Collective fluctuations are crucial for global biological response

Masa Tsuchiya[†,‡,*], Vincent Piras[†,‡], Masaru Tomita[†,‡] and Kumar Selvarajoo[†,‡]

[†] Institute for Advanced Biosciences, Keio University,
Tsuruoka, 997-0035, Japan.
[‡]Systems Biology Program, School of Media and Governance, Keio University,
Fujisawa, 252-8520, Japan.
*Email: tsuchiya@ttck.keio.ac.jp

Abstract– One fundamental mystery in biology is to grasp how a cell chooses a specific differentiation path in spite of having enormous number of molecules leading to complex multi-molecular interactions through DNA, RNA, proteins and metabolites. The basis for such deterministic cellular regulations may stem from the creation of attractor states in biological systems. To understand how attractor states are formed in complex biological responses, we investigated genome-wide expressions (mRNA expressions) dynamics for two distinct processes: i) the innate immune response of macrophages to lipopolysaccharide (LPS) stimulation and ii) the neutrophil differentiation process (HL-60 cell differentiation into neutrophil) using statistical correlation metrics defining the correlation space. For both processes, forming groups of genes reduced response fluctuations, revealing the hidden collective genome-wide (global) expression dynamics and their biological roles. We found roles of low expressed genes, which have been considered insignificant and noisy, as responsible for collective global dynamics. Moreover, in neutrophil differentiation, correlation distributions from the initial time point of two different stimuli (all-trans-retinoic acid (atRA) and dimethyl sulfoxide (DMSO) on HL-60 cells) overlap with their maximum probabilities in the correlation space to reveal the existence of a common neutrophil attractor. Defining attractor boundary as inflection curves of distributions, we found that there are specific gene ensembles (collectively named "genome vehicle") responsible for the neutrophil attractor and the collective motion of lowly and moderately variable genes within the genome vehicle plays an important role in the formation of the attractor. These findings may provide completely new comprehensive mechanistic view of cell fate decision. Our results suggest that the collective motion of the small fluctuating genes may drive the complex cell differentiation process.

1. Introduction

A large number of molecules constituting of DNA, RNA, proteins and metabolites in a cell interact with each other to execute numerous inter- and intracellular processes in response to environmental changes or physiochemical perturbations. Forming a large degree of interconnectivity, cells control and self-regulate vital cellular processes such as the immune responses, cell cycle, cell division and differentiation. It is truly intriguing to grasp how a specific path during cellular process such as differentiation can be chosen from vast number of possibilities that can arise through the complex multi-molecular interactions. The basis for such determinism may stem from the formation of attractor states in cellular processes [1,2]; these attractor states could drive orchestrated diverse regulations of biological processes through the averaging effect of response fluctuations at cell population level as suggested by our recent investigations on microarray datasets [3-5]. The averaging effect revealed emergent global regulation. The phenomena of averaging effect in physical many-body systems such as thermodynamics, condensed matter and fluid dynamics have been well studied and understood as mean field theory. However, it remains unclear how the complex and dynamically evolving molecular networks found in biological systems can give rise to a globally coherent orchestrated response such as cell fate decision of differentiation.

To grasp underlining mechanisms of how genome-wide expressions can be guided globally in complex cellular systems, we investigated two distinct processes, i) the innate immune response of macrophages to LPS [3] and ii) HL-60 cell differentiation into neutrophil cells [4] from a gene expression correlation point of view. In studies of large-scale high-throughput gene expression such as microarray, estimation of signal intensity is difficult because of unspecific binding affinity between probes and targets mRNAs, and especially in the low level expression changes, the effect of noises, compared with specific binding activity, is likely larger than that for highly variable genes. To overcome the difficulties of dealing with single gene expression noises, we grouped genes of whole genome into ensembles and analyzed their correlation dynamics based on temporal gene expressions using statistical correlation metrics (Pearson and mutual information). Pearson correlation is mathematically the cosine of the angle between two N dimensional gene expression vectors

when a gene expression value is evaluated from the average of N dimensional gene expressions (N = 22690 for the LPS response and N = 12625 for the neutrophil differentiation), whereas mutual information measures the mutual dependence of the two genome-wide gene expressions defined by joint and conditional entropies [4].

