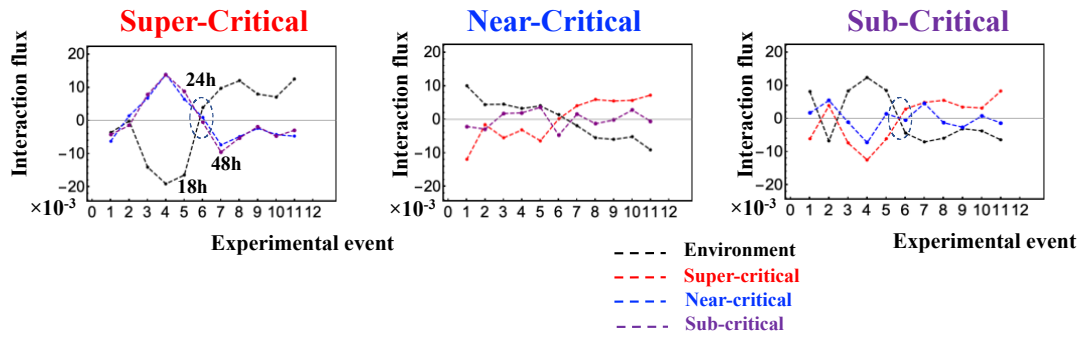
The genome-engine in control of dynamics of whole genome expression demonstrates how perturbation on the sub-critical state (generator of SOC-control) affects the entire genomic system. Therefore**, the elucidation of molecular mechanism underpinning the activation of the CP (involving more than hundreds of genes) through a transition of a compact globule to a swelled coil state is expected to uncover how the genome-engine is either enhanced or suppressed around the cell-fate change. Furthermore, this elucidation could predict when and how a cell-fate change occurs (i.e., a novel cell-fate control mechanism for cancers, iPS cells, stem cells and so forth)**.

atRA-stimulated HL-60 cells

A) Flux Dynamics around Cell-Fate Change



B) Interaction Expression Flux

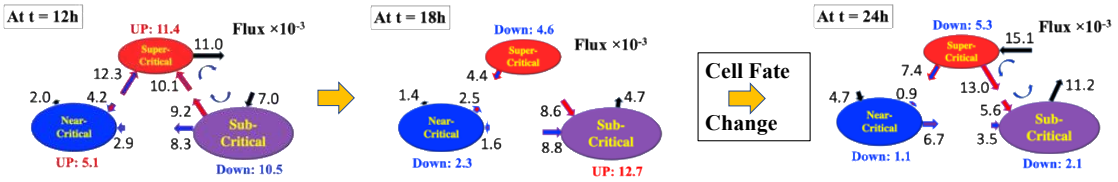


bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DMSO-stimulated HL-60 cells

C) Flux Dynamics around Cell-Fate Change

Enhancement of the cyclic flow → **Suppression of the cyclic flow**



D) Interaction Expression Flux

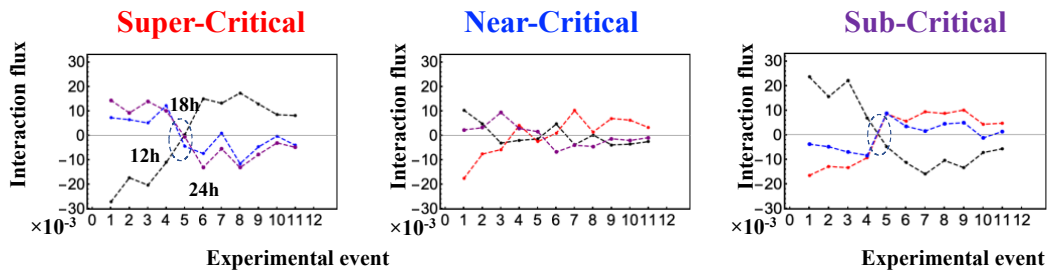
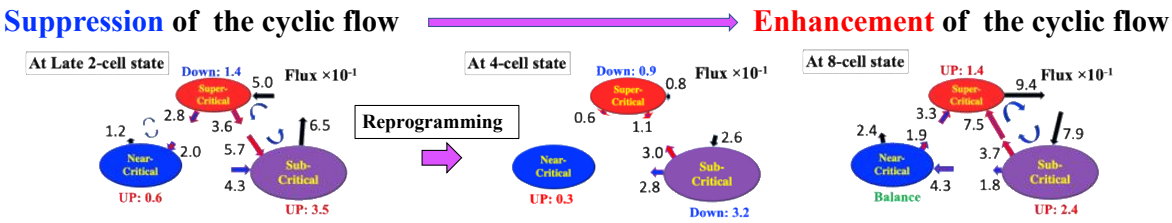


Figure 8: Cell-population - Cell fate mechanism through perturbation on the genome engine. For **A)** and **B)** the interaction flux dynamics (black: environment; red: super-critical state; blue: near-critical state; purple: sub-critical state) around cell-fate change show that cell-fate change in atRA-stimulated HL-60 cells occurs through enhancement (at 18h before the cell-fate change) - suppression (48h after the cell-fate) on the dominant cycle state flux (**Figure 7C**). Before the cell-fate change, the interaction flux dynamics (indicated by dashed cycle) intersect each other at 24h. This intersection (see also **Figures 9B,D,F** for single cell) marks the onset of the switch to coherent dynamics on the dominant cyclic flow (genome-engine). For **C)** and **D)**, the DMSO-stimulated HL-60 cells undergo an enhancement-suppression phase on the dominant cyclic flux at 12h (before the cell-fate change) and 24h (after the cell-fate change), respectively (**Figure 7D**). The intersection of interaction flux occurs around at 18h before the cell-fate change. Numerical values based on 10^{-3} represent interaction flux and any value less than 0.5 in the interaction flux is not shown; net flux (self-flux) of interaction fluxes at each critical state should be balanced. Note: For the genome-engine in HRG-stimulated MCF-7 cells, refer to details in [Tsuchiya, M., et al., 2016].

