Collective Dynamics of Specific Gene Ensembles Crucial for Neutrophil Differentiation: The Existence of Genome Vehicles Revealed

Masa Tsuchiya1,2*, Vincent Piras1,2*, Alessandro Giuliani3, Masaru Tomita1,2, Kumar Selvarajoo1,2*,

1 Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan, 2 School of Media and Governance, Keio University, Fujisawa, Kanagawa, Japan, 3 Environment and Health Department, Istituto Superiore di Sanità, Rome, Italy

Abstract

Cell fate decision remarkably generates specific cell differentiation path among the multiple possibilities that can arise through the complex interplay of high-dimensional genome activities. The coordinated action of thousands of genes to switch cell fate decision has indicated the existence of stable attractors guiding the process. However, origins of the intracellular mechanisms that create “cellular attractor” still remain unknown. Here, we examined the collective behavior of genome-wide expressions for neutrophil differentiation through two different stimuli, dimethyl sulfoxide (DMSO) and all-trans-retinoic acid (atRA). To overcome the difficulties of dealing with single gene expression noises, we grouped genes into ensembles and analyzed their expression dynamics in correlation space defined by Pearson correlation and mutual information. The standard deviation of correlation distributions of gene ensembles reduces when the ensemble size is increased following the inverse square root law, for both ensembles chosen randomly from whole genome and ranked according to expression variances across time. Choosing the ensemble size of 200 genes, we show the two probability distributions of correlations of randomly selected genes for atRA and DMSO responses overlapped after 48 hours, defining the neutrophil attractor. Next, tracking the ranked ensembles’ trajectories, we noticed that only certain, not all, fall into the attractor in a fractal-like manner. The removal of these genome elements from the whole genomes, for both atRA and DMSO responses, destroys the attractor providing evidence for the existence of specific genome elements (named “genome vehicle”) responsible for the neutrophil attractor. Notably, within the genome vehicles, genes with low or moderate expression changes, which are often considered noisy and insignificant, are essential components for the creation of the neutrophil attractor. Further investigations along with our findings might provide a comprehensive mechanistic view of cell fate decision.

Citation: Tsuchiya M, Piras V, Giuliani A, Tomita M, Selvarajoo K (2010) Collective Dynamics of Specific Gene Ensembles Crucial for Neutrophil Differentiation: The Existence of Genome Vehicles Revealed. PLoS ONE 5(8): e12116. doi:10.1371/journal.pone.0012116

Editor: Sangdun Choi, Ajou University, Republic of Korea

Received June 13, 2010; Accepted July 16, 2010; Published August 11, 2010

Copyright: © 2010 Tsuchiya et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study is funded by the CREST (Core Research of Evolutional Science & Technology) of Japan Science and Technology, Tsuruoka City, Yamagata prefecture government, GCOE (Global Center of Excellence) program of Keio University and JASSO (Japan Student Services Organization). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tsuchiya@ttck.keio.ac.jp (MT); kumar@ttck.keio.ac.jp (KS)

† These authors contributed equally to this work.

Introduction

Cell fate decision involves reprogramming of precursor cells into the differentiated state. It is intriguing to grasp how a specific path is chosen by a cell, among the several possibilities that can arise, through the complex multi-molecular interactions during differentiation. The understanding of such deterministic process, where the macroscopic stable cell fate transition requires the coordinated regulation of thousands of genes forming networks, could uncover mechanisms that control cell differentiation, as well as reveal better strategy to suppress disease progression, e.g., cancer proliferations.

The study of large-scale network dynamics has been investigated in a variety of fields including mathematics, physics, information sciences, ecology and biology ever since the onset of the nineties [1–2]. A large number of studies have already shown that the emergence of collective behavior, such as synchronization of processes, can arise due to the non-linear regulations of complex network systems with environmental perturbations. For example in biology, the secretion and detection of autoinducer molecules between bacteria enable a population of them to collectively regulate gene expression and, therefore, produce coordinated group behavior such as the formation of biofilm by Pseudomonas aeruginosa [3–4]. 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.

High-throughput omics (transcriptomics, proteomics & metabolomics) analyses have indicated that the molecular interactions within a living cell typically form a single, largely interconnected network [5–8]. It is, thus, necessary to have an integrated network view to understand cellular processes such as cell fate transitions or differentiations in which cells receive a broad range of biological signals or perturbations which influence gene expressions across the entire genome to produce reliable and robust outcome.

