Coherent organization in gene regulation: a study on six networks

Neşе Aral and Alkan Kabakçöglu
Department of Physics, Koç University, Rumelifeneri Yolu Sarıyer 34450, Istanbul, Turkey
E-mail: naral@ku.edu.tr

Keywords: gene regulation, coherence, Hebbian selection, Boolean network, cell cycle, differentiation

Abstract
Structural and dynamical fingerprints of evolutionary optimization in biological networks are still unclear. Here we analyze the dynamics of genetic regulatory networks responsible for the regulation of cell cycle and cell differentiation in three organisms or cell types each, and show that they follow a version of Hebb’s rule which we have termed coherence. More precisely, we find that simultaneously expressed genes with a common target are less likely to act antagonistically at the attractors of the regulatory dynamics. We then investigate the dependence of coherence on structural parameters, such as the mean number of inputs per node and the activatory/repressory interaction ratio, as well as on dynamically determined quantities, such as the basin size and the number of expressed genes.

List of abbreviations
GRN Gene regulatory network
Th T-helper lymphocyte

1. Introduction
Gene regulatory networks (GRNs) constitute the backbone of intracellular functional organization at the molecular scale. These interaction networks are the key to understanding the clockwork operation of a cell’s life cycle [1], mechanisms of response to environmental changes [2], robustness against random fluctuations [3, 4], the effects of mutations [5, 6], embryonic development in higher organisms [7], etc. Thanks to an enormous amount of data generated by recent experimental and computational efforts, we now have access to gene expression profiles in continuous time and can use them to deduce underlying regulatory interactions [8, 9], even speculate on the evolution of such interactions in historical time scales [10, 11].

Inferring the global GRN of an organism from time-resolved gene expression data is an ongoing challenge [12, 13]. Therefore, the past decade witnessed a growing interest in identifying the key principles that govern the structural organization of the GRNs [14–17]. Viewing the regulatory network as a collection of functional subunits is a popular paradigm [18–22], supported by the observation that certain motifs are frequently encountered in the regulatory networks of many organisms [22]. The GRN structure is ultimately determined and constrained by the requirement that the regulatory dynamics delivers a timely production of necessary proteins. The fact that some of the frequently encountered motifs promote dynamical stability and robustness to minor failures is therefore not surprising [23]. Controllability (the ability to direct the network to a specific trajectory by choosing the initial inputs appropriately) recently emerged as another defining feature of these complex systems [24], stressing the requirement for a better understanding of the interplay between the network architecture and its dynamical behavior. Despite the success of such approaches, it has been pointed out that there is need to develop new methods taking different edge signs into account [25]. Present investigation of coherent activity in gene regulatory networks is a progress in this direction.
Coherent regulation:
Protein production constitutes about one-half of raw material and energy consumption within a growing bacterial cell and one-third for a differentiating mammalian cell [26–28]. Therefore, it is plausible to ask whether the gene regulation hardware is wired in a way to achieve the desired functionality with minimal use of these resources. Considering the structure-dynamics relation from the perspective of energy efficiency, we propose and provide evidence that the GRN architecture has been partly shaped to promote utility of purpose among simultaneously expressed genes sharing a common regulatory target. We refer to such cooperative action of regulatory genes as ‘coherent regulation’ [29].

The idea that the evolutionary pressure for economy may have shaped regulatory interactions is not new; for example, it has been exploited earlier to identify the class of Boolean functions that better model regulatory dynamics [30], or to investigate the frequency of gene duplication in microbes [31]. We claim that, network structures with energy-optimal functionality should be wired to suppress the expression of ‘opposing minority’ regulators. These are transcription factors which, if expressed, would oppose but not significantly alter the target gene’s fate due to outweighing regulatory pressure favoring the status quo. Networks where such minority influences are suppressed would display a disproportionate degree of consensus among simultaneously expressed regulatory elements acting on a common target, i.e., exhibit coherent regulation.

Note that, the definition of coherent regulation here is different from that used in the context of robustness analysis [32]. Yet another use of similar terminology appears in the categorization of network motifs [33], where the coherence of a motif is determined according to the compatibility of alternative directed paths connecting two nodes. In contrast, the degree of coherence defined in the present work is not only a function of the interactions (edges) in the network, but also of the expression states of genes (nodes).

Investigation of coherence on GRNs requires information about the regulatory machinery in the cell; in particular the architecture (say, in the form of a directed graph) and the character (activation/inhibition) of interactions, as well as a detailed knowledge of the regulatory dynamics. Dynamical aspects of genetic regulation have been investigated both on small motifs composed of a few genes [34], and on larger networks [35, 36]. Depending on the desired resolution, Boolean models [37–40], Petri nets [41–43], and differential equation based continuum models [44, 45] are the typical approaches employed for this purpose. A continuum model is indispensable for a high (time-) resolution study of regulatory dynamics. Simpler Boolean models have also found a wide area of applicability, mostly in studies where a coarse characterization of the (quasi-)static stationary states is acceptable [38, 46]. These approaches have been successful in modelling the regulation of cell cycle [37, 47, 48], cell differentiation [49–51], circadian clocks [52, 53], etc.

We test our hypothesis on Boolean systems due to their simplicity and accessibility.