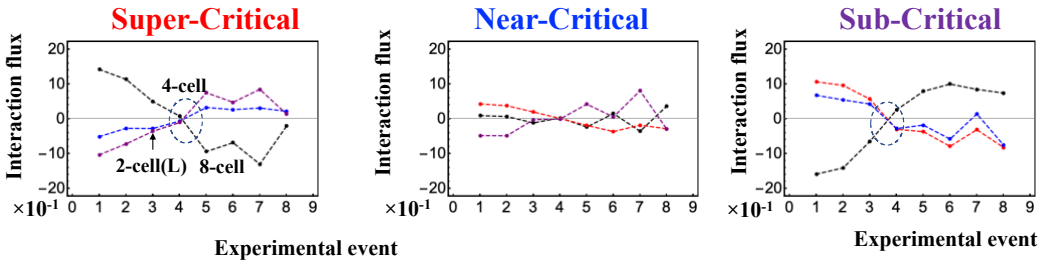
bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mouse Embryo Development

A) Flux Dynamics around Cell-Fate Change

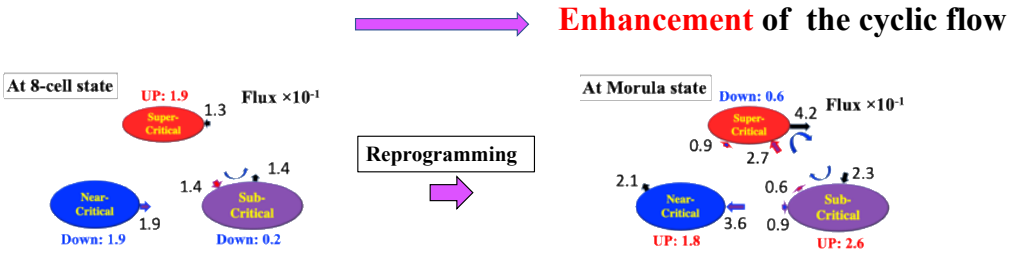


B) Interaction Expression Flux

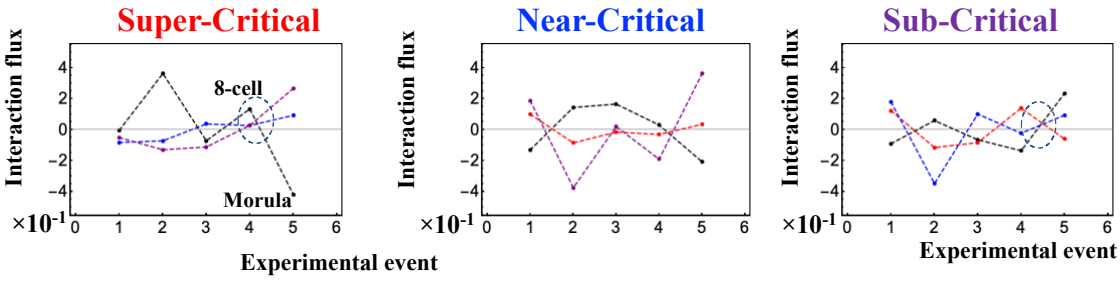


Human Embryo Development

C) Human Embryo Development



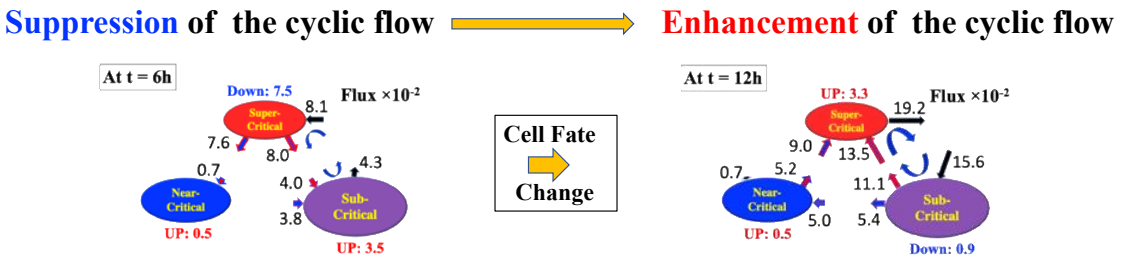
D) Interaction Expression Flux



bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Th17 Cell Differentiaiton

E) Th17 immune cell



F) Interaction Expression Flux

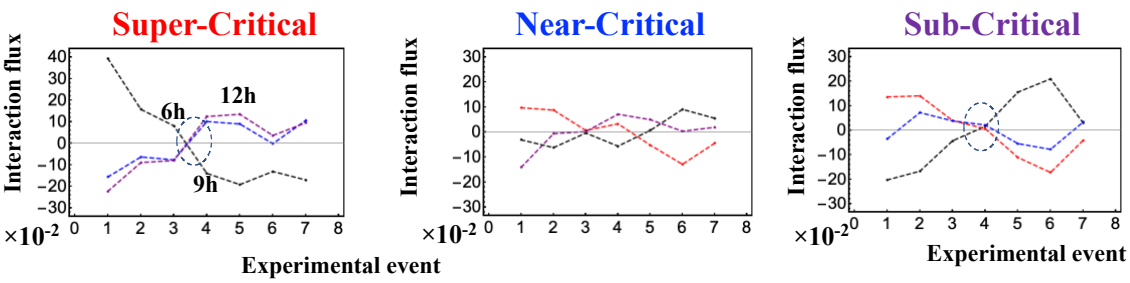


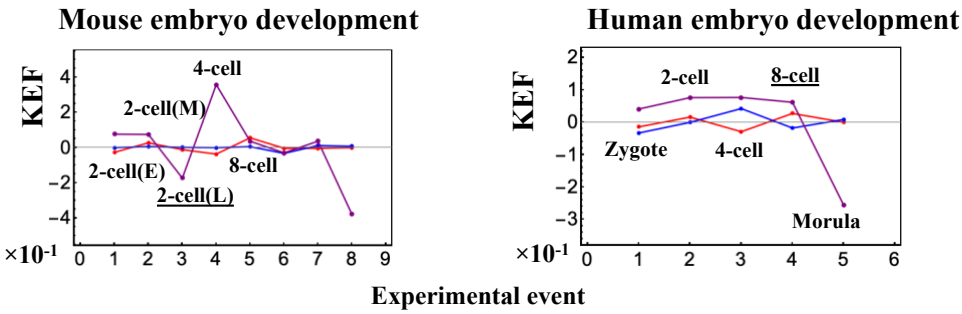
Figure 9: Single cell - Cell fate mechanism through perturbation on the genome engine. A) and **B)** Mouse embryo development**; C)** and **D)** Human embryo development; **E)** and **F)** Th17 immune cell differentiation. Notably, reprogramming on embryo development (more evident on mouse embryo) and cell differentiation on Th17 immune cell occurs through the suppression-enhancement on the dominant cycle state flux between super- and sub-critical states, which is opposite to the cell differentiation in HL-60 cells; for HRG-MCF-7 cell response refer to Fig. 12B in [Tsuchiya, M., et al., 2016]. Numerical values based on 10^{-1} for embryo cell and 10^{-2} for Th17 cell represent between-state expression flux, and are not shown for less than 0.5 value in interaction flux.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