To demonstrate the genome-wide integrated response for cell fate decision, Huang et al. investigated the differentiation of human pro-myelocytic leukemia HL-60 cells into neutrophil by the
action of two different reagents, DMSO and atRA [9]. Based on the 2773 highest expressed genes (based on two-fold expression changes), Huang et al. showed the convergence of cell fate despite different initial transcriptome dynamics arising from the different stimuli, thus suggesting the presence of stable multidimensional attractor states in biology [10–13]. Although this result is the first step towards understanding the existence of cell fate attractors, many other fundamental questions remain to be investigated. For example, what are the intracellular origins and mechanisms that instill genome-wide response? What form cellular attractors? How these emerge through the complex molecular networks? If the entire genome is linked through networks, is attractor state achieved by self-regulation [12]?

The majority of large-scale gene expression studies have focused on genes with high expression changes or variations to decipher key regulatory processes, since low-level expression changes of genes have been considered as noisy due to the issue of poor signal-to-noise ratio in microarray experiments. This is due to the difficulty in the estimation of unspecific binding abundance between probe and target in signal intensity [14–16], and especially for the low level expression changes, the effect of background noises, compared with specific binding activity, is likely larger than that for highly variable genes. However, in our recent study, we demonstrated that the splitting of whole genome into different ensembles to analyze their temporal expression changes from the initial time resulted in the reduction of their fluctuations as the ensemble size is increased. This resulted in collective genome-wide expression behaviors which exhibited local and global effects of lipopolysaccharide (LPS) stimulated macrophages; local being the well-known pro-inflammatory response of a small number of highly expressed genes, while global being the novel collective activation of diverse processes comprising the rest of the lowly expressed genes [17–18].

In this paper, we investigated the entire microarray data of HL-60 cells for atRA and DMSO stimuli including lowly variable signals over time [9]; DMSO is known to activate key transcription factors such as NF-kB [19], whereas atRA penetrates the nucleus and directly remodels chromatin structure [20]. To uncover the orchestrated gene expressions guiding cell fate decision, we used Pearson (linear) correlation and mutual information (nonlinear correlation) metrics to investigate the collective dynamics of gene expressions for each stimulus. To overcome the difficulties of dealing with single gene expression noises in microarray data, we formed grouping of genes (chosen randomly from the whole genome and ranked according to group expression changes across time) which showed the reduction of correlation noises as ensemble size is increased. From this, in contrast to a previous finding which suggested the whole genome's role in differentiation, we demonstrate that only selective portions of fractal-like gene ensembles are responsible for the neutrophil attractor. Notably, the removal of these specific gene ensembles from the whole genome, for both atRA and DMSO stimuli, destroys the attractor. Thus, for the first time, we reveal the existence of 'genome vehicle' and show that genes with low or moderate expression changes, contained within genome vehicles, are crucial for the neutrophil attractor.

Results and Discussion

Reduction of correlation noises when grouping genes

Previously, we have shown that the collective proinflammatory response of whole genome can be captured by random gene sampling of ensemble size above 80 [17–18]. Thus, to investigate the collective behavior of HL-60 cell differentiation, we randomly grouped genes from whole genome into different ensemble sizes $(n = 10, 50, 100, 200, 500, 1000)$ and evaluated their expression dynamics in the correlation space (see Methods, “Correlation analysis of gene expressions”). Both the temporal (modified) Pearson correlation of gene variation, $r_{n}$, and the corresponding temporal mutual information, $I$, distributions of the gene ensembles transited from scattered and incoherent ones to clear bell-shaped ones for $n \geq 100$, where $n \approx \sqrt{N}$ (N=12825, Figure 1). These result show that standard deviations of $r_{n}$ and $I$ distributions at each time point are reduced according to the $\sqrt{n}$ law with increasing $n$, where $n$ is the fitting coefficient. Thus, the ensemble size of $n=200$, with good resolution, was chosen to evaluate the probability distributions of $r_{n}$ and $I$ for each time point of the gene expression data, $\{V(t_{0}), V(t_{1}), ..., V(t_{12})\}$, where $V(t)$ is the whole genome expression deviation vector at $t$, $(i=0,1,..,12)$ (see Methods).