The organization of the paper is as follows: section 2 is a formal introduction to the Boolean network dynamics and coherent regulation. In section 3, we introduce six regulatory networks of different organisms or cell types, for which well-established Boolean models of regulation were adopted from the literature. Section 4 reports our results which suggest a bias towards high coherence in these systems, upon comparison with appropriately constructed random networks. Section 5 investigates structural and dynamical features associated with coherent regulation. Finally in section 6, we discuss our findings and provide motivation for further investigation of coherence in complex networks. Overall, the current work extends our earlier observation on a single GRN [29] to multiple organisms or cell types and suggest that a bias towards high coherence may be a generic feature of gene regulation in biological systems.

2. Computational framework

2.1. Time evolution

Following the standart notation, we describe the expression level of a gene by a binary variable with values 0 (silent) or 1 (expressed). Therefore, the state of a GRN composed of $n$ nodes at a given time $t$ is a binary vector $\sigma(t)$ of length $n$, where $\sigma_i(t)$ is the state of the $i$-th gene $(1 \leq i \leq n)$ at time $t$. Time evolution in a deterministic setting (which applies to all the models considered here) is described by an evolution operator $T: \sigma(t + 1) = T[\sigma(t)]$. Given an initial condition $\sigma(0)$, the ultimate fate (steady state) of the system is a cycle of length $q$, where a member state $\sigma^*$ of the cycle satisfies the condition $\sigma^* = T^q[\sigma^*]$. A fixed point is a trivial cycle with $q = 1$.

The time evolution equations of gene regulatory network models generally accommodate multiple attractors. Some of these attractors exhibit gene activity patterns which are in agreement with the experimentally observed protein levels in corresponding organisms. Hence we classify them as ‘biological attractors’. In contrast, remaining ‘non-biological attractors’ have no matching expression pattern observed. The numbers of all possible attractors and biological attractors for each cell type can be found in table 1. Similarly, we will refer to GRN models designed to reproduce the expression dynamics observed in a living organism (those listed in section 3)
Table 1. Structural/dynamical features associated with the six regulatory networks adopted from the literature. ‘c.c.’ and ‘diff.’ refer to regulation of cell cycle and differentiation respectively. The number of attractors are given in the last two columns, where ‘all’ refers to all attractors permitted by the dynamics and ‘bio’ indicates only the experimentally observed biological attractors. They both exclude the trivial point attractor \( \sigma = 0 \).

| Organism/Cell type     | Func.          | Network parameters | # of attr. |
|------------------------|----------------|--------------------|------------|
|                        |                | n     | k     | p     | all | bio. |
| *S. pombe*             | c.c.           | 10    | 2.3   | 0.35  | 13  | 1    |
| Mammalian cell         | c.c.           | 10    | 3.1   | 0.32  | 2   | 2    |
| *S. cerevisiae*        | c.c.           | 11    | 2.64  | 0.51  | 6   | 1    |
| *A. thaliana* whorls   | diff           | 11    | 2     | 0.55  | 6   | 4    |
| Myeloid progenitor     | diff           | 11    | 2.36  | 0.42  | 5   | 4    |
| Th-lymphocyte          | diff           | 12    | 1.67  | 0.65  | 6   | 2    |

The fixed points or limit cycles of the dynamics, in which a cell spends most of its time. Labelling different steady states with \( s \), the global coherence coefficient of the network is expressed as

\[
\alpha_c = \sum_i \frac{w_i}{q_i} \sum_j \alpha_i [\sigma^j] 
\]

(3)

where \( q_i \) is the length of the limit cycle \( s \) and \( w_i \) is a weight (such as the relative basin size, i.e., the fraction of initial states that end up in the given attractor) subject to the condition \( \sum_i w_i = 1 \).

Null-model ensembles:

Bias towards coherent regulation in a GRN can be assessed by comparing \( \alpha_c \) for the given network with the distribution of the same quantity in a representative ensemble of similar networks. We construct such a reference ensemble separately for each regulatory network detailed in the next section. The ensemble networks were generated by the edge shuffling method described in [57], which warrants a uniformly distributed ensemble of random networks when the edge-pairs are switched sufficiently many times. Hence the resulting network

1. is a connected graph,
2. strictly conserves the self edges along with the number of incoming and outgoing activatory/inhibitory interactions separately for each node,
3. statistically has no correlation with the original network, except for the local similarities imposed by the two constraints above.

While this is one of the simplest possible choices for a null-model, there exist other meaningful alternatives for ensemble selection (e.g., [58]).

By construction, each gene in these random ensembles is locally subject to the same number of repressory and activatory inputs as in the original network, albeit from possibly different regulatory
partners. Under the same rules of dynamics the fixed points and the corresponding $\alpha_c$ values are generally different for each random network, since they are determined by the new global structure. For each ensemble, we generated $10^4$ non-isomorphic networks and calculated their coherence coefficient both with and without a basin-size dependent weight ($w_b$) assigned to each dynamical attractor. In this procedure, isomorphism was ruled out by comparing the eigenvalue spectra of the (signed) adjacency matrices for different networks. If a previously encountered eigenvalue spectrum (recorded as a sorted list of eigenvalues truncated to 3 significant digits) was observed, the corresponding network was eliminated and a new one was generated. We have not enumerated the frequency of such occurrences, however a test run yielded 3 out of $10^4$, suggesting that they are rare.