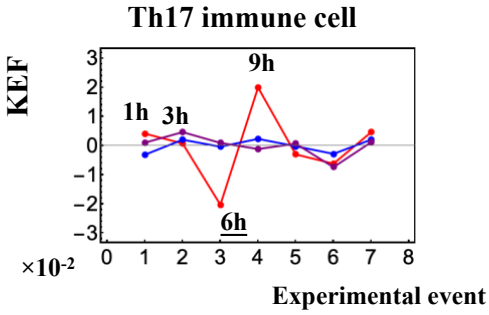
Kinetic Energy Flux (KEF)

Single Cell

A) Embryo Development

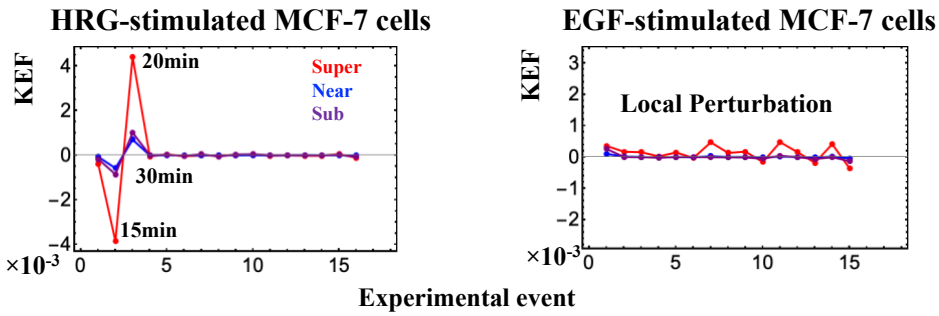


B) Immune Cell



Cell Population

C) MCF-7 Cells



D) HL-60 Cells

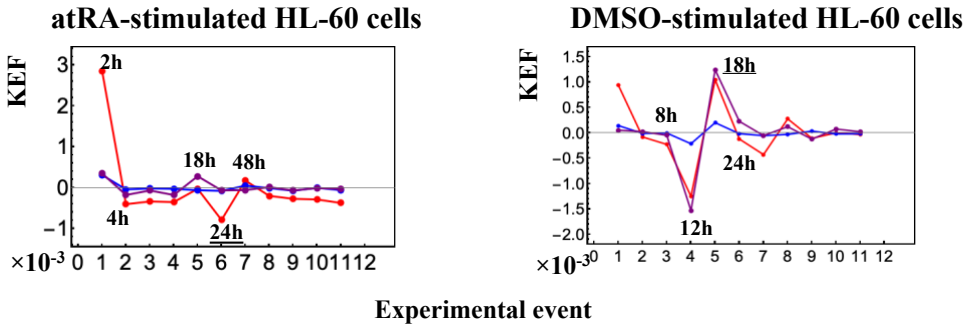


Figure 10: Kinetic Energy Flux (KEF). Global perturbation on genome-engine is well-manifested in the kinetic energy flux (**Methods**) examining single-cell responses in **A):** Embryo development and **B):** Immune cell, and population responses in **C):** MCF-7 cells and D): HL-60 cells. At the single cell level, the timing of global perturbation (involvement of more than one critical state) coincides with that of cell-fate change (**Figures 4C, 5C**).

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

At the population level, kinetic energy flux is damping. In MCF-7 cells (**C**), the timing of the activation of the CP (ON: 10-15min: **Figure 2A**) is within the global perturbation (10-30min) in the HRG-stimulation (intersection of interaction flux at 20-30min in Fig [Tsuchiya, M, et al., 2016], whereas in the EGF-stimulation (no cell-differentiation [Saeki, Y., et al., 2009]), only local perturbation occurs. On the other hand, in HL-60 cells (**D**), the timing of the perturbation on the genome engine (at 18h-24h-48h intervals) corresponds to the end of the damping (second largest perturbation) for atRA-stimulated HL-60 cells and the largest global perturbation (at 12h-18h-24h intervals) for DMSO-stimulated HL-60 cells. Underlined numbers represent the timing just before the cell-fate change.

III. Discussion

In this report (dealing with different systems from embryo to cancer development), we demonstrated the existence of a shared underlying genomic mechanism for cell-fate decision. The key finding is that the CP (critical gene set) corresponds to the center of mass (CM) of genome expression and reveals fixed point behavior according to temporal expression variance (*nrmsf*). On the basis of this fact, the temporal CM correlation, correlation between different genome expressions (baseline the CM of genome expression) in time, reveals active or inactive state of the CP in cell population. On the other hand, in single cell level, activation of the CP for cell-fate change corresponds to the timing of erasure of initial-state sandpile criticality. In embryonic reprogramming, the erasure of the zygote-state criticality points to the complete erasure of the memory of the zygote CP through reprogramming.

This state of the CP is supported by singular behavior of the CP: cell differentiation for both cell population (**Figure 3D)** and single cell (**Figure 5D**) exhibits clear swelled coil (ON) - compact globule (OFF) transition of the CP, whereas in embryonic reprogramming (single cell: **Figure 4D**) inversion of singular behaviors of the CP is revealed, where inversion of the intra-chain phase segregation of coil and globule states exhibits. These singular behaviors are clear transitional behaviors, which are closely concerned with the intrinsic characteristics of first-order phase transition inherent in genome sized DNA molecules. The activation of the CP is essential for the occurrence of cell-fate change; conversely inactivation of the CP does not undergo cell-fate change.

Regarding mega base-pairs size of DNA phase transition, up until the late 20th century, it had been believed that a single polymer chain, including DNA molecule, always exhibit cooperative but mild transition between elongated coil and compact globule, which is neither a first-order nor a second-order phase transition [Flory, P., 1953; Gennes, P.G. 1979]. Nowadays, it has become evident that long DNA molecules above the size of several tens of kilo base-pairs exhibit the characteristics to undergo large discrete transition, i.e., first-order property on coil-

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

globule transition [Yoshikawa, K, et al., 1996; Yoshikawa, K and Yoshikawa, Y., 2002; Zinchenko A, et al., 2008]. Such first-order characteristics are rather general for a semi-flexible polymer, especially polyelectrolyte chains such as giant DNA molecules. It is also noted that insufficient charge neutralization on the globule state causes instability and leads to the generation of intra-chain segregation [Sakaue, T., Yoshikawa, K. 2006; Shew, C.Y., Yoshikawa, K., 2007] (see e.g., **Figure 4D**). When such instability with long-range Columbic interaction is enhanced, the characteristic correlation length tends to become shorter, corresponding to the generation of the critical state in the transition. Such behavior accompanied with the folding transition of DNA is also observed for reconstructed chromatin [Schiessel, H., 2003; Nakai, T., et al., 2005; Suzuki, Y., et al., 2011].