Localizing and overlapping of probability distributions of correlations for atRA and DMSO responses indicate neutrophil attractor

Utilizing the noise reduction by grouping genes, we plotted the probability distributions of $r_{n}$ and $I$ versus time, and observed that as the ensemble size is increased, the distributions localized to specific points ($r_{n}$, $I$), especially after 48h (Figure 2A–B). These localizations may be derived by the presence of neutrophil attractor. To test whether the localization of probability distributions indicate attractor state, we analyzed the superposition of $r_{n}$ and $I$ distributions after 48h for both atRA and DMSO and found they possess distinct peaks for both the atRA and DMSO responses (Figure 2C). Moreover, the superposition of the probability distributions (SPD) of $r_{n}$ 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.

To define 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. Note that due to the limited temporal data points, we are unable to determine the attractor basin for neutrophil differentiation as defined in continuous dynamics. Thus, we evaluated the gradients of the SPD for $r_{n}$ and $I$ to determine the inflection points for atRA and DMSO responses and the resultant plots reveal distinctive crater-like feature with the rings indicating inflection points (Figure 2D) and the common overlapping area of the inflection points of the SPDs, i.e., the SPD boundaries for the atRA and DMSO responses was defined as the neutrophil attractor (Figure 2E).

As a further test, the attractor boundary also encompasses the convergence of the atRA- and DMSO-trajectories (Figure 2E, right panel). To check the statistical significance of the localized SPD of $r_{n}$ and $I$ within the attractor, we verified that its standard deviation of both $r_{n}$ and $I$ distributions for each stimulus also scales with $\sqrt{n}$ as $n$ is increased (Figure 2F). Note that for the other localized SPD of $r_{n}$ and $I$ before 48h, it coincided with the whole genome trajectory loops indicating intermediary cell differentiation states [21] (Figure 2E, left panel).

Emergence of asymptotic whole genome collective behaviors

To investigate the whole genome collective behavior, we grouped genes according to their variance across time. The whole genome deviation vector, $V(t_{0})$, was sorted from the highest to the lowest standard deviation $\sigma$ (see Methods, “Ranking gene ensembles”). This sorting order at $t_{0}$ was retained for all other
time points \( t = 1, \ldots, 12 \). Next, we split the ranked whole genome into \( p \) groups, where \( p \) is the integer values of \( N/n \) for \( n = 10, 50, 100, 200, 500, 1000 \). Similarly for the random selection of genes, we checked whether expression noises can be reduced for the ranked groups as the group size is increased. We plotted the set of mean values of gene deviations for \( p \) groups, \( G_k(t) \) versus \( G_{k+1}(t) \), \( k = 1, 2, \ldots, p \) (see Methods, “Ranking gene ensembles” and Figure 3). The plots show the group’s mean values transited from scatter to the emergent asymptotic curves, \( f_i(G_k(t_0)) \), as \( n \) is increased for all \( t_0 \), \( \sigma = 1 \) for \( r_v \), and \( \sigma = 0.3 \) for \( I \). Note that this transition also occurred for all time points (data not shown).

doi:10.1371/journal.pone.0012116.g001

These asymptotic curves suggest that the genome-wide averaging behavior of collective expression dynamics exists. Once again \( n_i = 200 \) genes produced acceptably good resolution. Thus, \( n_i \) is the basic size of the genome element, \( G_k(t_0) \), and for the whole genome, we obtained 63 genome elements, totaling 12600 genes. The remaining 25 genes with very low \( s \) were discarded. Note: we used \( \sigma \) instead of coefficient of variation \( \text{CV} = \sigma/\mu \) for ranking genome elements as we are dealing with trajectories of ensemble of genes, rather than normalized form as often used in conventional approaches. Nevertheless, we compared \( \text{CV} \) versus \( \sigma \) and found linear relationships between them (Figure S1), ruling out any possible trivial scale effect as explanation of our results.

Figure 1. Transition from scattered to smooth bell shaped distributions of \( r_v \) and \( I \) when grouping genes. Distributions of (A) \( r_v \) and (B) \( I \) for ensembles of \( n \) randomly chosen genes from whole genomes \( n = 10, 50, 100, 200, 500, 1000 \), estimated by Gaussian kernel with 100 repeats at a representative \( t = 48 \)h (similar profiles are obtained for all time points), left panels for atRA and middle panels for DMSO response. Standard deviation of \( r_v \) and \( I \) distributions (right panels of (A) and (B)) at \( t = 48 \)h decreases as \( n \) is increased, following a \( 1/\sqrt{n} \) law, \( \sigma = 1 \) for \( r_v \), and \( \sigma = 0.3 \) for \( I \). Note that this transition also occurred for all time points (data not shown).

