

Interplay of digital and analog control in time-resolved gene expression profiles

Moritz E. Beber^{1,2} , Patrick Sobetzko^{1,3}, Georgi Muskhelishvili¹ and Marc-Thorsten Hütt^{1*}

*Correspondence:

m.huett@jacobs-university.de

¹Department of Life Sciences and Chemistry, Jacobs University, Campus Ring 1, 28759 Bremen, Germany

Full list of author information is available at the end of the article

Abstract

Background: Measuring the agreement between a gene expression profile and a known transcriptional regulatory network is an important step in the functional interpretation of bacterial physiological state. In this way, general design principles can be explored. One such interpretive framework is the relationship of digital control, that is, the impact of sequence-specific interactions, and analog control, i.e., the extent of the influence of chromosomal structure.

Methods and Results: Here, we present time-resolved gene expression profiles of *Escherichia coli*'s growth cycle as measured by RNA-seq. We extend methods which have been developed for discrete sets of differentially expressed genes and apply them to the wild type and two mutant time-series for which the global transcriptional regulators *fis* and *hns* were inactivated. We test our continuous methods using simulated 'expression profiles' generated from random Boolean network dynamics where we observe a clear trade-off between maximum response and level of detail included. In the real time-course expression data, we find strong interdependent changes of digital and analog control during the exponential growth phase and a dominance of analog control during the stationary phase.

Conclusions: Our investigation puts forward a simple and reliable method for quantifying the match between time-resolved gene expression profiles and a transcriptional regulatory network. The method reveals a systematic compensatory interplay of digital and analog control in the genetic regulation of *E. coli*'s growth cycle.

Keywords: Random Boolean networks, *Escherichia coli*, Time-resolved expression, Bacterial gene regulation, Fis, Hns

Background

Even in times, when Systems Biology is coming close to simulating complete organisms in silico [1], bacterial gene regulation is far from being understood. In order to organize gene expression in time and space, i.e., in the right order and quantities across chromosome locations, several evolutionary targets are plausible:

- (1) long-range interactions between genes implemented by dedicated transcription factors (TFs),
- (2) stabilization of mesoscopic or macroscopic DNA structures and
- (3) binding site affinity.

Just as regulation by TFs, nucleation of large-scale chromosome structures require specific DNA sequences. Dynamic DNA loops, conducive to transcription, and stiff, tightly interwound regions that are largely inaccessible, are induced and stabilized by nucleoid associated proteins (NAPs). The local shape of the binding energy landscape therefore has a direct impact on the global chromosome conformation.

Furthermore, the gene expression patterns serve cellular functions. Qualitatively speaking, metabolic function largely dictates the requirements for the spatiotemporal organization of gene expression. The patterns of expression changes have evolved to match the metabolic needs of the organism under the conditions at hand.

Over the last ten years a wide range of studies provided first evidence that all these components are strongly interlinked and shape the observed gene expression patterns (see, e.g., [2–4]). The interplay of genome architecture and global regulators [5], the genome's spatiotemporal organization [6], as well as the relationship of sequence information and genome architecture [7] are challenging to understand.

On a more general level, the task of comparing dynamical information, e.g., time series available for each node, with a given network occurs in different variations in a wide range of disciplines: In neuroscience the problem of relating functional connectivity, i.e., a network derived from correlations in node activities, with structural connectivity, e.g., a network of cortical areas connected by fibers, is currently of high relevance [8–10]. In systems biology, the comparison of metabolomics data with metabolic networks belongs to the same class of statistical problems (see, for example, [11, 12]). A related question is *network inference*, i.e., how to estimate interaction networks from dynamical data. For gene regulatory networks this is exemplified in [13, 14] and for the reconstruction of microbial interaction networks from microbiome compositions in [15]. In this context, the whole research field of Bayesian networks needs to be mentioned [16]. A recent review of a wide range of such network reconstruction, inference, and modeling methods is given in [17].

At the core of many of these research directions are methods from the statistical physics of complex networks since the organization of dynamical processes on graphs belongs to the most prominent research questions in this field (see, e.g., [18–20] for reviews). In our investigation we resort to essentially the same set of concepts and methods: The study of random Boolean networks has a rich history in statistical physics (see, e.g., [21]), even before it gained such recent prominence in systems biology [22–24].

The results of a wide range of our own investigations over the last years have their foundations in statistical physics and information theory [2, 4, 25–31]. They have cemented the notion of a tight interplay of the regulatory network implemented via TFs and their binding sites (digital control); and the regulation implemented via alterations of chromosomal configuration and DNA compaction (analog control) in bacterial gene regulation (depicted in Fig. 1). Our findings show that this interplay is not only revealed by the dynamical quantities, i.e., the control strengths [25, 32], but also clearly visible on the structural level [33]. The distribution patterns of genes under regulation by TFs and of genes without (known) regulation by TFs are fundamentally different. Using methods from point process statistics, we observed systematically shorter distances among the latter class of genes, suggesting a higher importance of (distance-driven) analog control [33].