2. Global LPS simulated innate immune responses

To investigate gene expression dynamics in LPS simulated innate immune response on macrophages, we evaluated the correlation structure of Affymetrix mouse expression data for wildtype, MyD88 knockout (KO), TRIF KO, Double KO (DKO) at 0, 1 and 4 hours. Upon LPS binding, TLR4 triggers two major intracellular pathways, the MyD88-dependent and TRIF-dependent pathways to generate innate immune response. Thus, in case of the double knockout of MyD88 and TRIF molecules, it was believed that no immune response would occur [6]. However, we found that temporal Pearson correlation on DKO still has progressive response, suggesting DKO still possesses LPS response contrary to a current belief.

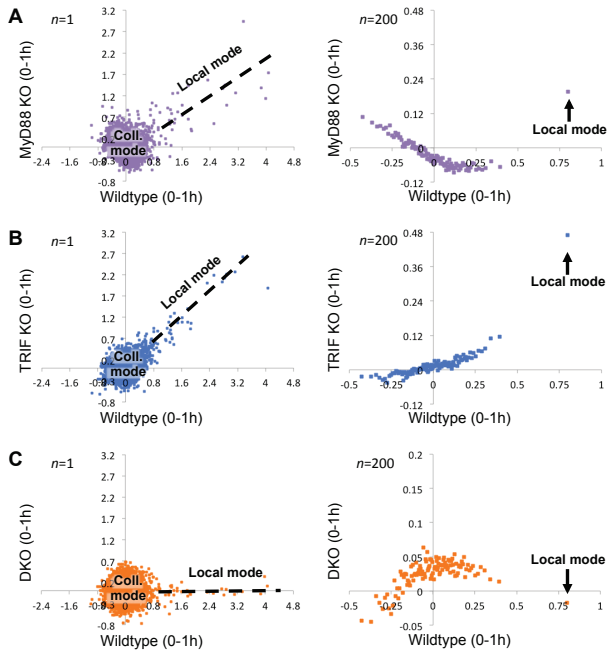

Figure 1. Emergence of asymptotic whole genome collective behaviors in LPS stimulated immune response. Large scatter in collective mode and linear distribution in local mode. Genome-wide single ORFs (left panels) expression changes for 0–1 h between genotypes: wildtype vs. A) MyD88KO, B) TRIF KO, C) DKO. Right panels: corresponding plots for group of 200 ORFs, sorted by their expression change in the corresponding genotype (x-axis). Arrows indicate groups containing local mode (immune genes).

Additionally, we confirmed that the DKO correlation response on 157 immune genes has a flat profile, almost perfect correlation between different times, i.e., the lack of any classical immune response to LPS [3]. Thus, as a result, the global response on DKO is different from the local innate immune response (see also Figure 1C).

3. Emergent genome-wide control in LPS stimulated Macrophages

To understand the temporal progress of biphasic (local and global) responses, we investigated the changes of whole genome expression for 0-1h in each genotype. We plotted the expression change (Δx) of single ORFs (wildtype for x-axis and other genome types for y-axis) (Figure 1) and the average value of open reading frames (ORFs) sorted from highest to lowest expression changes for each group. Remarkably, we observed the transition from a large scatter in expression distributions around the origin for single ORFs to smooth linear curves for group of 80 (ORFs) onwards for all genotypes; grouping of expression distribution for 80 ORFs or above forms Gaussian distribution and the average value of each group follows linear lines. These linear curves reveal the emergence of regulatory signature working at the level of groups of genes. Furthermore, the fluctuations from these asymptotic lines reduce as the grouping size is increased. Notably, for DKO, the result reinforced that DKO possesses genome-wide LPS response through MyD88- and TRIF- independent manner. These emergent linear asymptotic curves can be linearly superposed to decipher the global gene regulatory differential control principle of the transcriptional and mRNA decay machineries between the wildtype and mutant genomes [3]. These works also have shed new light on innate immune response, providing a significant role for the lowly expressed genes in the diverse collective mode, which are often considered noisy and insignificant in microarray experiments.