### 3. Investigated genetic regulatory networks

In this paper, we investigate the degree of coherent regulation in six GRNs associated with different organisms or cell types: cell-cycle networks in *Saccharomyces cerevisiae* (budding yeast) [46], *Schizosaccharomyces pombe* (fission yeast) [37] and mammals [38], cell differentiation networks of *Arabidopsis thaliana* whorls [49], Th-lymphocyte [59] and myeloid progenitors [55]. A graphical description of each GRN is given in figure 2. These dynamical models were chosen from the literature, subject to the criterion that they reproduce the experimentally observed steady-state expression profile(s) after truncation to Boolean variables. Below, we give a brief description of each GRN.

**Schizosaccharomyces pombe** (fission yeast) cell cycle: The fission yeast cell-cycle network was modeled by Davidich and Bornholdt [37] as a network with 10 nodes (figure 2(a)). The dynamics is governed by threshold functions which can also be interpreted as majority rules (see table 7) and yield 12 fixed points along with a fixed cycle. The fixed point with the largest basin matches the biological $G_1$ phase.

**Mammalian cell cycle**: Fauré et al [38] analysed the regulation dynamics with synchronous, asynchronous, and mixed updating schemes in this GRN model composed of 10 key regulatory elements (figure 2(b)). Regulation dynamics is given in terms of logical expressions (see table 6) which are determined according to available experimental evidence for each node. The resulting dynamical attractors are independent of the updating scheme and include a fixed point and a limit cycle, in agreement with the experimental expression data. Note that, the visual depiction of the GRN given in reference [38] is inconsistent with the
used logic update functions. We here remained faithful to the given logical expressions, after verifying that they reproduce the reported steady states. The structure of the GRN consistent with the interactions in table 6 is given in figure 2(b).

Saccharomyces cerevisiae (budding yeast) cell cycle:
The model proposed by Li et al [46] is composed of 11 nodes (figure 2(c)). This popular model reproduces the G1 phase as the dominant attractor of a simple Boolean dynamics, as well as the intermediate phases of cell division. Time evolution is governed by threshold functions (majority) rule (Table 7), which yield 7 fixed point attractors. The attractors other than G1 have relatively small basins and, to our knowledge, no clear biological interpretation.

Arabidopsis thaliana whorl differentiation: Mendoza et al [54] use the network in figure 2(d) in order to model the dynamics of flower morphogenesis. The interactions in the 11-node model network are again inferred from experimental data. Different initial conditions evolved by the proposed (majority) rules (table 7) of dynamics yield 6 point attractors, 4 of which have a clear biological interpretation.

Myeloid differentiation: A Boolean model was set up by Krumsieck et al [55] in order to understand the mechanisms underlying myeloid differentiation from common myeloid progenitors to megakaryocytes, erythrocytes, granulocytes and monocytes. The 11-node model network (figure 2(e)) is composed of relevant transcription factors which evolve (in time) under separate logical update functions (see Table 5), again inferred from the available experimental evidence. The dynamics gives rise to 5 point attractors, where 4 are in agreement with microarray expression profiles of the mature cell types. It is pointed out that the fifth attractor cannot be realized during physiological hematopoietic differentiation.

Th-Lymphocyte differentiation: Remy et al [56] proposed this regulatory network model for the differentiation of T-helper lymphocytes (Th0) cells into Th1 and Th2 in the vertebrate immune system. Discrete-time evolution of each node in the model network is governed by node-specific logical rules given in table 8. The network structure in figure 2(f) is dominated by activatory interactions. The dynamics settles into 3 steady states, in agreement with the gene expression levels in the Th0, Th1 and Th2 cells.

It was shown earlier that structural parameters such as the mean number of incoming edges per node (k) and the fraction of up-regulating interactions (p) in the network are important determinants of global coherence, while the size of the network for fixed (k, p) only weakly influences $\alpha_c$ [29]. Observed values of $(n, k, p)$ for each network are listed in table 1. For the GRNs we consider here, k varies within $[1.67, 3.1]$, while p changes in the interval $[0.32, 0.65]$ (self edges are excluded). All models have roughly the same network size ($\sim$11 nodes).

In the present study we analyze random networks of sizes 10-12 since these are the sizes of the biological networks considered. Nonetheless, the coherence of larger-size Boolean networks can be computed with equal ease. We generated and calculated the attractors and coherence values of thousands of random Boolean networks with more than 100 nodes and $k \leq 10$ without any computational bottleneck in coherence calculations. While generalizing equation (3) for continuous systems is straightforward [29], constructing a random network ensemble with continuous dynamics for comparison may involve its own challenges depending on which constraints one chooses to impose on the ensemble.

### 4. Coherence in biological networks

Table 2 and figure 3 summarize the outcome of our coherence analysis on the GRNs depicted in figure 2 and listed in section 3. We separately calculated the degree of coherence of each GRN over the full set of fixed points/cycles ($\alpha_c$) and over the biologically

| Organism/Cell type | $\alpha_c$ (uniformly weighted) | $\alpha_c$ (basin-size weighted) |
|--------------------|-------------------------------|---------------------------------|
|                    | $\alpha_c$ | $\alpha_c^{\text{bin}}$ | $\alpha_c^{\text{rand}}$ | $\alpha_c$ | $\alpha_c^{\text{bin}}$ | $\alpha_c^{\text{rand}}$ |
| S. pombe            | 0.95        | 1                       | 0.94 ± 0.04               | 0.99       | 1                       | 0.93 ± 0.05               |
| Mammalian cell      | 0.88        | 0.88                    | 0.85 ± 0.07               | 0.88       | 0.88                    | 0.86 ± 0.07               |
| S. cerevisiae       | 0.97        | 1                       | 0.80 ± 0.08               | 0.99       | 1                       | 0.75 ± 0.10               |
| A. thaliana whorls  | 1           | 1                       | 0.97 ± 0.04               | 1          | 1                       | 0.96 ± 0.04               |
| Myeloid progenitor  | 0.89        | 0.89                    | 0.82 ± 0.09               | 0.90       | 0.88                    | 0.84 ± 0.09               |
| Th-lymphocyte       | 0.99        | 0.96                    | 0.94 ± 0.05               | 0.94       | 0.92                    | 0.94 ± 0.05               |