doi:10.1371/journal.pone.0012116.g001

Genome Vehicles Revealed

PLoS ONE | www.plosone.org 3 August 2010 | Volume 5 | Issue 8 | e12116
Specific genome elements fall into the attractor in a fractal-like manner

We investigated trajectories of the 63 ranked genome elements (Figure 4A, left panel) by creating a sequence of binary numbers where 1 and 0 indicate genome elements falling into and not falling into the attractor, respectively, against their standard deviation, $\sigma$ (Figure 4B, upper panels). The result showed that the genome elements falling into the attractor are non-continuous in $\sigma$. To understand the discontinuity, we checked the sensitivity of genome elements falling into the attractor, i.e., changing from 0 to 1 or vice-versa for single-gene shift (Figure 4A, right panel). We found that even a single replacement of the highest $\sigma$ gene from a genome element with the highest $\sigma$ gene of the next lower ranked genome element results in its destiny change of falling or not falling into the attractor (see e.g., $G_1(t_0)$ for DMSO, Figure 4B, lower panels). This expansion of a genome element shows fractal-like binary distributions and the sensitivity of single-gene shift within a genome element demonstrates the non-linear nature of gene expression dynamics.
Note that the expansion processes are limited by the lack of continuous data to show true fractal characteristics [22–23].

Next, we evaluated trajectories of the 63 genome elements and compared each with the whole genome trajectory, in terms of the Euclidean distance, $d$. We observed that 12 genome elements for atRA and 20 for DMSO, fall into the attractor, and among them more than 50% (with $0.24 \leq \sigma \leq 0.40$ for atRA and $0.20 \leq \sigma \leq 0.59$ for DMSO) are close to the whole genome trajectory with minimum distance ($d = 0.11$) (Figure 4C). This indicates that the genome elements falling into attractor scale with the whole genome trajectory. Overall, these results suggest that whole genome responses possess fractal-like nature for neutrophil differentiation.

To exhaustively search for more possible genome elements that can enter the attractor, we performed an iterative procedure where we removed the initial elements into attractor and shifted the remaining genomes by 50 genes from the highest $\sigma$ values to create new genome elements (Figure S2). Through this, we obtained additional 9 elements for atRA and 8 elements for DMSO.

The loss of the attractor when specific genome elements are removed

We evaluated the SPDs of $\tau$ and $I$ of all the genome elements into the attractor and found the SPD boundaries of atRA and DMSO responses overlapped to maintained the attractor (Figure 5A), while those for the rest of genome elements did not (Figure 5B). Moreover, for the genome elements falling into attractor, the corresponding trajectories of both atRA and DMSO converged, but not for the trajectories of the rest of genome elements (Figure 5A–B). These results indicate that the rest of genomes for both atRA and DMSO stimuli failed to demonstrate the convergence and to form the neutrophil attractor.

Previously, Huang et al. indicated the convergence of atRA and DMSO trajectories in the space spanned by the first two principal components based on 2773 high expression genes (2-fold change in expression values from $t_0$) [9]. Notably, our analysis shows that for the high expression genes, neither their correlation trajectory converged nor their SPD boundaries overlapped (Figure 5C). Furthermore, SPD boundaries of the specific genome elements...
Figure 4. Specific genome elements guiding neutrophil cell fate in fractal manner. (A) Sensitivity of genome elements falling to the attractor against standard deviation: left panel, schematic trajectories of genome elements, $G_k(t_0)$ ($k = 1, \ldots, 63$), into (blue) and not into attractor (red), sorted by standard deviation $\sigma$. Right panel, schematic of genome element with single-gene shift can change its fate to the attractor. Illustration shows the process of single-gene shift, i.e., removing the highest $\sigma$ gene from an element $G_k(t_0)$ not falling into attractor, and adding the highest $\sigma$ gene of the next lower rank group $G_{k+1}(t_0)$, creates a new genome element, $G_0(t_0)$, that fall into the attractor. (B) Upper panels: the binary sequence of 63 genome elements atRA and DMSO responses represented by blue and light red, for 1 and 0 respectively. Lower panels, binary sequence of 199 additional genome elements were created using single-gene shift as in (A) for $G_6(t_0)$ for atRA and $G_{11}(t_0)$ for DMSO. (C) Euclidean distance $d$, between the whole genome’s trajectory and each genome element’s trajectory. More than 50% of genome elements that fell into the attractor are close to the minimum distance ($d=0.11$, indicated by a dotted line).