4. Emergence of asymptotic whole genome collective behaviors in HL-60 cell differentiation into neutrophil

To investigate the collective behavior of gene expressions in HL-60 cell differentiation [4], we grouped genes according to their expression variance (gene variation) across time and split the whole genome into p groups, where p is the integer values of N/n for n = 10, 50, 100, 200, 500, 1000 (N = 12625). We plotted the set of mean values of gene expressions for p groups, which show that the groups' mean values transited from scatter to the emergent (temporally oscillating) asymptotic curves at each time point (Figure 2). Again, the fluctuations from these asymptotic lines reduce following the inverse square root law as the grouping size is increased. This suggests that the

temporal genome-wide averaging behavior of collective expression dynamics exists. We set n = 200 genes since it produced acceptably good resolution.

5. The existence of genome vehicles as collective dynamics of specific gene ensembles crucial for neutrophil differentiation

To elucidate further understanding of the significant role of the global response [5,7], we randomly grouped genes from whole genome (N = 12625) into different ensemble sizes (n = 10, 50, 100, 200, 500, 1000) with 100 repeats at each time point (t = 0, 2, 4, 8, 12, 18, 24, 48, 72, 96, 120, 144, 168h) to evaluate the dynamics of correlation distributions of the gene ensembles in the correlation space, (r_v, I), that is defined by Pearson correlation, r_v and mutual information, I of gene variation from the initial time point. Plotting average correlation values at each time confirms the formation of Gaussian distribution (due to the central limit theorem). The ensemble size of n = 200, with good resolution, was chosen to evaluate the probability distributions of r_v and I for each time point of the gene expression data.

Utilizing noise reduction by grouping genes, we plotted the probability distributions of r_v and I versus time, and observed that as the ensemble size is increased, the distributions localized to specific points (r_v, I), especially after 48h (Figure 3A). The superposition of the probability distributions (SPD) of r_v and I of atRA and DMSO responses overlap indicating the presence of cell fate attractor, as it corresponds to the fact the two stimuli elicit the same biological end-point, the generation of a mature neutrophil cell. Regarding the attractor region, we adopted the concept of critical (inflection) points as used in phase transitions in thermodynamic systems to determine the boundary of the neutrophil attractor; unlike continuous dynamics, we are unable to determine the attractor basin for neutrophil differentiation due to the limited temporal data points. Common inflection curves between atRA and DMSO responses, evaluated from the gradients of the SPD for r_v and I, determine the neutrophil attractor (Figure 3A). Tracking the ensembles' trajectories, we noticed that only certain, not all, fall into the attractor in a fractal-like manner [4]. Particularly, the removal of these genome elements from the whole genomes, for both atRA and DMSO responses, no longer have the common region, i.e., destroys the attractor (Figure 3B). This provides evidence for the existence of specific genome elements (named ''genome vehicle'') responsible for the neutrophil attractor. Selective portions of fractal-like gene ensembles are responsible for the neutrophil cell fate decision acting as 'drivers' of reaching a genome-wide characteristic profile.