In relation to the classical concept of SOC (**see Introduction**), the activation and inactivation of the CP suggests that there may be another layer of a macro-state (genome state) composed of distinct micro-critical states (found by us). The activation of the CP makes the genome-state to be considered 'super-critical' to guide the cell-fate change - super-critical after activation of the CP: HRG-stimulated MCF-7 cells, and DMSO- and atRA-stimulated HL-60 cell. Whereas, prior to the activation of their CPs, the genome states are considered 'sub-critical', and the genome state of EGF-stimulated MCF-7 cells remains 'sub-critical' (no cell-fate change) all the way.

These findings on the CP allow for a systematic determination of critical states for both cell population and single cell regulation. This implies that expression flux analysis among critical states through the cell nuclear environment provides a potential universal model of self-organization as the genome-engine mechanism, where a highly coherent behavior of low-variance genes (sub-critical state) generates a dominant cyclic expression flux with high-variance genes (super-critical state) to develop autonomous critical control system. This explains the coexistence of critical states (distinct expression response domains) through critical point; *nrmsf* (temporal expression variance) acting as the order parameter for the self-organization - *self-organized criticality (SOC) control of genome expression*. It is intriguing to see that the location of the CP moves from a higher *nrmsf* value in single cell to a lower value in cell population: the CP locates at the boundary between super- and near-critical states in a single cell, whereas in cell population, the CP exists at the boundary between near- and sub-critical states.

The genome-engine mechanism rationalizes how the change in criticality affects the entire genome expression to drive cell-fate change. This driving happens through enhancement (before cell-fate change) - suppression (after cell-fate change) or suppression-enhancement on the genome-engine. This different sequence of perturbation might stem from different stage of

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

repressive pressure against cell-differentiation while competing with cell-proliferation. Regarding genome-reprogramming, our results provide further insight of the reprogramming event in mouse embryo development [Tsuchiya, M., et al., 2017]: the activation of the CP from OFF to ON state changes the genome-engine from a suppressed to enhanced state, and this drives the genome to pass over a critical transition state (SOC landscape) right after the late 2-cell state (note: regarding critical state transition in single cell level, also see [Mojtahedi, M., et al., 2016]). The genome-engine suggests that the activation mechanism of the CP should elucidate how the global perturbation occurs on self-organization through change in signaling by external or internal stimuli into a cell. Recent study shows that the dynamics of high order structure of chromatin exhibits liquid like behavior [Maeshima, K., et al., 2016], which could be crucial characteristic in enabling the genome to conduct SOC gene expression control for cell-fate determination.

Further studies on these matters are needed to clarify the underlying fundamental molecular mechanism, and the development of a theoretical foundation for the autonomous critical control mechanism in genome expression as revealed in our findings is expected to open new doors for a general control mechanism of the cell-fate change and genome computing (see Discussion in [Tsuchiya, M, et al, 2015, 2016]), i.e., the existence of '*genome intelligence*'.

As for now, we can safely affirm that the strong interaction among genes with very different expression variance and physiological roles push for a complete re-shaping of the current molecular-reductionist view of biological regulation looking for single 'significantly affected' genes in the explanation of the regulation processes. The view of the genome acting as an integrated dynamical system is here to stay.

IV. Methods

Biological Data Sets

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

Cell population:

1. 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 [Saeki Y, et al., 2009]) at 18 time points: $t_1 = 0$, $t_2 = 10$,15, 20, 30, 45, 60, 90min, 2, 3, 4, 6, 8, 12, 24, 36, 48, $t_{T=18} = 72$h.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Microarray data of the induction of terminal differentiation in human leukemia HL-60 cells by DMSO and atRA; GEO ID: GSE14500 (N = 12625 mRNAs; details in [Huang, S., et al., 2005]) at 13 time points: t_1= 0, t_2 = 2, 4, 8, 12, 18, 24, 48, 72, 96, 120, 144, $t_{T=13}$ =168h.

Single cell:

3. 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) with experimental details in [Yan, L., et al., 2013] and [Deng, Q., et al., 2014], respectively.

We analyzed 7 human and 10 mouse embryonic developmental stages listed below:

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.

4. RNA-Seq data of T helper 17 (Th17) cell differentiation from mouse naive CD4+ T cells in RPKM (Reads Per Kilobase Mapped) values, where Th17 cells are cultured with anti-IL-4, anti-IFNγ, IL-6 and TGF-β, (details in [Ciofani, M., et al., 2012]; GEO ID: GSE40918 (mouse: N = 22281 RNAs) at 9 time points: t_1 = 0, t_2 = 1, 3, 6, 9, 12, 16, 24, $t_{T=9}$ = 48h. For each time point, the reference sample numbers are listed: GSM1004869-SL2653 (t= 0h); GSM1004941-SL1851 (t =1h); GSM1004943-SL1852 (t = 3h); GSM1005002-SL1853 (t= 6h); GSM1005003-SL1854 (t= 9h); GSM1004934-SL1855 (t= 12h); GSM1004935,6,7-SL1856, SL8353, SL8355 (t= 16h; average of three data); GSM1004942-SL1857 (t = 24h); GSM1004960-SL1858 (t= 48h).

In reference to the colors used in the various plots throughout this report, they are based on the experimental events and have been assigned as the following: black as the initial event, purple as the 2nd event, and the subsequent events as blue, dark cyan, dark green, dark yellow, brown, orange, red, dark pink, and pink.

For microarray data, the Robust Multichip Average (RMA) was used to normalize expression data for further background adjustment and to reduce false positives [Bolstad, B. M., et al., 2003; Irizarry, R. A., et al, 2003; McClintick, J. N., Edenberg, H. J., 2006].

For RNA-Seq data, RNAs with RPKM values of zero over all of the cell states were excluded.

In the analysis of sandpile criticality, random real numbers in the interval [0, a] generated from

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a uniform distribution were added to all expression values (only in **Figures 4C,5C**). This procedure avoids the divergence of zero values in the logarithm. The robust sandpile-type criticality through the grouping of expression was checked by changing a positive constant: a ($0 < a < 10$); we set $a = 0.001$. Note: The addition of large random noise ($a >> 10$) destroys the sandpile CP.