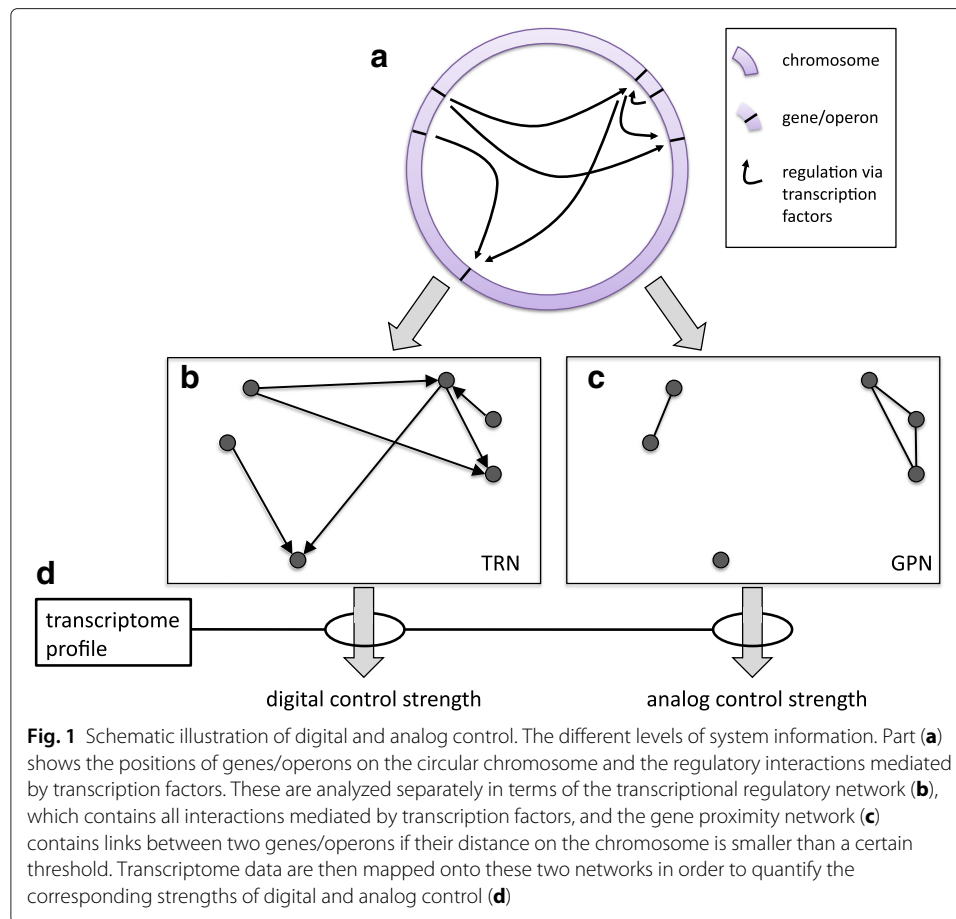


Fig. 1 Schematic illustration of digital and analog control. The different levels of system information. Part (a) shows the positions of genes/operons on the circular chromosome and the regulatory interactions mediated by transcription factors. These are analyzed separately in terms of the transcriptional regulatory network (b), which contains all interactions mediated by transcription factors, and the gene proximity network (c) contains links between two genes/operons if their distance on the chromosome is smaller than a certain threshold. Transcriptome data are then mapped onto these two networks in order to quantify the corresponding strengths of digital and analog control (d)

In [27] the two arms of the chromosome from chromosomal replication origin (*OriC*) to terminus (*Ter*) have been established as an appropriate coordinate system to interpret the spatiotemporal changes in gene expression. When going to a more localized version of the formal control strength, strong shifts of importance between digital and analog control are observed across different cellular conditions and growth phases. Generally, analog control is organized *along* these linear chromosomal segments from *OriC* to *Ter*, while digital control is organized *across* these segments.

Finally, by applying the methods of quantifying digital and analog control to the subset of gene expression changes consistent with predicted metabolic flux patterns (via flux-balance analysis, see [33]), we found that these expression changes, which are apparently instrumental in generating a coherent metabolic state, are predominantly under analog control.

In [32], we established methods for quantifying the amount of digital and analog control contained in a set of significantly, differentially expressed genes for the bacterium *Escherichia coli*. In that way, we have discovered a tight interplay between digital and analog control governing gene expression patterns in response to diverse perturbations of the gene regulatory machineries.

Our general view is that gene expression data can be decomposed into ‘sub-patterns’ compliant with (or dictated by) a certain regulatory mechanism, like the transcriptional

regulatory network (TRN) or DNA topology. The distinction of (and coupling between) digital and analog control has also been supported by a statistical analysis of gene locations [33].

In order to further test and validate the previously formulated hypothesis of a buffering between digital and analog control [32], we have measured gene expression time courses for wildtype *E. coli* and two mutants. This allows us to verify the previous findings based on a more modern platform, RNA-seq, and to include the temporal dimension in our, now continuous, assessment of digital and analog control.

Before applying the new methods to real expression profiles, we assess the performance of digital control strength quantifiers using the framework of random Boolean networks. Lastly, we discuss the implications of our findings for our understanding of bacterial gene regulation.

Methods

Cell growth conditions and mRNA isolation

The *E. coli* CSH50 Δfis and Δhns strains were grown in 4 L of double rich medium (dYT) in a fermenter under constant pH 7.4 and high aeration (500 rpm stirring, 5 L air per min) at 37 °C. The culture was inoculated from 16 h overnight cultures at an initial OD₆₀₀ of 0.1. Cells were grown for 7 h and samples for RNA-seq were taken at 1, 2, 3, 5 and 7 h after inoculation (see Table 1). Each sample was immediately dissolved in ice-cold ethanol-phenol (5 % phenol) solution to prevent mRNA degradation. RNA was extracted using the RNeasy Mini kit (Qiagen) and treated with Turbo DNase (Life Technologies). Subsequent rRNA depletion was carried out using the MicroExpress kit (Life Technologies) and 0.5 μ g of enriched mRNA of each sample were sent for RNA-seq (Illumina HiSeq 2000).