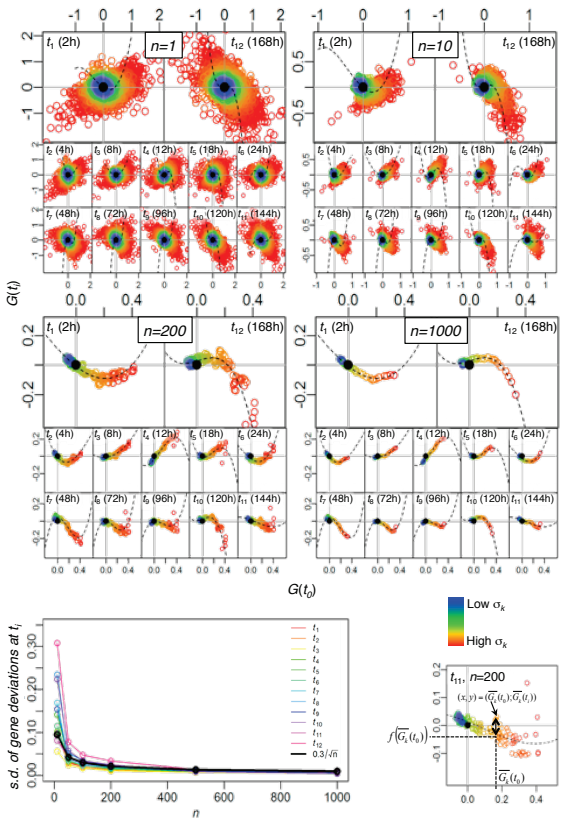

Figure 2. Emergence of asymptotic whole genome collective behaviors in neutrophil differentiation process of HL-60 cells. 12625 genes (N) are grouped according to their variance across time, where G stands for atRA whole genome and the ensemble size is n. The whole genome was sorted from the highest to the lowest standard deviation. As n is increased, the standard deviation of the ranked groups at t_i (i = 0, 2, 4, 8, 12, 18, 24, 48, 72, 96, 120, 144, 168h) decreases and mean values of ensembles approach to the asymptotic curve (dashed lines), obeying the inverse square root law (center panel, thick black line) with a transition occurring around $\sqrt{N}$. Note for DMSO response, the similar transition also occurs (not shown).

6. Conclusions and Discussions

In this paper, we show the existence of averaging effect of large numbers in two distinct immune processes. In LPS stimulated innate immune response, our works reveal the presence of a highly ordered, coordinated, genome-wide expression dynamics, thereby requesting the need to consider global phenomena when interpreting immune response, and not restricting to a few highly expressed immune genes. On the other hand, in the neutrophil differentiation, we show that the self-regulation of genome vehicles leads to the formation of a common attractor for neutrophil differentiation. In addition, we demonstrate that the collective motion of lowly and moderately variable genes within the genome vehicle (GV), play an important role in the formation of