relevant subset \((\alpha_{\text{bi}})\), both adopted from respective references. We compared them with the mean \((\alpha_{\text{rand}})\) obtained from the associated ensemble of random networks described in section 2.2.

Our central observation is that, despite the variability in \(k\) and \(p\), all biological networks are more coherent than random networks constructed with similar structural parameters. Figure 3 also shows the distribution of \(\alpha_{c}\) over the random ensembles for a better judgment of the bias towards coherence. The difference in coherence between the biological network and the random ensemble is visible but within acceptable bounds for most isolated examples. However, cumulatively they suggest an overall preference towards coherence. We quantified the statistical significance of this bias by using Fisher’s method [58] (see section 7) on the present data which returns a combined P-value for the observed coherence values in the background of respective distributions. The results are given in Table 3 for both uniformly and basin weighted cases. For each case we quote two P-values obtained by considering the average coherence value for (i) all of the attractors, and (ii) biological attractors only. The values 0.044 and 0.074 calculated over all the attractors already suggest a bias towards high coherence. This observation is accentuated by the fact that the subset of attractors which are found in nature (biological attractors) display a higher preference for coherence, as quantified by the smaller P-values 0.016 and 0.034. The degree of selection is appreciable, considering that the compared ensemble networks have not only the same number of nodes, the same number of interactions, and the same proportion of repressory/activatory regulation globally, but also the same local pattern of incoming and outgoing edges for each node.

Robustness of our results were checked by a comparison of the distributions in figure 3 for sizes \(N = 100, 500, 1000, 2000, 5000, 10000\). Note that the shapes of the distribution are determined by the structural features that characterize the ensembles and the rules of dynamics, both of which differ for each ensemble. Therefore, there is no a priori reason for them to converge to a Gaussian or any other universal distribution, especially when the network size is kept finite. Instead, stabilization of our results for \(N = 10^4\) was verified by means of Jensen-Shannon (JS) divergence, which is a measure of

### Table 3. P-values for the observed data.

|                      | Uniformly weighted | Basin weighted |
|----------------------|--------------------|----------------|
| All attr.            | 0.044              | 0.016          |
| Bio. attr.           | 0.074              | 0.034          |

Figure 3. Degree of coherent regulation for biological versus random networks. The histograms of \(\alpha_{c}\) for random ensembles \((\alpha_{c}^{\text{rand}})\) of \(10^4\) networks having similar structure is shown with their mean values (green line). The red line indicates the \(\alpha_{c}\) of the biological networks \((\alpha_{c})\) and the blue line corresponds to the \(\alpha_{c}\) values of the biological attractors only \((\alpha_{c}^{\text{bio}})\), whose exact values can be seen on Table 2.
the difference between two distributions [60]. For each value of \(N\) above, we generated 1000 random pairs of ensembles of size \(N\) and compared the mean JS divergence between such pairs. The result is given in figure 4 which shows that the distributions in figure 3 and therefore the results in table 2 stabilize for \(N \geq 2000\).

5. Coherence versus structural/dynamical network properties

The bias observed in biological regulatory networks above provides motivation to investigate the relationship between coherence and other architectural or dynamical determinants of a network. Identifying such connections could help one recognize coherent systems from certain telltale patterns instead of requiring detailed information about their function, and/or design them by means of simple guiding principles.

**Edge number & type:**

We first investigated the dependence of average coherence \(\tau_c\) on the number of incoming/outgoing edges per node, \(k\), and the fraction of activatory interactions, \(p\). To this end, we simulated the majority rule dynamics in [46] on different \((k, p)\) pairs equally spaced in the rectangle \(0.909 \leq k \leq 10 \quad \otimes \quad 0 \leq p \leq 1\), with \(n = 11\). The result is shown in figure 5. Each data point in figure 5 is an average over \(10^3\) random networks which are generated by random shuffling of edges. No constraint was imposed on the shuffling process, apart from the requirement of connectedness (every node is accessible from every other node in the undirected network).

One observes that, independent of the average connectivity \(k\), random networks display minimum coherence when approximately two-thirds (60%—72%) of the interactions are activatory. This behaviour around \(p = 2/3\) is the result of two factors that influence

---

**Figure 4.** The Jensen-Shannon divergence between two different distributions of the same sample size \((N)\), decreases to zero for \(N > 2000\), indicating that ensemble size \(N = 10^4\) yields a stabilized distribution.