Gene expression analysis

The 50 bp Illumina HiSeq reads were mapped on the *E. coli* MG1655 genome (NCBI). Chromosomal repeats were masked. Gene expression was determined by normalizing the coding sequence (CDS) reads with the total number of reads as well as the length of the CDS. The expression values of each gene at a given time point were normalized to the sum of expression of all genes at that time point. The expression curve of each gene was subsequently interpolated by a natural spline method. Resulting expression curves were verified by fluorescence measurements of yfp-coupled promoters exhibiting different temporal patterns. The wild type data was first published in [31]. The *fis* and *hns* mutant strain data is new to this work.

Table 1 The table denotes the time in minutes after inoculation of *E. coli* cultures in fresh medium when the cells were harvested for sequencing. It shows this information for the wildtype (wt), *fis* and *hns* mutant

Time [min]	wt	<i>fis</i>	<i>hns</i>
60	x	x	x
120	x	x	x
180	x	x	x
300	x		
420	x	x	x

Although RNA-seq generally allows for a higher confidence in its outcomes than microarray data, the unequal spacing of the time points and a lack of replicates for the time points pose a challenge to adequately handle the data. In the original method [32], significantly differentially expressed genes were determined by T-test and then used to determine the number of connected versus disconnected nodes,

$$R = \frac{N(k > 0)}{N(k = 0)}. \quad (1)$$

Assessing the differentially expressed genes in the current data set is much more difficult, however. It is impossible to assess the variance; and applying an approximate fold change of 2 per hour is not possible since more than just the exponential growth phase is observed and the time intervals are uneven. Additionally, the mutant strains have one less time point of data collection. Since RNA-seq data is more reliable, however, we decided to simply use the continuous expression levels directly.

Transferring continuous expression levels to the aforementioned method is relatively straightforward. We are generally only interested in the response of a certain gene and not the absolute value. In addition, all genes should be comparable within a network. For this reason the expression levels of each gene over time were normalized between zero and unity.

There are interpolated expression levels available for almost all genes such that the basis for computing analog or digital control are the complete GPN or TRN, respectively. It is also clear that there are certain expectations for the distribution of relative expression levels. When considering the chromosome structure, it is obvious that genes within a transcription unit (TU) must have the same expression level. Similarly, all genes within one operon should have approximately the same expression level. For this reason we convert the networks under investigation to an aggregated form where nodes are no longer genes and TFs but operons with mean relative expression levels of all genes involved. The links in the network are then from operon node to operon node if there is a link from a gene or TF in one operon to a gene in another operon.

Instead of counting connected differentially expressed genes versus disconnected ones, as in Eq. (1), we now compute a control strength per link that takes into account the relative expression level of each node involved. The sum over all terms per link is then normalized by the total number of links. The continuous control ratio $R_{\text{cont.}}$ as its name suggests is thus no longer discrete.

$$R_{\text{cont.}} = \frac{\sum_{(j,i) \in M} C(i,j)}{|M|} \quad (2)$$

The pairs (j, i) are part of the existing set of links M . The TRN and GPN used for the subsequent analysis are listed as Additional Information: Additional file 1: GPN, gene level; Additional file 2: GPN, operon level; Additional file 3: TRN, gene level; Additional file 4: TRN, operon level.

Control type confidence

In order to evaluate whether an observed control strength is unexpectedly high or low, in [32], a set of control strengths was computed for comparison. Those control strengths were computed in a population of networks that was generated by randomly choosing

the same number of nodes from the complete TRN or GPN as selected by the differentially expressed genes. From those data a z-score was computed which is the control type confidence (CTC) discussed previously.

$$CTC = \frac{R_{cont.} - \mu_{R_{cont.}}}{\sigma_{R_{cont.}}}, \quad (3)$$

where $\mu_{R_{cont.}}$ and $\sigma_{R_{cont.}}$ are the mean and standard deviation of $R_{cont.}$ in the random sample. Here, we proceed similarly but a few distinctions are of note. The RNA-seq expression data include most every gene in the TRN and GPN. In order to introduce any randomness at all and since we use the continuous expression levels, we simply shuffled the index i of all genes (or transcription units, or operons). This resulted in each node of the TRN or GPN ending up with different expression levels. Considering that we actually have an expression matrix over all genes and time points, and considering that each gene's expression profile over time was normalized between zero and unity, we always shuffled and kept intact the entire time series. We measure the control ratio $R_{cont.}$ differently in the case analog or digital control.

Absolute control strength

Computing analog control strength is the more obvious scenario. Links are based on proximity of genes and we expect neighboring genes to have similar relative expression values due to a high likelihood of being located in the same region of analog control. It seems sensible to regard relative expression levels at the same time point and compute the control strength of a link as follows:

$$C_A(i, j) = 1 - |e_i - e_j|. \quad (4)$$

That means, there is a high absolute control strength when relative expression levels e_i and e_j of the neighboring genes i and j are of similar magnitude. For the undirected GPN the distinction of direction has no impact.

Functional control strength

The regulatory function of links in the TRN and thus digital control present much more of a challenge to measure appropriately. The regulatory interaction is mediated by transcription factors which may be liable to any of the following situations:

- (i) Transcription factors are proteins, that means, they are the result of first transcription, then translation and potentially post-translational modification (PTM).
- (ii) They may also depend on co-factors for the right conformation.
- (iii) All these steps may cause a time delay between activation of gene i and a regulatory action at gene j and strongly depend on copy numbers and binding dynamics of the TF.