Normalized Root Mean Square Fluctuation (nrmsf)

Nrmsf (see more Methods in [Tsuchiya, M., et al., 2015]) is defined by dividing *rmsf* (root mean square fluctuation) by the maximum of overall $\{rmsf_i\}$:

$$rmsf_i = \sqrt{\frac{1}{S}\sum_{j=1}^{S}\left(\varepsilon_i(s_j) - \langle\varepsilon_i\rangle\right)^2}, \quad (1)$$

where $rmsf_i$ is the *rmsf* value of the ith RNA expression, which is expressed as $\varepsilon_i(s_j)$ at a specific cell state s_j or experimental time (e.g., in mouse embryo development, $S = 10$: s_1 = zygote, early 2-cell, middle 2-cell, late 2-cell, 4-cell, 8-cell, morula, early blastocyst, middle blastocyst and s_{10} = late blastocyst), and $\langle\varepsilon_i\rangle$ is its expression average over the number of cell states. Note: *nrmsf* is a time-independent variable.

CM correlation analysis

To investigate the transition dynamics, the correlation metrics based on the center of mass (CM) grouping, the *CM correlation* is built upon the following basic statistical formalization:

1) CM grouping: genome expression is considered as a N-dimensional vector, where each expression is subtracted by the average value of the whole expression at $t = t_j$. Next, the whole expression is sorted and grouped according to the degree of *nrmsf*, where CM grouping has K groups and within each group there are n number of expressions ($N = K{\cdot}n$): N-dimensional CM grouping vector, $\boldsymbol{C}(t_j) = \left(\boldsymbol{c}^1(t_j), \boldsymbol{c}^2(t_j), \ldots, \boldsymbol{c}^k(t_j), \ldots, \boldsymbol{c}^K(t_j)\right)$; $\boldsymbol{c}^1(t_j)$ and $\boldsymbol{c}^K(t_j)$ are the highest and lowest group vectors of *nrmsf*, respectively. Here, the unit vector of kth vector $\boldsymbol{c}^k(t_j)$ is defined as $\hat{\boldsymbol{c}}^k(t_j) = \frac{\boldsymbol{c}^k(t_j)}{|\boldsymbol{c}^k(t_j)|}$. Note that the less than n elements in the last group (the lowest *nrmsf*) have been removed from the analysis.

2) Keeping in mind, correlation corresponds to the cosine of angle between unit vectors, i.e., inner product of unit vectors ($\cos\theta = \hat{\boldsymbol{a}}.\hat{\boldsymbol{b}}$, θ: angle; $\hat{\boldsymbol{a}}.\hat{\boldsymbol{b}}$: dot product (scalar) of unit vectors: $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{b}}$).

Two different CM correlations are considered:

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

i) **Spatial CM correlation**: for a given time point ($t = t_j$), development of CM correlation between the first group (highest *nrmsf* group) and other vectors: $\hat{\boldsymbol{c}}^1(t_j).\hat{\boldsymbol{c}}^k(t_j)$; ($k$ = 2,3,..K).
ii) **Temporal CM correlation**: for a given group (k), development of CM correlation between the initial and other experimental points: $\hat{\boldsymbol{c}}^k(t_1).\hat{\boldsymbol{c}}^k(t_j)$ (k= 1,2,3,..K) over experimental time points, t_j (see **Biological Data Sets**).

Expression Flux Analysis

Here, to describe genome engine mechanism on both single cell and population genome expression, the key fact is that dynamics of coherent behavior emerged from stochastic expression in distinct critical states (coherent-stochastic behavior: CSB) follows the dynamics of the CM of a critical state. We have developed the expression flux approach to reveal dynamic interaction flux between critical states [Tsuchiya, M., et al., 2016-2018].

The CSB in a critical state corresponds to the scalar dynamics of its CM. The numerical value of a specific critical state (i.e., super-, near- or sub-critical state) is represented by $X(s_j)$ at a specific experimental event (s_j), where an experimental event (s_j) corresponds to a cell state or an experimental time point. The expression flux between critical states is interpreted as a non-equilibrium system and evaluated in terms of a dynamic network of effective forces, where interaction flux is driven by effective forces between different critical states and can be described by a second-order time difference. It is important to note that the oscillatory phenomenon interpreted using a second-order difference equation with a single variable is equivalent to inhibitor-activator dynamics given by a couple of first-order difference equations with two variables. Flux dynamics approach is further developed to analyze quantitative evaluation of the degree of non-harmonicity and time reversal symmetry breaking in nonlinear coupled oscillator systems [Tsuchiya, M., et al., 2018].

Basic formulas of expression flux dynamics are given as follows:
Net self-flux of a critical state: The net self-flux, the difference between the IN flux and OUT flux, describes the effective force on a critical state. This net self-flux represents the difference between the positive sign for incoming force (net IN self-flux) and the negative sign for outgoing force (net OUT self-flux); the CM from its average over all cell states represents up- (down-) regulated expression for the corresponding net IN (OUT) flux.
The effective force is a combination of incoming flux from the past to the present and outgoing flux from the present to the future cell state:

$$f\left(X(s_j)\right) = \frac{\Delta P}{\Delta s} = \frac{1}{\Delta s}\left\{\frac{\left(X(s_j)-X(s_{j-1})\right)}{\Delta s_j} - \frac{\left(X(s_{j+1})-X(s_j)\right)}{\Delta s_{j+1}}\right\} - \langle f(X)\rangle$$

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$= (\text{IN flux} - \langle \text{IN flux} \rangle) - (\text{OUT flux} - \langle \text{OUT flux} \rangle), \qquad (2)$$

where ΔP is the change in momentum with a unit mass (i.e., the impulse: $F\Delta s = \Delta P$) and natural log of average (<…>) of a critical state, $X(s_j) = ln\,\langle \frac{1}{N_C}\sum_{i=1}^{N_C} \varepsilon_i(s_j)\rangle$ with the ith expression $\varepsilon_i(s_j)$ at the jth experimental event, $s = s_j$ (N_C = the number of RNAs in a critical state; refer to **Tables 1,2**); the average of net self-flux over the number of critical states, <*f*(X)> = <INflux> - <OUTflux>.

Here, scaling and critical behaviors occur in log-log plots of group expression, where the natural log of an average value associated with group expression such as *ln<nrmsf>* and *ln<expression>* is taken. Thus, in defining expression flux, the natural log of average expression (CM) of a critical state is considered.

It is important to note that each embryo cell state is considered as a statistical event (note: a statistical event does not necessarily coincide with a biological event) and its development as time arrow (time-development) when evaluating the average of group expression: fold change in expression and temporal expression variance (*nrmsf*). This implies that an interval in the dynamical system (Equation (2)) is evaluated as difference in event, i.e., $\Delta s_j = s_j - s_{j-1} = 1$ and $\Delta s = s_{j+1} - s_{j-1} = 2$ in embryo development, as well as difference in experimental times such as in cell differentiation (note: actual time difference can be considered as scaling in time). Then, we evaluate a force-like action in expression flux.