**Figure 5.** Dependence of coherence value \((\tau_c)\) on the network parameters \((k, p)\) with \(n = 11\). As \(k\) increases from 1 to 10, the value of \(p\) where \(\tau_c\) is a minimum for fixed \(k\) shifts from 0.72 to 0.6. Coherence values corresponding to the biological networks are shown by spheres.
coherence: (i) the number of active neighbors per node (the larger it is, the higher the chances of receiving conflicting inputs), (ii) the ratio of activatory/inhibitory incoming edges ($p$). If the number of active nodes were determined independent of $p$, one would expect a minimum of coherence at $p = 1/2$, where a typical node is most likely to have two inputs of both types. However, $p$ also indirectly determines the number of active nodes in the system (larger $p$ values mean a larger fraction of activatory interactions, therefore a larger number of active nodes on average). Both factors above will work towards increasing coherence if $p$ is reduced from $1/2$. However, they will compete for $p > 1/2$. Therefore, one expects the coherence minimum to be located at a value $p^* > 1/2$, where the two factors above cancel each other for deviations around $p^*$.

Coordinates corresponding to the studied biological networks are shown by colored spheres on the same figure. It is interesting that all of them are situated on the low-$p$ slope of the minimum coherence valley, where a plateau of high $\alpha_c$ allows for variability in $k$ and $p$ without sacrificing coherence.

**Number of active genes:**

We next asked if the number of active genes ($n_a$) at a fixed point/cycle is a determining factor for coherence. After

---

**Figure 6.** Average ratio of active nodes in the point attractors versus average coherence in ensembles of model networks. The average is calculated over $10^4$ networks chosen randomly from the null-model ensemble described in section 2.2

**Figure 7.** Average coherence versus average basin size for the attractors (point or cycles) of $10^4$ networks with similar structure. Note that proposed rules of dynamics in some models appear to prohibit appearance of excessively dominant attractors.
all, coherence reflects the harmony between the regulatory messages originating from these genes. To this end, we considered the attractors found within the ensembles generated by shuffling the edges of each GRN in figure 2. We grouped them according to \( n_a \) and calculated the mean coherence within each group. The results in figure 6 show that attractors with higher number of active nodes are generally less coherent. This behavior may be understood by the intuitive fact that it is more difficult to reach consensus in a large group than in a small one. We remind that, this relationship is valid only within the restricted ensembles generated by the shuffling process described in section 2. One can not speak of a monotonic dependence of coherence on \( n_a \) if, for example, the difference in \( n_a \) between two networks is due to different \( p \) values. This is evident from the fact that \( n_a (p) \) calculated over the network ensembles used in figure 5 trivially increases with \( p \) while \( \pi_k (p) \) is nonmonotonic for any fixed \( k \).

**Basin size:**
It is implicit from table 2 and figure 3 that, coherent attractors are not always those with larger basins. We explicitly examined the variation in coherence as a function of the relative basin size of the attractors. The results (figure 7), obtained over the networks in the randomized ensembles of each GRN separately, suggest no consistent relation between the basin size and coherence. Likewise, biologically relevant attractors [37, 38, 46, 54–56] of GRNs in figure 2 are not always those with the largest basins. This is hardly surprising, since the initial state prior to differentiation or cell division is never determined randomly.

### 6. Conclusion
We investigated the degree of coherent regulation in several regulatory networks responsible for cell cycle and differentiation in various organisms or cell types, by means of a recently proposed measure of coherence. We found that, even though most networks are moderately more coherent than expected (in reference to architecturally similar network ensembles), cumulatively there is a statistically meaningful bias towards coherence. Our findings lend support to the thesis that pressure for coherent regulation is one of the factors driving the evolution of GRNs. Coherence in gene regulation networks resembles Hebb’s rule [61] for neural connections in the brain, which is commonly stated as ‘cells that fire together, wire together’ [62]. Accordingly, it is not difficult to imagine a Hebbian-like mechanism at work in the present context, where incompatible regulatory interactions get eliminated over time. The fact that transcriptional regulatory networks have been observed to mutate faster than a broad set of biological networks [63] suggests that the gene regulatory networks we examined (especially in Lymphocyte, Myeloid and S. cerevisiae with an abundance of transcriptional regulation) may have benefited from such Hebbian selection. Elimination of incompatible interactions may also be achieved by suppressing the expression of proteins involved in such interactions. The negative correlation observed in figure 6 between coherence and the number of expressed genes supports this picture and suggests that while increasing coherence the system simultaneously evolves in a direction which reduces the consumption of energy and material resources within the cell.

We also showed, on randomly generated model GRNs, that coherence is harder to achieve with increasing number of interactions per node, and with an inhibitory interaction ratio around 1/3. Furthermore, coherent networks typically involve a smaller number of active genes at their steady states, compared to arbitrary networks with similar edge composition and local connectivity. It could be interesting to focus on networks following the identical expression pattern in time as, say, the yeast cell-cycle (see e.g., [58]) and check if those with higher coherence are more similar to their biological counterpart. Another question is whether coherent networks encompass other desirable properties such as better controllability, higher robustness in the sense that a temporary change in the state of a node does not drive the system to a different attractor (see for example [64] on S. cerevisiae), or better synchronization ability as observed in yeast cell colonies [65]. Finally, from an architectural perspective, coherence can be viewed as a design choice for any natural or artificial network where inhibitory and activatory interactions coexist. One might then ask how to build a coherent system from scratch, or how to enhance coherence in an existing system with minimal intervention. We hope our findings to trigger further theoretical investigations in these directions.