Obviously, Eq. (4) is a poor approximation of digital control. We can improve it by taking into account the regulatory function of a link. We ignore links that have an unknown or dual role since only a handful of each exist anyway. The following function is applied separately for links A_{ij} that are activating or inhibiting. The matrix A is a variant of the adjacency matrix, which also incorporates the type of the link. The element A_{ij} , denoting

the influence of the j th gene on the i th gene, has the following possible entries: 0 (no link), 1 (activating) and -1 (inhibitory).

$$C_F(i, j) = \begin{cases} 1 - |e_i - e_j|, & \text{if } A_{ij} = 1 \\ |e_i - e_j|, & \text{if } A_{ij} = -1 \end{cases} \quad (5)$$

Random Boolean network model

In order to test and ‘calibrate’ the statistical methods described above, we use a simple model, random Boolean networks, for generating artificial gene expression data. We generate a random Erdős-Rényi (ER) graph with N nodes and M directed links. After that, M_A links are declared activating, while the other $M_I = M - M_A$ links are declared inhibitory. Following the prescription from [23], we then add self-inhibitory links for all nodes which have no incoming inhibitory links. Again, in accordance with the model from [23], the following update rule is used to generate dynamics,

$$x_i(t+1) = \begin{cases} 1, & \text{if } \sum_j A_{ij}x_j(t) > 0 \\ 0, & \text{if } \sum_j A_{ij}x_j(t) < 0 \\ x_i(t), & \text{if } \sum_j A_{ij}x_j(t) = 0 \end{cases}, \quad (6)$$

where $x_i(t)$ denotes the binary, 1 = ON- or 0 = OFF-state of the i th gene at time t .

After a small number of time steps (typically five), such an RNB will settle into a fixed point or a cyclic attractor. These asymptotic states are ideal for analysis with the classical digital control strength defined in [32]. In order to have a longer time series of continuous values available to our analysis, we simulate a large number of short time series (thus sampling transients leading towards these asymptotic states), concatenate them and then compute the average activation of a gene in a certain time window. We use 10^4 short time series, each 10 time steps long and average activity over windows of 10^3 time steps.

The main assumption of the minimal model we used to calibrate our quantification methods is that on a microscopic time scale, ON and OFF are meaningful states of individual genes. We thus neglect (at such a microscopic level) a more gradual description of gene activity. In our RBN formalism, continuous gene expression levels arise as time averages: On this coarse-grained time scale, a high ‘expression level’ means that in a given time window the gene under consideration has been very often in the ON state. A second processing step (in addition to averaging the binary gene states over time windows) is the normalization of the data. In principle, two normalizations are possible here: (1) normalizing the sum over all time points for each individual gene; (2) for each time point, normalizing the sum over all gene activities. In order to match the decision made for normalizing the experimental data (time course of each gene normalized individually) we here select the first normalization variant.

Results and discussion

On the methodological side, the main result of our investigation is to provide and test a new method for analyzing time-resolved gene expression data. With the wide availability of next-generation sequencing methods, high-quality time-resolved data are now becoming rapidly available for diverse biological and biomedical situations.

In the case of time-resolved data, several extensions of the original method from [32] (see also Eq. 1) are possible. For measuring analog control type confidence (CTC) we will use the *absolute control strength* C_A (Eq. 4). This choice is motivated by the expectation that in a neighborhood structure like the gene proximity network (GPN), most genes will be within the same domain of analog control and are expected to have very similar expression levels.

For digital control, represented by the TRN, the situation is less clear. Each link has a regulatory function, an activating or inhibiting interaction, and in reality this function is performed by a transcription factor which may introduce a delay between a signal arriving at one gene and the result of its regulatory function at another gene. Hence, we decided to test these digital control extensions using synthetic data.

Simulated expression data

In order to understand the capability of each of the four definitions of control strengths for time-resolved data (see below), we employ numerical simulations using random Boolean networks (RBNs) to generate synthetic data sets in which all systematic information is generated by the underlying regulatory network. RBNs have been very successful in describing the dynamics of small-scale regulatory systems (see, e.g., [22, 23, 34]). Variants of RBNs have also been used as minimal models of signaling pathways [24, 35]. We therefore expect a very high match between the ‘transcriptome data’ and the network architecture, i.e. high CTC values (see definition in Eq. 3).

Clearly, real transcriptome profiles differ dramatically from the output of such RBNs. Gene expression data are not binary (ON or OFF) but have a broad distribution of values. Also, the actual switching events arising from the regulation are more gradual and also affected by a multitude of other factors beyond the regulation by transcription factors, for example, external stimuli, noise or signaling.

We have therefore applied a sequence of processing steps to the binary data, which are intended to mimic some of those effects. By running the RBN for a few time steps starting from random initial conditions and then putting these runs into a longer time series, we have a stylized version of a system kept in a perpetual transient due to external influences. The details can be found in the Section “Random Boolean network model”.

The networks are equal in size to the latest TRN published by RegulonDB [36] with $N = 1791$ genes, $M_A = 2453$ activating and $M_I = 2095$ inhibitory interactions. Additionally, in accordance with the scheme described in [23], inhibitory self-loops are introduced at nodes that have only incident activating links. Whereas the TRN has a broad, heterogeneous degree distribution, the RBNs’ degree distributions resemble a normal distribution. A sample network and time series of ON/OFF states can be found in Additional file 5: Figure A.1.

Arriving at a RBN prediction of the continuous ‘expression level’ of a node required normalizing the sum of its ON states within a window that includes multiple short time series simulated from different random initial states. The larger the window, the broader the possible levels. Even for small window sizes, the range of values is much wider than for a uniform random choice of states of equal dimension.

In this calibration step using RBNs, several technical aspects of the data processing can be manipulated and their effect on the control strengths can be studied. Such technical aspects are the normalization, the number of initial conditions used, the

simulation time, the size of the averaging window, and the amount of fluctuations in the simulated ‘data’. The key finding of this calibration step is that the absolute control strength C_A (see Eq. 2) is apparently too simplistic to capture the underlying regulatory network, while the three other control strengths clearly identify a high match between the ‘transcriptome data’ and the regulatory network, with all three producing approximately the same levels of CTC values. In our subsequent analyses, we use the simple, best performing *functional control strength* C_F as a measure for the digital component of control. It is described in detail in the Section “Functional control strength”. The full comparison of the different control strengths can be found in the supplements (Section “Definition of control strengths”).