The interaction flux of a critical state: The interaction flux represents flux of a critical state $X(s_j)$ with respect to another critical state (Super, Near, Sub) or the environment (E: milieu) Y_j can be defined as:

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

where, again, the first and second terms represent IN flux and OUT flux, respectively, and the net value (i.e., IN flux- OUT flux), represents incoming (IN) interaction flux from *Y* for a positive sign and outgoing (OUT) interaction flux to *Y* for a negative sign. *Y* represents either the numerical value of a specific critical state or the environment, where a state represented by *Y* is deferent from one by *X*.

With regard to the global perturbation event, the net **kinetic energy flux** [Tsuchiya, M., et al., 2016] clearly reveals it in a critical state (**Figure 10**):

$$K\left(X(s_j)\right) = \frac{1}{2}\left\{\left(\frac{(X(s_j)-X(s_{j-1}))}{\Delta s_j}\right)^2 - \left(\frac{(X(s_{j+1})-X(s_j))}{\Delta s_{j+1}}\right)^2\right\} - \langle K(X)\rangle, \qquad (4)$$

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

$$f\left(X(s_j)\right) = \sum_{i=1}^{M=2} f\left(X(s_j); A_i\right) + f\left(X(s_j); E\right), \quad (5)$$

where state $A_i \in \{\text{Super}, \text{Near}, \text{Sub}\}$ with $A_i \neq X$, and M is the number of internal interactions ($M = 2$), i.e., for a given critical state, there are two internal interactions with other critical states. Equation (5) tells us that the sign of the difference between the net self-flux and the overall contribution from internal critical states, $f\left(X(s_j)\right) - \sum_{i=1}^{M=2} f(X(s); A_i)$, reveals incoming flux (positive) from the environment to a critical state or outgoing flux (negative) from a critical state to the environment.

Here, we need to address the previous result of expression flux dynamics in mouse single-cell genome expression [Tsuchiya, M., et al., 2017], where expression of a critical state was taken as $X(s_j) = \frac{1}{N_C}\sum_{i=1}^{N_C} ln\langle\varepsilon_i(s_j)\rangle$, which has different ordering of operations: first taking the natural log of expression and then, average operation. Hence, in flux dynamics, we examine whether or not mathematical operation between averaging and natural log, i.e., operation between $ln\,\langle\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(s_j)\rangle$ and $\frac{1}{n}\sum_{i=1}^{n} ln\langle\varepsilon_i(s_j)\rangle$ can be exchanged (mathematically commuted). In microarray data, flux behaviors do not change much between these action ordering (almost the same: commuted). Whereas in RNA-Seq data, they are not commuted due to its data structure with lots of zero values; adding small random noise into log of expression, $ln\langle\varepsilon_i(s_j)\rangle$ (previous result) makes good effect (noise-sensitive), but not in $ln\,\langle\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(s_j)\rangle$): noise-insensitive (this report). Although detail dynamics of interaction flux changes by taking different action-orderings in RNA-Seq data (e.g., Fig. 6 in [Tsuchiya, M., et al., 2017]), two important characteristics in genome-engine: the formation of dominant cyclic flux between super- and sub-critical states and the generator role of the sub-critical state do not change (invariant features). Thus, we conclude that the concept of the genome-engine is quite robust.

Contributions

MT initiated the project; MT, AG and KY designed the study; MT developed the study and analyzed data; AG and KY provided theoretical support; MT, AG and KY wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

MT sincerely thanks the following institution and individuals who helped complete this research project: the SEIKO Life Science Laboratory, Osaka, Japan, his family (particularly, his daughters, Drs. Kimiko and Kazumi Tsuchiya with any editing), and Drs. Andrzej Kasperski and Jekaterina Erenpreisa for fruitful discussions.