### 7. Methods
We measure the statistical significance of our results by using the Fisher’s Method [66] which combines the outcomes of several independent tests of significance, in order to determine the probability that the cumulative data is the result of chance [66], given as \( P \)-value. To this end, first one calculates the \( P \)-value, the probability for selecting (from the hypothesized distribution) a value which is equal to or larger than the one observed, for each of the \( k \) individual tests. Then, the calculated \( P \) values (in our case 6 of them) are combined to give the test statistic \( \chi^2_1 = -2 \sum_{k=1}^{6} \ln (p_k) \), noting that the outcome has a \( \chi^2 \) distribution with \( 2k \) degrees of freedom. By evaluating \( \int_{\chi^2_1}^{\infty} \chi^2_1 (x) \, dx \) or by looking at a percentage point table for \( \chi^2 \) distribution one obtains the cumulative \( P \)-value for the observed data.

### Appendix A
We also repeated our analyses by excluding the nodes with zero input while calculating the coherence for an attractor (in the text, these nodes where assumed to be
coherent since they do not receive conflicting inputs). The resulting coherence distributions are shown in figure 8 and the corresponding P-values are given in table 4. We find quantitatively similar results, except for the uniformly weighted estimation of coherence over all attractors, where a significant increase (0.21) in the P-value is observed. A closer look into the data reveals that this outcome is solely due to those attractors in S. pombe which have no biological counterparts, yet they result in a lower-than-mean coherence.

**Figure 8.** Degree of coherent regulation with only input receiving nodes for biological versus random networks. The histograms of $\alpha_c$ for random ensembles ($\alpha_{c\text{rand}}$) of $10^5$ networks having similar structure is shown with their mean values (green line). The red line indicates the $\alpha_c$ of the biological networks ($\alpha_{c\text{bi}}$) and the blue line corresponds to the $\alpha_c$ values of the biological attractors only ($\alpha_{c\text{bi}}$).

**Figure 9.** Comparison of coherence values for the 13 attractors of S. pombe network. The yellow and blue bars represent the calculation over all nodes and input receiving nodes respectively. The only biologically relevant attractor is the first one indicated with red dots.

**Table 4.** P-values for the observed data when only input receiving nodes are taken into account.

| P-value | Uniformly weighted | Basin weighted |
|---------|-------------------|----------------|
| All attr. | Bio. attr.       | All attr. | Bio. attr. |
| 0.210 | 0.018 | 0.081 | 0.040 |
for the proposed GRN when all attractors are treated with an equal weight. The change of the coherence for these attractors is shown in figure 9. Accordingly, these results overall continue to support central observation of our paper that the dynamics of GRNs in biological organisms is biased towards high coherence. We thank an anonymous referee for suggesting the repetition of our analysis after excluding the zero input nodes.

Appendix B

Below, we list the rules of regulatory dynamics for each GRN model adopted from the literature. Note that all the rules listed in tables 5–8 fall into the general class of bipolar Boolean functions [67], for which the regulatory role of each input is unambiguously defined as activatory or repressory.

Table 5. Update functions for myeloid differentiation.

| Network             | Gene   | Boolean Update Function |
|---------------------|--------|-------------------------|
| Myeloid differentiation | GATA-2 | GATA-2 ∧ (GATA-1 ∧ FOG-1) ∧ PU.1 |
|                     | GATA-1 | (GATA-1 ∨ GATA-2 ∨ Fli-1) ∧ PU.1 |
|                     | FOG-1  | GATA-1                  |
|                     | EKLF   | GATA-1 ∧ Fli-1          |
|                     | Fli-1  | GATA-1 ∧ EKLF           |
|                     | SCL    | GATA-1 ∧ PU.1           |
|                     | C/EBP  | C/EBP ∧ (GATA-1 ∧ FOG-1 ∧ SCL) |
|                     | PU.1   | (C/EBP ∧ PU.1) ∧ (GATA-1 ∨ GATA-2) |
|                     | cJun   | PU.1 ∧ Gfi-1            |
|                     | Gfi-1  | C/EBP ∧ Gfi-1           |

Table 6. Update functions for mammalian cell cycle.

| Network            | Gene   | Boolean Update Function |
|--------------------|--------|-------------------------|
| Mammalian cell cycle | CycD   | CycD                    |
| Rb                 | (CycD ∧ CycE ∧ CycA ∧ CycB) ∨ (p27 ∧ CycD ∧ CycB) |
| E2F                | (E2F ∧ CycA ∧ CycB) ∨ (p27 ∧ E2F ∧ CycB) |
| CycE               | E2F ∧ Rb |
| CycA               | (E2F ∧ Rb ∧ Cdc20 ∧ (Cdh1 ∧ Ubc)) ∨ (CycA ∧ Rb ∧ Cdc20 ∧ (Cdh1 ∧ Ubc)) |
|                    | p27    | (CycD ∧ CycE ∧ CycA ∧ CycB) ∨ (p27 ∧ (CycE ∧ CycA ∧ CycB) ∧ CycB) |
| Cdc20              | CycB   |
| Cdh1               | (CycA ∧ CycB) ∨ Cdc20 ∨ (p27 ∧ CycB) |
| UbcH10             | Cdh1   | (Cdh1 ∧ Ubc ∧ (Cdc20 ∨ CycA ∨ CycB)) |
| CycB               | Cdc20 ∧ Cdh1 |
Table 7. Update rules of the regulatory dynamics for the budding yeast (Saccharomyces cerevisiae), the fission yeast (Saccharomyces pombe), and the flower Arabidopsis thaliana.