Time-series gene expression data

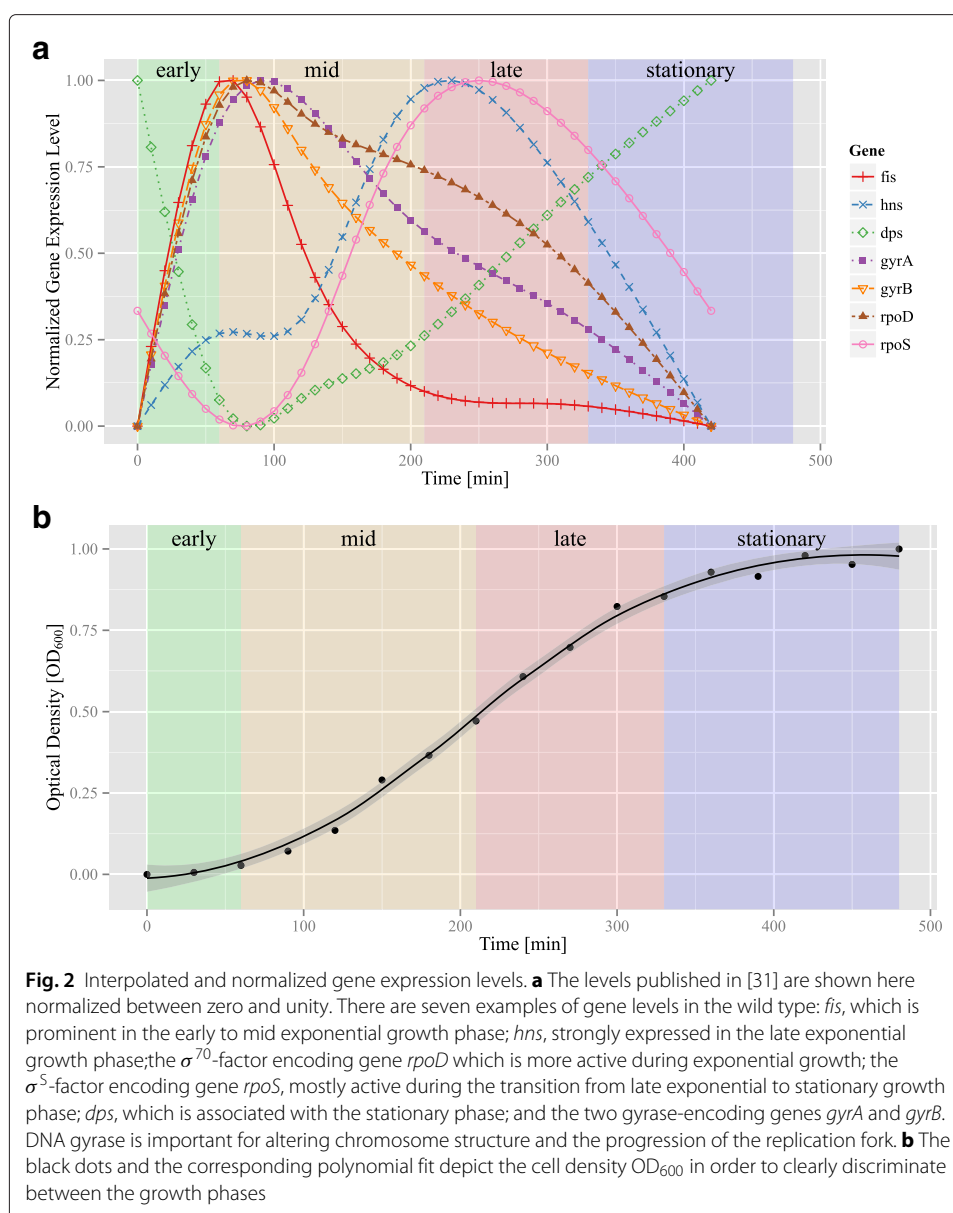
Before application of the newly defined methods, we present an impression of the time-resolved RNA-seq data measured in the wild type (first published in [31]) and the two mutants *fis* and *hns*. The relevant experiments and data transformations are described in Sections “Cell growth conditions and mRNA isolation” and “Gene expression analysis”.

In Fig. 2, we show some examples of gene expression in the wild type. They are either directly relevant to our study, i.e., *fis* and *hns*, or known to be active during different phases of the growth cycle. *dps* is a marker gene of the stationary phase. *gyrA* and *gyrB* encode the subunits of DNA gyrase which is a member of the topoisomerase family and can increase DNA supercoiling. In addition, it is required for DNA synthesis and replication fork progression [37–39]. *rpoD* encodes the σ^{70} factor which is the major constituent of RNA polymerase during exponential growth. *rpoS* encodes the σ^S factor which is abundant during the transition to the stationary phase. Overall, the changes in gene expression levels along the growth curve of the cell culture (black points) shown in Fig. 2 are consistent with established knowledge.