doi:10.1371/journal.pone.0012116.g004
Figure 5. The loss of the attractor when genome vehicles are removed. The SPD boundaries and trajectories for atRA (plain lines with lighter tone) and DMSO (dark dashed lines) responses of (A) genome elements falling into attractor (i.e., genome vehicles) overlap and converge, indicating the formation of neutrophil attractor. (Insert shows overlapping SPD boundaries of atRA and DMSO responses of the whole genomes, indicated by red and blue dotted polygons respectively), (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. Data points are represented by circles and last time
without these highly expressed genes maintained a common neutrophil attractor, albeit with less area (Figure 5D). Thus, the lowly and moderately variable genes within genome vehicles play an important role in the formation of the neutrophil attractor (Figure 5E). These results demonstrate that the collective dynamics of specific gene elements for both atRA and DMSO responses are responsible for the cell fate decision and we call these genome elements that effectively drive cells toward the attractor for each stimulus as “genome vehicle”.

In summary, we show the existence of genome vehicles is responsible for neutrophil differentiation. Despite initial differences of the transcriptional program induced by atRA and DMSO stimulations, the self-regulation of the genome vehicles leads to the formation of a common neutrophil attractor. In addition, we demonstrate that the collective motion of lowly and moderately variable genes within the genome vehicle, which are often considered as noisy and insignificant, play an important role in the formation of the neutrophil attractor, perhaps indicating the non-instructive signalling of genes related to small-amplitude DNA fluctuations within chromatin structure, for the motion of the genome vehicle might decipher dynamics of small-amplitude DNA fluctuations within chromatin structure, connected with the dynamics of chromatin structural changes, non-instructive signaling of genes related to small-amplitude DNA fluctuations, and whole) is defined by

\[ I(x, y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \]

where \( X_i = x_i - \bar{x} \) and \( Y_i = y_i - \bar{y} \), \( x_i \) and \( y_i \) are gene expression of the \( i \)-th gene of expression vectors \( X \) and \( Y \), respectively, \( \bar{x} \) and \( \bar{y} \) are mean values, and \( \theta \) is the angle between the two \( N \)-dimensional expression vectors from the center of mass. Thus, this form of analysis of gene expressions reveals linear relationship, e.g., \( \theta = 0 \) for perfect positive linear relationship, \( \theta = \pi/2 \) for perfect negative (anti-correlated) linear relationship and \( \theta = \pi/2 \) for linearly independent relationship.

However, if the relationship between the two vectors is non-linear, then Pearson correlation analysis is insufficient. In such cases, the use of mutual information has been instrumental in biology [29–30]. In this paper, we adopted both a modified version of Pearson correlation (see below) and mutual information, to investigate the whole genome collective dynamics in the process of neutrophil differentiation to two distinct stimuli (atRA and DMSO), revealing the existence of neutrophil attractor as well as the whole genome expression dynamics toward the attractor.

### Modified Pearson correlation \( r_v \) for measuring expression variation

Each stimulus’s dynamic genome expression activity (partial and whole) is defined by \( N \)-dimensional gene deviation-from-average expression vectors at time \( t_i \), \( i = 0, 1, \ldots, \mathcal{M} \),

\[ \overrightarrow{V}(t_i) = (v_{11}(t_i), v_{12}(t_i), \ldots, v_{1M}(t_i)) \]

where \( v_{ij}(t_i) \) is expression value of the \( j \)-th gene at \( t_i \), \( \overrightarrow{V} \) is its average expression over \( \mathcal{M} + 1 \) discrete time points, and \( v_{ij}(t_i) = x_{ij}(t_i) - \bar{V} \) is its deviation from the average gene expression (called gene deviation). In our study, \( N = 12,625 \) genes/ORFs and \( M = 12 \), where \( t_i = 0, 2, 4, 8, 12, 18, 24, 48, 72, 96, 120, 144, 168 \)h. We modified the typical Pearson \( r \) (see Eq.1) to be \( r_v(\overrightarrow{V}(t_i); \overrightarrow{V}(t_0)) = \frac{\overrightarrow{V}(t_i) \cdot \overrightarrow{V}(t_0)}{|\overrightarrow{V}(t_i)||\overrightarrow{V}(t_0)|} \) (see Figure S3 for \( r_v \) vs. \( \varphi \)) by subtracting their average expression value, \( \overline{\overrightarrow{V}} \), from each expression value at all time points, instead of subtracting the mean of whole genome expression, \( \overrightarrow{\bar{V}} \). This index thus measures the temporal correlation of genome-wide expression deviations from their average values so as to allow discriminating gene expressions with different amplification but similar temporal profiles. For simplicity we included ORFs as genes, and for the microarray data, we applied RMA normalization which is known to produce robust reproducible results for all range of expression units [31].