| Network                        | Node i | Gene       | Update Function |
|--------------------------------|--------|------------|-----------------|
| S. cerevisiae cell cycle       | 1      | Cln3       | \( S_i(t+1) = \begin{cases} 1, & I > 0 \\ 0, & I \leq 0 \end{cases} \) |
|                                | 2      | Cln1,2     |                 |
|                                | 3      | Cdc20      |                 |
|                                | 4      | Mcm        |                 |
|                                | 5      | Swi5       |                 |
|                                | 6      | SBF        |                 |
|                                | 7      | MBF        |                 |
|                                | 8      | Std        | \( S_i(t+1) = \begin{cases} 1, & I > 0 \\ 0, & I < 0 \end{cases} \) |
|                                | 9      | Cdc5       |                 |
|                                | 10     | Cdc11      |                 |
|                                | 11     | Cdc1       |                 |
| S. pombe cell cycle            | 1      | SK         | \( S_i(t+1) = \begin{cases} 1, & I > 0 \\ 0, & I \leq 0 \end{cases} \) |
|                                | 2      | SLP        |                 |
|                                | 3      | PP         |                 |
|                                | 4      | Ste9       | \( S_i(t+1) = \begin{cases} 1, & I > 0 \\ 0, & I \leq 0 \end{cases} \) |
|                                | 5      | Rum1       |                 |
|                                | 6      | Cdc2       |                 |
|                                | 7      | Cdc2       | \( S_i(t+1) = \begin{cases} 1, & I > \theta \\ 0, & I < \theta \end{cases} \) |
|                                | 8      | Wee1       |                 |
|                                | 9      | Cdc25      |                 |
| A. thaliana whorl differentiation | 1    | EMF1       |                 |
|                                | 2      | TFL1       |                 |
|                                | 3      | LFY        |                 |
|                                | 4      | AP1        |                 |
|                                | 5      | CAL        |                 |
|                                | 6      | LUG        |                 |
|                                | 7      | UFO        |                 |
|                                | 8      | B FU       |                 |
|                                | 9      | AG         |                 |
|                                | 10     | AP3        |                 |
|                                | 11     | PI         |                 |
|                                | 12     | SUP        |                 |

\(^*\) in the update functions \( I = \sum_{j=1}^{N} A_{ij} S_j(t) \) where \( A \) is the weighted adjacency matrix and \( \theta \) indicates the threshold values for the nodes. For detailed explanations for these values one can refer to [37, 46, 54].

References

[1] Csikász-Nagy A, Battogtokh D, Chen K C, Novák B and Tyson J 2006 Biophys. J. 90 4361–79
[2] Kashwagi A, Urabe I, Kaneko K and Yomo T 2006 PLoS One 1 e49
[3] Chaves M, Albert R and Sontag E D 2005 J. Theor. Biol. 235 431–49
[4] Garg A, Mohranam K, Di Cara A, De Micheli G and Xenarios I 2009 Bioinformatics 25 1011–9
[5] Sevim V and Rikvold P A 2008 J. Theor. Biol. 253 323–32
[6] Kaneko K 2007 PLoS One 2 e434
[7] Zhou Q, Chipperfield H, Melton D A and Wong W H 2007 Proceedings of the National Academy of Sciences 104 16458–43
[8] Liang K C and Wang X 2008 EURASIP Journal on Bioinformatics and Systems Biology 2008 1–4
[9] Zhou X, Wang X and Dougherty E R 2003 Signal Process. 83 745–61
[10] Kaneko K 2007 PLoS One 2 e434
[11] Brooks A N, Turkarshian S, Beer K D, Yin Lo F and Baliga N S 2011 Wiley Interdiscip. Rev.: Syst. Biol. Med. 5 544–61
[12] Wang Y R and Huang H 2014 Journal of Theoretical Biology 362 53–61
[13] Chai L E, Loh SK, Low S T, Mohamad M S, Deris S and Zakaria Z 2014 Comput. Biol. Med. 48 55–65
[14] Davidson E H 2010 Nature 468 911–20
[15] Peter I S and Davidson E H 2009 FEBS Letters 583 3948–58
[16] Thomas R and Kaufman M 2001 J. Theor. Biol. 21 170–9
[17] Klemm K and Bornholdt S 2003 Proc. Natl Acad. Sci. USA 102 19414–9
[18] Thieffry D 2007 Briefings Bioinf. 8 220–5
[19] Thomas R, Thieffry D and Kaufman M 1995 Bull. Math. Biol. 57 247–76
[20] Kaufman M and Thomas R 2003 C. R. Biol. 326 205–14
[21] Burda Z, Krawczyki A, Martin O and Zagorski M 2011 Proc. Natl Acad. Sci. USA 108 17263–8
[22] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 Science Signalling 288 824
[23] Cinquin O and Demongeot J 2002 C. R. Biol. 325 1085–95
[24] Liu Y Y, Slotine J J and Barabási A L 2011 Nature 473 167–73
[25] Wu Y, Zhang X, Yu Y and Ouyang Q 2009 PLoS Comp. Bio. Comp. Bio. 5 e1000442
[26] Li G W, Burbhardt D, Gross C and Weissman J S 2014 Cell 157 624–35
[27] Russell J B and Cook G M 1995 Microbiol. Rev. 59 48–62
[28] Buttgereit F and Brand M D 1995 Biochem. 312 163–7
[29] Aral N and Kabakçoğlu A 2015 Phys. Biol. 12 036002
[30] Ramôo P, Kessel J I and Yli-Harja O 2005 15 034101
[31] Wagner A 2005 Mol. Biol. Evol. 22 1365–74
[32] Willadsen K, Triesch J and Wiles J 2008 Understanding Robustness in Random Boolean Networks ALIFE 694–701
[33] Alon U 2006 An Introduction to Systems Biology: Design Principles of Biological Circuits (Oxford: CRC press)
[34] Alon U 2007 Nat. Rev. Genet. 8 450–61
[35] Huang S, Eichler G, Bar-Yam Y and Ingber D E 2005 Phys. Rev. Lett. 94 128701