References

1. Bak, P., Tang, C. Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.* 59: 381–384; doi:10.1103/PhysRevLett.59.381.
2. Bolstad, B. M., Irizarry, R. A., Astrand, M., Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-193; doi:10.1093/bioinformatics/19.2.185.
3. Censi, F., Giuliani, A., Bartolini, P., Calcagnini, G. (2011). A Multiscale Graph Theoretical Approach to Gene Regulation Networks: A Case Study in Atrial Fibrillation. *IEEE Trans. Biomed. Eng*. 99, 1–5.; 2015; doi: 10.1109/TBME.2011.2150747.
4. Ciofani, M., Madar, A., Galan, C., Sellars, M., Mace, K., Pauli, F., Agarwal, A., Huang, W., Parkhurst, C. N., Muratet, M., et al. (2012). A validated regulatory network for Th17 cell specification. *Cell* 151, 289–303; doi: 10.1016/j.cell.2012.09.016.
5. 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; doi: 10.1126/science.1245316.
6. Flory, P. (1953). Principles of Polymer Chemistry, Cornell University Press.
7. Gennes, P.G. (1979). Scaling Concepts in Polymer Physics, Cornell University Press, Ithaca, NY.
8. Giuliani, A., Tsuchiya, M., Yoshikawa, K. (2018). Self-Organization of Genome Expression from Embryo to Terminal Cell Fate: Single-Cell Statistical Mechanics of Biological Regulation *Entropy* 20, 13; https://doi.org/10.3390/e20010013.
9. Halley, J. D., Burden, F. R., Winkler, D. A. (2009). Summary of stem cell decision making and critical-like exploratory networks. *Stem Cell Res.* 2, 165-177; https://doi.org/10.1016/j.scr.2009.03.001.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10. Huang, S., Eichier, G., Bar-Yam, Y., Ingber, D. E. (2005), Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701-128705;
doi: 10.1103/PhysRevLett.94.128701.
11. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249-264; DOI: 10.1093/biostatistics/4.2.249
12. Jensen, H. J. (1998). Self-Organized Criticality. *Cambridge Univ. Press*, UK.
13. Kauffman, S. A. (1993). The Origins of order self-organization and selection in evolution. *Oxford Univ. Press*, New York; https://doi.org/10.1142/9789814415743_0003.
14. Kim, K.Y., Wang, J. (2007). Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *PLoS Comp. Biol*. 3, e60;
https://doi.org/10.1371/journal.pcbi.0030060.
15. Langton, C. G. (1990). Computation at the edge of chaos - phase transitions and emergent computation. *Physica D* 42, 12-37; https://doi.org/10.1016/0167-2789(90)90064-V.
16. MacArthur, B. D., Ma'ayan, A., Lemischka, I. R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* 10, 672-681; doi: 10.1038/nrm2766.
17. Maeshima, K., Ide, S., Hibino, K., Sasai, M. (2016). Liquid-like behavior of chromatin, Curr. Opi. Gene. Dev. 37, 36-45; https://doi.org/10.1016/j.gde.2015.11.006.
doi: 10.1103/PhysRevLett.94.128701.
18. Marković, D., Gros, C. (2014). Power laws and self-organized criticality in theory and nature. *Phys. Rep.* 536, 41-74; https://doi.org/10.1016/j.physrep.2013.11.002.
19. McClintick, J. N., Edenberg, H. J. (2006). Effects of filtering by present call on analysis of microarray experiments. *BMC Bioinformatics* 7, 49; doi: 10.1186/1471-2105-7-49.
20. Mojtahedi, M., Skupin, A., Zhou, J., Ivan Castaño, I. G., Rebecca Y.Y. Leong-Quong, R. Y.Y., Chang, H., at al. (2016). Cell Fate Decision as High-Dimensional Critical State Transition. *PLoS Biol* 14: e2000640; doi:10.1371/journal.pbio.2000640.
21. Muñoz, M. A. (2018). Colloquium: Criticality and dynamical scaling in living systems. *Rev. Mod. Phys*. 90, 031001-031030; https://doi.org/10.1103/RevModPhys.90.031001.
22. Nakai, T., Hizume, K., Yoshimura, S. H., Takeyasu, K., Yoshikawa, K. (2005). Phase transition in reconstituted chromatin *Europhys. Lett*. 69, 1024-1030; https://doi.org/10.1209/epl/i2004-10444-6.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. Raser, J. M., O'Shea, E. K. (2005). Noise in gene expression: Origins, consequences, and control. *Science* 309, 2010–2013; doi: 10.1126/science.1105891.
24. 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; doi: 10.1186/1471-2164-10-545.
25. Sakaue, T., Yoshikawa, K. (2006). On the formation of rings-on-a-string conformations in a single polyelectrolyte chain: A possible scenario. *J. Chem. Phys*. 125, 32767-6(2006); doi: 10.1063/1.2244555.
26. Schiessel, H. (2003). The physics of chromatin. *J. Phys. Condes. Matter*, 15, R699-R774; doi: 10.1088/0953-8984/27/6/060301.
27. Shew, C.Y., Yoshikawa, K. (2007). Mean field theory for the intermolecular and intramolecular conformational transitions of a single flexible polyelectrolyte chain Journal of Chemical Physics, 126, 32767-9.
28. Suzuki, Y., Yoshikawa, Y., Yoshimura, S.H., Yoshikawa, K., Takeyasu, K. Unraveling DNA dynamics using atomic force microscopy. (2011). *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.*, 3, 574-588; doi: 10.1002/wnan.150.
29. Takahashi, K., Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency, *Nature Rev. Mol. Cell Biol.* 17, 183–193; doi: 10.1038/nrm.2016.8.
30. Tsuchiya, M., Hashimoto, M., Takenaka, Y., Motoike, I. N., Yoshikawa, K. (2014). Global genetic response in a cancer cell: Self-organized coherent expression dynamics. *PLOS One* 9: e97411; https://doi.org/10.1371/journal.pone.0105491.
31. 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; https://doi.org/10.1371/journal.pone.0128565.
32. Tsuchiya, M., Giuliani, A., Hashimoto, M., Erenpreisa, J., Yoshikawa, (2016). K. Self-organizing global gene expression regulated through criticality: Mechanism of the cell-fate change. *PLoS ONE* 11: e0167912; https://doi.org/10.1371/journal.pone.0167912.
33. Tsuchiya, M., Giuliani, A., Yoshikawa, K. (2017). Single-Cell Reprogramming in Mouse Embryo Development through a Critical Transition State. *Entropy* 2017, 19, 584; https://doi.org/10.3390/e19110584.
34. Tsuchiya, M., Giuliani, A., Yoshikawa, K. (2018). A Quantitative Evaluation of Symmetry Breaking In Nonlinear-Oscillatory System - Based on Flux Dynamics (Effective force) View Point. Presentation; doi: 10.13140/RG.2.2.34048.74240.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

35. Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131-1139; doi:10.1038/nsmb.2660.
36. Yoshikawa, K. (2002). Field hypothesis on the self-regulation of gene expression. *J. Biol. Phys*. 28: 701-712; doi: 10.1023/A:1021251125101.
37. Yoshikawa, K., Takahashi, M., Vasilevskaya, V.V., Alexei R. Khokhlov, A.R. (1996). Large Discrete Transition in a Single DNA Molecule Appears Continuous in the Ensemble. *Phys. Rev. Lett*., 76, 3029-3031; 10.1103/PhysRevLett.76.3029.
38. Yoshikawa, K., Yoshikawa, Y. (2002). Compaction and Condensation of DNA, in Pharmaceutical Perspective of Nucleic Acid-Base Therapy, Tayler & Francis, Abingdon, UK, 137-163.
39. Zinchenko, A., Pyshkina, O., Lezov, A., Sergeyev, V., Yoshikawa, K. (2008). Single DNA molecules: compaction and decompaction (chapter 3), in DNA interactions with polymers and surfactants. Wiley-Blackwell.

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables:

Table 1: Critical States of Single-Cell Genome Expression

Mouse Embryo Development	Human Embryo Development	Th17 cell differentiation
Critical States:	**Critical States:**	**Critical States:**
Super-critical: 3583 RNAs **-5.6*** < *ln<nrmsf>*	**Super-critical**: 2607 RNAs **-5.9*** < *ln<nrmsf>*	**Super-critical**: 1253 RNAs **-4.4*** < *ln<nrmsf>*
Near-critical: 5922 RNAs -7.5 < *ln<nrmsf>* < **-5.6**	**Near-critical**: 6707 RNAs -8.0 < *ln<nrmsf>* < **-5.9**	**Near-critical**: 1452 RNAs -6.0 < *ln<nrmsf>* < **-4.4**
Sub-critical: 7636 RNAs *ln<nrmsf>* < -7.5	**Sub-critical**: 7345 RNAs *ln<nrmsf>* < -8.0	**Sub-critical**: 10429 RNAs *ln<nrmsf>* < -6.3

***Critical Point at the boundary between Super- and Near-critical states**

Table 2: Critical States of Cell Population Genome Expression

HRG-MCF-7 cells	EGF-MCF-7 cells	atRA-HL-60 cells	DMSO-HL-60 cells
Critical States:	**Critical States:**	**Critical States:**	**Critical States:**
Super-critical: 4756 mRNAs -2.0 < *ln<nrmsf>*	**Super-critical**: 4210 mRNAs -2.1 < *ln<nrmsf>*	**Super-critical**: 2187 mRNAs -2.1 < *ln<nrmsf>*	**Super-critical**: 1721 mRNAs -2.0 < *ln<nrmsf>*
Near-critical: 7649 mRNAs **-2.5*** < *ln<nrmsf>* < -2.0	**Near-critical**: 8667 mRNAs **-2.6*** < *ln<nrmsf>* < -2.1	**Near-critical**: 2935 mRNAs **-2.6*** < *ln<nrmsf>* <-2.1	**Near-critical**: 2663 mRNAs **-2.5*** < *ln<nrmsf>* <-2.0
Sub-critical: 9872 mRNAs *ln<nrmsf>* < **-2.5**	**Sub-critical**: 9400 mRNAs *ln<nrmsf>* < **-2.6**	**Sub-critical**: 7503 mRNAs *ln<nrmsf>* < **-2.6**	**Sub-critical**: 8241 mRNAs *ln<nrmsf>* < **-2.5**

***Critical Point at the boundary between between Near- and Sub-critical states**

Table 3: Single Cell: Activation of the CP associated with Coherent Perturbation on the Genome Engine.