### Mutual information \( I \)

Nonlinear dependency between vectors \( \overrightarrow{V}(t_i) \) and \( \overrightarrow{V}(t_0) \) is checked by mutual information [32–33]

\[ I(\overrightarrow{V}(t_i); \overrightarrow{V}(t_0)) = \sum_{x \in \overrightarrow{V}(t_i)} \sum_{y \in \overrightarrow{V}(t_0)} p(x,y) \log(p(x,y)) - \sum_{x \in \overrightarrow{V}(t_i)} p(x) \log(p(x)) - \sum_{y \in \overrightarrow{V}(t_0)} \log(p(y)) + \sum_{x \in \overrightarrow{V}(t_i)} \sum_{y \in \overrightarrow{V}(t_0)} p(x,y) \log(p(x,y)) = -\epsilon \]

where the joint probability distribution function \( p(x,y) \), and marginal probability distribution functions, \( p(x) \) at \( t_i \) and \( p(y) \) at \( t_0 \) are estimated by means of an histogram-based approach by discretizing the gene expression into \( K = 10 \) bins [32]. Note: due to the discretization, mutual information \( I \) incurs a systematic error \( \epsilon \) [32]. Since randomly ordered data should destroy correlations, we expect \( I \) to be close to zero, therefore, we calculated the minimum \( I \) for \( 100 \) random permutations of gene deviation vectors \( \{ \overrightarrow{V}(t_i) \} \). However, we found a positive value for minimum \( I \) instead of zero, and so subtracted this minimum positive constant value from the final \( I \).

For comparing \( I \) of atRA and DMSO response, we used the normalized \( I(\overrightarrow{V}(t_i); \overrightarrow{V}(t_0)) = \frac{I(\overrightarrow{V}(t_i); \overrightarrow{V}(t_0))}{I(\overrightarrow{V}(t_i); \overrightarrow{V}(t_0))} \) and called \( \tilde{I} \) as \( I \) throughout the text.
Ranking gene ensembles

The whole genome deviation vector at \( t_0 \)
\[ V(t_0) = (v_1(t_0), v_2(t_0), \ldots, v_n(t_0)) \]
(\( N = 126225 \)) was sorted according to the standard deviation, \( \sigma = \sum_{j=1}^{n} \sigma_j \), from the highest to the lowest, where \( \sigma_j = \sqrt{\frac{1}{M+1} \sum_{i=0}^{M} v_j(t_i)} \) (\( M = 12 \)). The resultant ranked whole genome vector at \( t_0 \) is represented by
\[ S(t_0) = (x_1(t_0), x_2(t_0), \ldots, x_j(t_0), \ldots, x_N(t_0)), \]
where \( \sigma_j < \sigma_i \) for \( i > j \) and \( \sigma_j \) is the standard deviation of the \( j^{th} \) deviation.

Next, we split the whole genome into \( p \) groups (each having \( n \) genes) at \( t_0 \) so that the whole genome at \( t_0 \) is represented by
\[ G(t_0) = \bigcup_{k=1}^{p} G_k(t_0), \]
where
\[ \sigma_k = \sqrt{\frac{1}{M+1} \sum_{i=0}^{M} v_j(t_i)} \] \( \in \{ k \} \) and \( \sigma_k \) is its standard deviation. Note that we choose \( p \) to be an integer value of \( \sqrt{n} \) for \( n = 10, 50, 100, 500, 1000, \) and the residual genes were not evaluated. From here onwards, we simplified all notations without \( \sigma \) symbols, e.g., \( G_k(t_0) = G_k(t_0) \).

The set of \( p \) groups’ average gene deviation from the whole genome at \( t \) is represented by \( \overline{G}_k(t) = \bigcup_{k=1}^{p} \overline{G}_k(t) \), where
\[ \overline{G}_k(t) = \sqrt{\frac{1}{n} \sum_{i=0}^{n} v_j(t_i)} \] \( \overline{G}_k(t) \) and \( i = 0, 1, \ldots, 12 \).