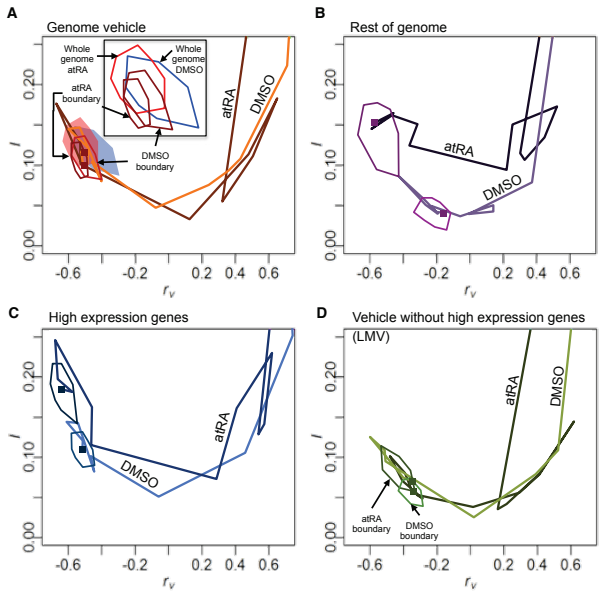

Figure 3. **The loss of the attractor when genome vehicles are removed.** The SPD (superposition of the probability distributions) boundaries (defined by their joining inflection points, see details in [5]) and trajectories for atRA (plain lines with darker tone) and DMSO (lighter tones) responses of (A) genome elements falling into attractor (i.e., genome vehicles, GV) overlap and converge, indicating the formation of neutrophil attractor. (Overlapping SPD boundaries of atRA and DMSO responses of the whole genomes are indicated by red and blue polygons respectively, insert is a zoom of the attractor region, red and blue lines indicate atRA and DMSO SPDs respectively, brow lines indicate GV SPDs), (B) rest of genome elements without genome vehicles do not overlap and converge, (C) high expression genes (2-fold change from t_0 for at least one time point) of the genome vehicles do not overlap and converge, (D) lowly and moderately variable (LMV) genes of the genome vehicles still overlap and converge, retaining the neutrophil attractor. Last time point is represented by a square.

the neutrophil attractor, perhaps indicating the non-instructive signaling of genes related to small-amplitude DNA motions in chromosome dynamics regulating gene expressions. Ample recent evidence [8] reveals that a chromatin epigenetic state [9] may determine a cellular state, that is, a genome-wide epigenetic modification pattern of histone and DNA with other proteins packing the genome inducing dynamic change of DNA-histone interaction and structure. Furthermore, for cell differentiations, there exist distinct bivalent domain patterns, where bivalent gene expressions are silenced while remaining poised for activation. Determining the relationship between GV genes responsible for the emergent attractor states and epigenetic modifications especially on bivalent genes may provide a dynamic interacting picture of how bivalent domains forming expression active and repressive proteins such as trithorax group (TrxG) and polycomb group (PcG) proteins, respectively, determine cell commitments. An important future direction is to understand global self-regulation under an attractor state, which may reveal the epigenetic language with rules/grammars and indicate that the whole epigenetic system is following a simple nonlinear governing rule.

Acknowledgments

We would like to thank NOLTA organizing committee members for their fruitful suggestions and comments. This study is funded by the CREST of Japan Science and Technology, Tsuruoka City, Yamagata prefecture government, GCOE program of Keio University and JASSO.

References

[1] S.A. Kauffman, The Origins of Order, New York, Oxford University Press, 1993.

[2] S. Huang, G. Eichler, Y. Bar-Yam, D.E. Ingber, “Cell fates as high-dimensional attractor states of a complex gene regulatory network”, Phys. Rev. Lett., Vol.94, 128701, 2005.

[3] M. Tsuchiya, V. Piras, S. Choi, S. Akira, M. Tomita, A. Giuliani and K. Selvarajoo, “Emergent genome-wide control in wildtype and genetically mutated lipopolysaccarides stimulated macrophages”, PLoS ONE , vol.4, e4905, 2009.

[4] M. Tsuchiya, V. Piras, A. Giuliani, M. Tomita and K. Selvarajoo, “Collective dynamics of specific gene ensembles crucial for neutrophil differentiation: the existence of genome vehicles revealed”, PLoS ONE, vol.5, e12116, 2010.

[5] M. Tsuchiya, K. Selvarajoo, V. Piras, M. Tomita and A. Giuliani, “Local and Global responses in complex gene regulation networks”, Physica A, vol.388, pp.1738-1746, 2009.

[6] T. Hirotani, M. Yamamoto, Y. Kumagai, et al., “Regulation of lipopolysaccharide-inducible genes by MyD88 and Toll/IL-1 domain containing adaptor inducing IFN-beta”, Biochem Biophys Res Commun vol.328 pp. 383–392, 2005.

[7] M. Tsuchyia, S. T. Wong, Z. Y. Yeo, et al., “Gene expression waves: cell cycle independent collective dynamics in cultured cells”, FEBS J. vol.274, pp. 2878-2886, 2007

[8] S. Yamanaka and H. M. Blau, “Nuclear reprogramming to a pluripotent state by three approaches”, Nature vol.465, pp. 704-710, 2010.

[9] T. S. Mikkelsen, M. Ku, D. B. Jaffe, et al., “Genome-wide maps of chromatin state in pluripotent and lineage-committed cells”, Nature, vol.448, pp. 553-560, 2007.