	Location of the CP	**Activation** of the CP	**Global Perturbation** on Genome-Engine
Mouse Embryo Development	$ln\langle nrmsf\rangle_{CP} \sim -5.6$	**ON** after **Late 2-cell**	**Suppression-Enhancement**
Human Embryo Development	$ln\langle nrmsf\rangle_{CP} \sim -5.9$	**ON** after **8-cell**	**Suppression-Enhancement**
Th17 Immune Cell Differentiation	$ln\langle nrmsf\rangle_{CP} \sim -4.4$	**ON after 6h**	**Suppression-Enhancement**

bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

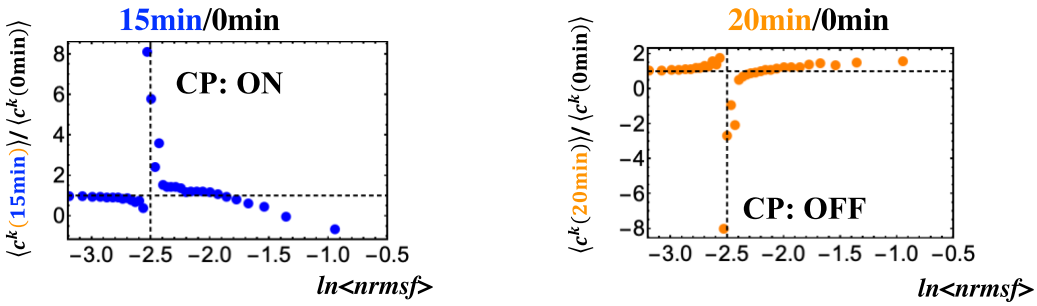
Table 4: Cell Population: Activation of the CP associated with Coherent Perturbation on the Genome Engine. *Refer to Fig. 12B in [Tsuchiya, M., et al., 2016].

	Location of the CP	**Activation** of the CP	**Global Perturbation** on Genome-Engine	**Cell Fate Change**
EFG-stimulated **MCF-7 cells**	$ln\langle nrmsf\rangle_{CP} \sim -2.6$	**OFF** state	**Local Perturbation**	**NO**
HRG-stimulated **MCF-7 cells**	$ln\langle nrmsf\rangle_{CP} \sim -2.5$	**ON** at **10-15min**	**Suppression-Enhancement***	**2-3h**
DMSO-stimulated **HL-60 cells**	$ln\langle nrmsf\rangle_{CP} \sim -2.5$	**ON** at **12-18h**	**Enhancement-Suppression**	**18-24h**
atRA-stimulated **HL-60 cells**	$ln\langle nrmsf\rangle_{CP} \sim -2.6$	**ON** at **24-48h**	**Enhancement-Suppression**	**24-48h**

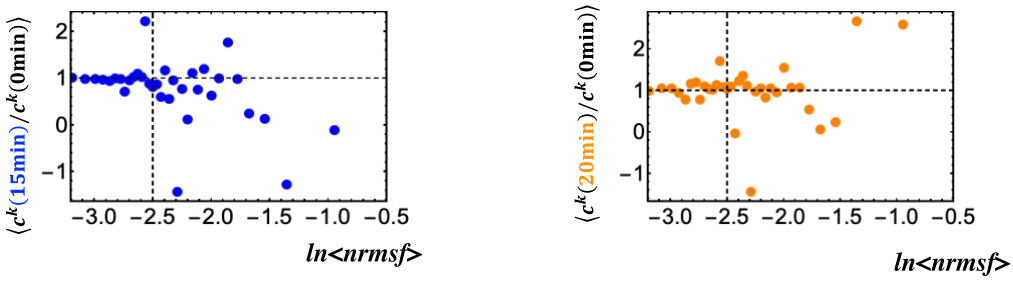
bioRxiv preprint doi: https://doi.org/10.1101/637033; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure S1:

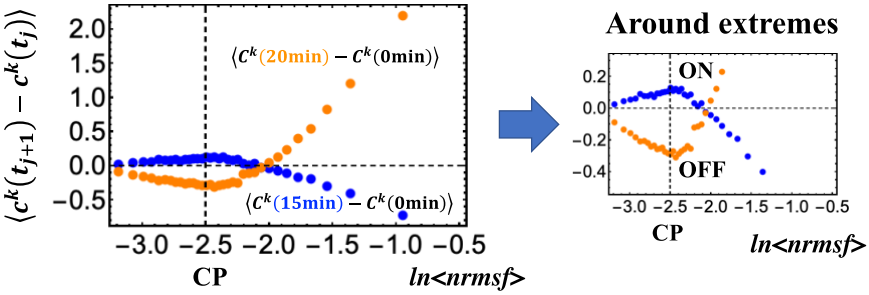
A) Fold Change in Ensemble Averages



B) Ensemble average of fold change in individual expression



C) Ensemble average of time difference in expression group



Supplementary Figure S1 (HRG-stimulated MCF-7 cells):

A) Fold change in ensemble (group) averages between expression groups shows that coherent-stochastic behavior (emergent coherent behavior from stochastic expression) represented by the center of mass (average) of group reveals a clear coil-globule transition.

B) Ensemble average of fold change in individual expressions between two temporal groups, $<\boldsymbol{c}^k(t_{j+1})/\boldsymbol{c}^k(t_j)>$, does not reveal any transitional behavior that is attributable to the stochastic behavior of expression (sensitive in fold change).

C) Ensemble average of time difference in expression groups, $<\boldsymbol{c}^k(t_{j+1})- \boldsymbol{c}^k(t_j)>$, supports the coherent scenario in (**A**), where at around the CP (*ln<nrmsf>*~ -2.5) there is a positive maximum at 0-15min and negative minimum at 0-20min that corresponds to either an ON or OFF state, respectively.