Supporting Information

Figure S1 Comparing CV with \( \sigma \), Standard deviation, \( \sigma \) versus CV for genome elements of \( n \) = 200 genes sorted by \( \sigma \).

Found at: doi:10.1371/journal.pone.0012116.s001 (0.18 MB TIF)

Figure S2 Identifying genome elements that form genome vehicles. (A) Schematic of iterative procedure to exhaustively determine genome elements falling into attractor (i.e., the genome vector, see main text). (B) Number of genome elements falling into the attractor with respect to the number of iterations. We terminated the iteration procedure until the 4 consequential iterations do not constitute any genome element falling into the attractor. Since the number of gene shift other than 50 is not sensitive to the characteristics of the genome vehicle, we chose 50 genes shift to save the computational time. We obtained a total of 21 and 20 genome elements constituting the genome vehicles for arRA for DMSO, respectively.

Found at: doi:10.1371/journal.pone.0012116.s002 (0.56 MB TIF)

Figure S3 Relationship between \( r_1 \) and \( r_2 \) and \( r_3 \) are obtained for \( n = 200 \) randomly selected genes with 3000 repeats (each represented by a dot) from the entire data containing 13 time points (\( i = 0, \ldots, 12 \)).

Found at: doi:10.1371/journal.pone.0012116.s003 (0.30 MB TIF)

Acknowledgments

We thank Ken Showalter of the West Virginia University, Midori Hashimoto and Shohei Sonohara of the Keio University for stimulating discussions. We dedicate this work to our late friend and mother (of KS), Lilly Pillay Selvarajao. We appreciate our family and the system immunology group members at IAB, Keio University.

Author Contributions

Conceived and designed the experiments: MT AG MT KS. Analyzed the data: MT VP KS. Wrote the paper: MT VP KS.

References

1. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: Structure and dynamics. Physics Reports 424: 175–308.
2. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell’s functional organization. Nat Rev Genetics 5: 101–113.
3. Nadell CD, Xavier JB, Levin SA, Foster KR (2008) The Evolution of Quorum Sensing in Bacterial Biofilms. PLoS Biol 6: e14.
4. Danino T, Mondragon-Palomino O, Tuining L, Hasty J (2010) A synchronized quorum of genetic clocks. Nature 463: 326–330.
5. Gehlenborg N, O’Donoghue SJ, Baliga NS, Goessmann A, Hubs MA, et al. (2010) Visualization of omics data for systems biology. Nat Methods 7(Supp1): S56–68.
6. Park G, DeSantis T, Farkas I, Voeck T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435: 814–818.
7. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, et al. (2007) Multiple High-Throughput Analyses Monitor the Response of E. coli to Perturbations. Science 316: 593–597.
8. Tsuchiya M, Wong ST, Yeo ZX, Colosimo A, Palumbo MC, et al. (2007) Gene expression waves. Cell cycle independent collective dynamics in cultured cells. FEBS J 274: 2078–2089.
9. Huang S, Eichler G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys Rev Lett 94: 128701.
10. Waddington CH (1957) The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology. New York: Macmillan.
11. Kaufman SA (1969) Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets. J Theor Biol 22: 437–467.
12. Kaufman SA (1993) The Origins of Order. New York: Oxford University Press.
13. Huang S (2009) Reprogramming cell fates: reconciling rarity with robustness. Biosays 31: 546–549.
14. Chodavarapu MS, Case-Green SC, Southern EM (1997) Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. Nucleic Acids Res 25: 1155–1161.
15. Forman JE, Walton ID, Stern D, Rava RP, Trulson MO (1998) Thermodynamics of duplex formation and mismatch discrimination on photolithographically synthesized oligonucleotide arrays. In: Leuton NB, JohnSantaLucia, ed. Molecular Modeling of Nucleic Acids. Washington DC: American Chemical Society. pp 206–229.
16. Southern E, Mir K, Shchepinov M (1999) Molecular interactions on microarrays. Nat Genet 21: 5–9.
33. Steuer R, Daub CO, Selbig J, Kurths J (2004) Measuring Distances between Variables by Mutual Information. In: Baier D, Wernecke KD, eds. Innovations in Classification, Data Science, and Information Systems, Proceedings of the 27th Annual GiKI Conference. Springer Verlag. pp 81–90.

34. Venables WN, Ripley BD (2002) Modern Applied Statistics with S Springer Verlag.