Application to real expression data

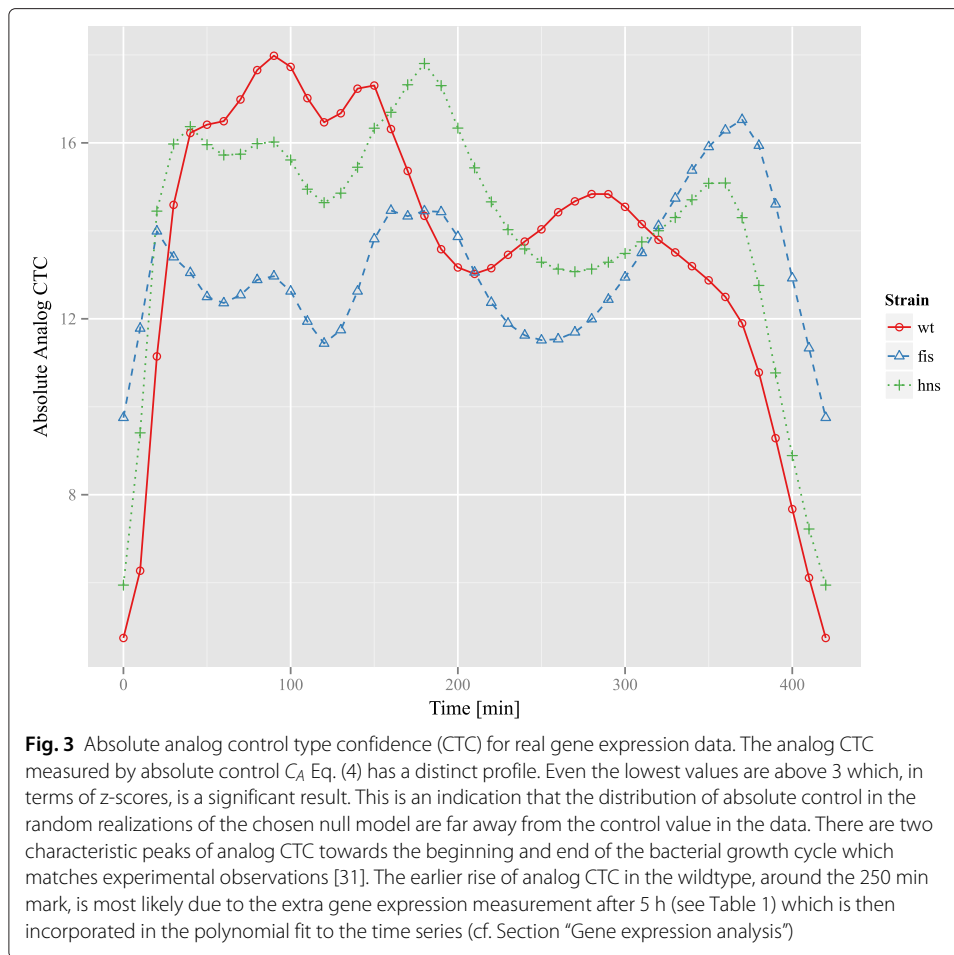
Our first aim is to qualitatively reproduce the (static) observations from [32]. The results in [32] are based on gene expression data obtained during the exponential growth phase which corresponds to the 60 – 120 min time window in the time-resolved data analyzed here. The main goal of our investigation is to evaluate the hypothesis formulated in [32] of a balancing of digital and analog control.

With these general growth cycle-dependent expression results in mind, we can interpret the outcomes of analog and digital control. Importantly, our results are based on the operon projections of the TRN and GPN (see Section “Gene expression analysis” for details). Figure 3 shows the analog CTC for the absolute control strength C_A . The analog CTC measured by *absolute control strength* C_A has a distinct profile. First of all, even the lowest values are above 3 which, in terms of z-scores, is a significant result. This is an indication that the distribution of absolute control in the random realizations of the chosen null model are far away from the control value in the data. Our null model does take into account operon structure but otherwise simply distributes the gene expression values randomly. However, the characteristic peaks of analog CTC in the wild type towards the beginning and end of the bacterial growth cycle match our previous observations presented in Fig. 2. A possibly confounding



factor in the earlier rise of analog CTC, around the 220 min mark, in the wild type as compared to the mutant strains, is the extra gene expression measurement after 5 h (see Table 1) which is then incorporated in the polynomial fit to the time series (cf. Section “Gene expression analysis”).

We only show the functional control C_F results for digital CTC since that is the measure suggested by the RBN results. The complete results can be seen in Additional file 5: Figures A.3, A.4 and A.5 but will not be discussed here. Digital CTC based on C_F (Fig. 4) has a fairly similar profile as compared to analog CTC but shifted by approximately -50 minutes, a sharper first peak and actually fluctuates around zero which is an indicator of a reasonable null model. The digital CTC values at exactly 120 min are in good qualitative agreement with the result in [32] Figure 3b. This is clearly shown in Fig. 5. Both results show the increasing order of wild type, *hns* and *fis* mutant. Effects of the additional