Table 8. Update functions for lymphocyte differentiation.

| Network | Gene       | Boolean Update Function |
|---------|------------|-------------------------|
| IL-4    | \( K_i(9), K_i(11), K_i(9, 11) \) |
| IL-12   | \( K_i(12) \) |
| IFN-\( \gamma \)R | \( K_i(1) \) |
| IL-4R   | \( K_i(2) \) |
| IL-12R  | \( K_i(3) \) |
| STAT1   | \( K_i(4) \) |
| STAT6   | \( K_i(5) \) |
| STAT4   | \( K_i(6) \) |
| SOCS1   | \( K_{SO}(7), K_{SO}(11), K_{SO}(7, 11) \) |
| T-beta  | \( K_{SO}(7), K_{SO}(11), K_{SO}(7, 11) \) |
| GATA-3  | \( K_i(8) \) |

\(^*\) \( K_i(\{ j \}) \) is the set of nodes that when simultaneously active turn the node on. For details one can refer to [36].
[36] Luscombe N M, Babu M M, Yu H, Snyder M, Teichmann S A and Gerstein M 2004 *Nature* **431** 308–12
[37] Davidich M I and Bornholdt S 2008 *PLoS One* **3** e1672
[38] Fauré A, Naldi A, Chaouiya C and Thieffry D 2006 *Bioinformatics* **22** e124–31
[39] Fauré A and Thieffry D 2009 *Mol. BioSyst.* **5** 1569–81
[40] Lau K Y, Ganguli S and Tang C 2007 *Phys. Rev.* **E75** 051907
[41] Murata T 1989 *Proc. IEEE* **77** 541–80
[42] Chaouiya C, Remy E, Ruet P and Thieffry D 2004 Applications and Theory of Petri Nets 2004 137–56
[43] Karlebach G and Shamir R 2008 *Nat. Rev. Mol. Cell Biol.* **9** 770–80
[44] Chen K C, Wang T Y, Tseng H H, Huang C Y F and Kao C Y 2009 *Bioinformatics* **25** 2883–90
[45] Chen T, He H L, Church G M et al 1999 Modeling gene expression with differential equations *Proc. Symp. BioComput.* **5** 44
[46] Li F, Long T, Lu Y, Ouyang Q and Tang C 2004 *Proc. Natl Acad. Sci. USA* **101** 4781–6
[47] Chen K C, Calzone L, Guias N-A, Cross F R, Novak B and Tyson J I 2004 *Mol. Biol. Cell.* **15** 3841–62
[48] Davidich M and Bornholdt S 2008 arXiv:0807.1013
[49] Mendoza L, Thieffry D and Alvarez-Buylla E R 1999 *Bioinformatics* **15** 593–606
[50] Albert R and Othmer H G 2003 *J. Theor. Biol.* **223** 1–8
[51] Gursky V V, Reinitz J and Samsonov A M 2001 *Chaos* **11** 132–41
[52] Leloup J C and Goldbeter A 1998 *J. Biol. Rhythms* **13** 70–87
[53] Akman O E, Watterson S, Parton A, Binns N, Millar A J and Ghazal P 2012 *J. R. Soc., Interface* **9** 2365–82
[54] Mendoza L and Alvarez-Buylla E R 1998 *J. Theor. Biol.* **193** 307–19
[55] Krumrieck J, Marr C, Schröder T and Theis F J 2011 *PLoS One* **6** e22649
[56] Remy E, Ruet P, Mendoza L, Thieffry D and Chaouiya C 2006 *Trans. Comp. Syst. Biol. VII* 56–72
[57] Neto P E, Randrup Y and Stone L 2005 *Phys. Rev.* **E72** 056708
[58] Lau K Y, Ganguli S and Tang C 2007 *Phys. Rev.* **E75** 051907
[59] Remy E and Ruet P 2008 *Bioinformatics* **24** i220–6
[60] Lin J 1991 *IEEE, Trans. Inf. Theory* **37** 145–51
[61] Morris R 1999 *Brain Research Bulletin* **50** 437
[62] Schatz C J 1992 *Sci. Am.* **267** 60–7
[63] Shou C, Bhardwaj N, Larn H Y, Yan K K, Kim P M, Snyder M and Gerstein M B 2011 *PLoS Comput Biol* **7** e1001050
[64] Oguz C, Laomettachit T, Chen K C, Watson L T, Baumann W T and Tyson J I 2013 *BMC Syst. Biol.* **7** 53
[65] Oguz C, Palmisano A, Laomettachit T, Watson L T, Baumann W T and Tyson J I 2014 *PLoS One* **9** e96726
[66] Fisher R A 1950 *Statistical Methods for Research Workers* (Edinburgh: Oliver And Boyd)
[67] Breukw G and Woltran S 2010 *Abstract dialectical frameworks In Proc. KR-2010* pp 102–111