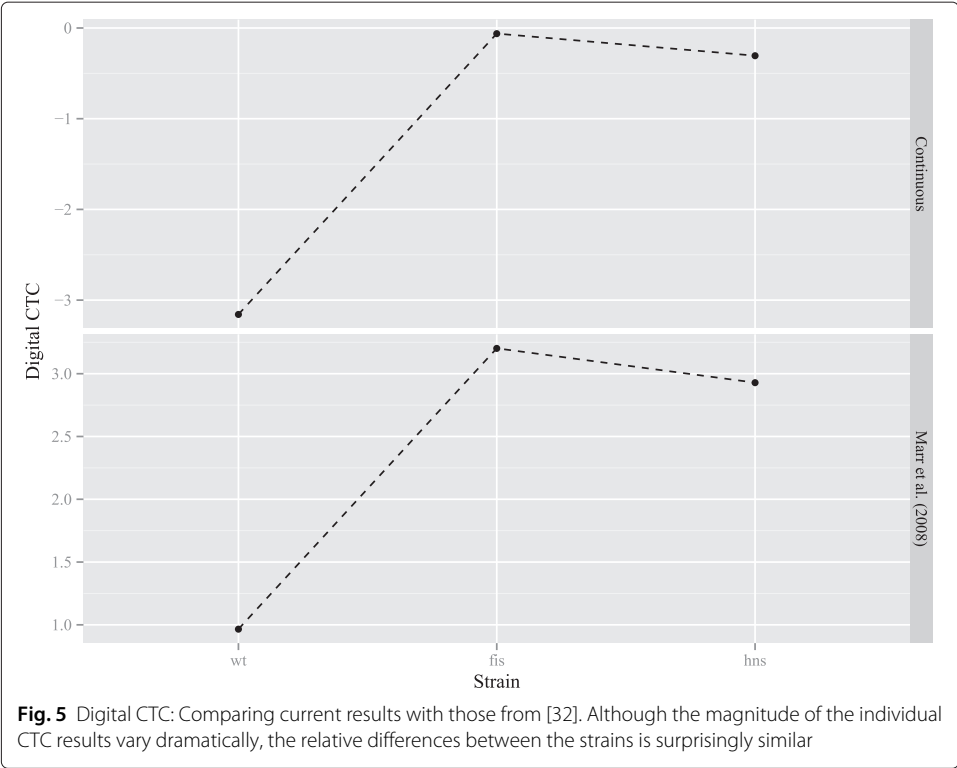
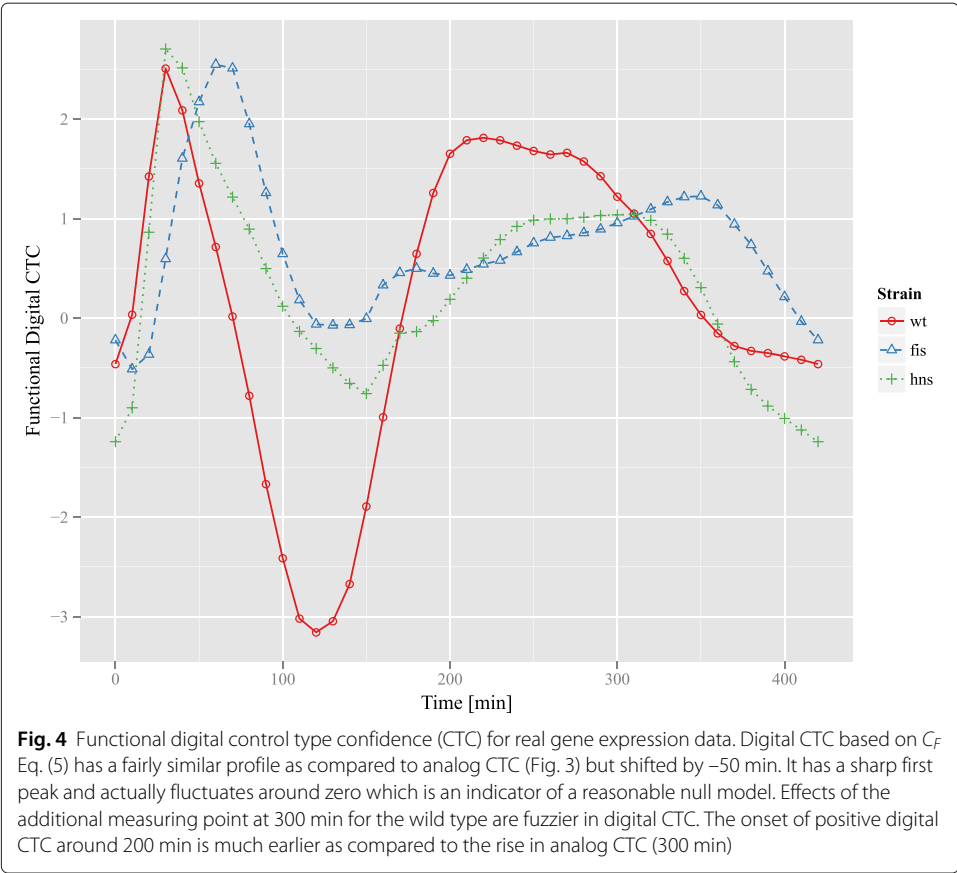


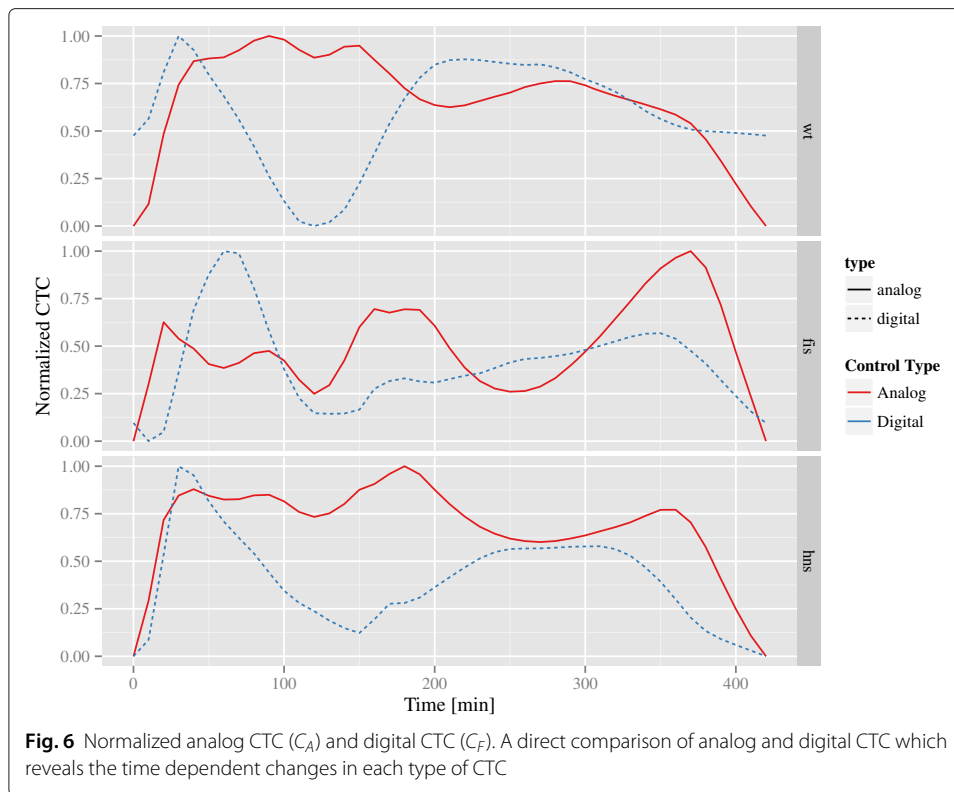
measuring point for the wild type are possibly apparent in digital CTC, too. The onset of positive digital CTC around 200 min is much earlier as compared to analog CTC (300 min) and may suggest an initiation of the global gene expression pattern change by the digital component which is followed by the analog (structural) component approximately 100 min later. The trend of digital and analog CTC is directly compared in Fig. 6 which makes it easier to follow the above description.

Conclusions

For static transcriptome data it has been shown that control strengths are a useful method for evaluating the agreement between the pattern of gene expression and the underlying transcriptional regulatory network [25, 30, 32]. Here we have extended these methods to the time domain and applied them to a novel set of gene expression profiles. The results obtained in a simple model of random Boolean dynamics have helped us to evaluate the different types of digital control strengths considered.

Conceptually, this approach is reminiscent of the comparison of structural connectivity (SC: the underlying interaction graph) and functional connectivity (FC: a network derived from similarities in dynamical behavior among nodes) in computational neuroscience





[8, 10, 40]. As in this case of SC/FC correlations, we analyze the ‘effective network’, i.e., the dynamical usage pattern of a given static interaction network.

The results for real time-series RNA-seq expression data show, on one hand, that analog CTC, over the course of the bacterial growth cycle, has two phases when it is strongest. This observation is in accordance with another evaluation of the same data [31]. The first phase, which is around the 60 min. measuring point, coincides with a peak for digital CTC. This suggests that both types of control are involved in shaping the regulatory patterns that are characteristic of the exponential growth phase.

The early peaks of analog and digital CTC (around 40 min) closely coincide with the rise of *rpoD* and *fis* expression. The *rpoD* gene encodes the RNA polymerase major σ^{70} subunit forming the vegetative RNAP σ^{70} holoenzyme, which transcribes the strong ribosomal operons organized around the chromosomal replication origin (OriC). These operons are activated by FIS during exponential growth and accumulate RNA polymerase in transcription foci formed in the vicinity of OriC [41] and delimiting the chromosomal *rrn* functional domain [3]. Accordingly, during this stage extensive communications are observed between the functionally related (primarily anabolic) genes across the chromosomal arms in the OriC end of the chromosome [30]. These observations are wholly consistent with the high analog CTC observed during the early stage of growth. At the same time *fis* acts as a hub regulating numerous genes in the TRN, and so the early activation of *fis* expression is also consistent with the early peak of the digital CTC.

The second peak of digital CTC (around 200 min) coincides with transition to stationary phase and corresponding the increase in *rpoS* and *hns* expression. The *rpoS* gene encodes the stationary phase σ^S subunit of the RNA polymerase involved in transcription

of catabolic genes [30], whereas *hns* acts as a hub in the TRN increasingly binding its genomic targets on transition to stationary phase [42]. The two (early and late) peaks of the digital CTC thus closely coincide with sequential activation of anabolic and catabolic genes, and thus mark the commencement of exponential growth and transition to stationary phase, respectively. Previous findings that the vegetative RNAP σ^{70} and the stationary phase RNAP σ^S holoenzymes cooperate with different sets of transcription factors [3] are fully consistent with this notion. Moreover, activation of RNAP σ^S holoenzyme on transition to stationary phase is associated with DNA relaxation and expression of other abundant NAPs such as IHF, Lrp and Dps followed by morphological reorganisation of the nucleoid and activation of communications between the chromosomal arms in the replication terminus [26, 39, 43], again in keeping with the second peak of analog CTC observed in our study.

The second phase of strong analog CTC is preceded by a rise in digital CTC that occurs at least an hour earlier. This may be an indication of the digital component inducing strong changes in the analog component. We should be able to identify characteristic signal on other levels of the growth cycle experiment. These results are further evidence for the compensatory interplay of digital and analog control and show interesting interactions over the growth cycle of *E. coli*.

In future studies two steps will be necessary:

- (i) Consideration of a calibration model that is based both on an artificial digital and analog component rather than just the one.
- (ii) The type of evaluation performed here should be applied to other time series data and if replicates are available compared with the results for a discrete evaluation based on differentially expressed genes.

Additional files

Additional file 1: GPN edgelist on the level of genes. An edgelist representing the undirected GPN at the level of genes using RegulonDB ECK12 identifiers. (TSV 374 kb)

Additional file 2: GPN edgelist on the level of operons. An edgelist representing the undirected GPN at the level of operons using RegulonDB ECK12 identifiers. (TSV 295 kb)

Additional file 3: TRN edgelist on the level of genes. An edgelist including regulatory interactions representing the directed multigraph TRN at the level of genes using RegulonDB ECK12 gene and TF identifiers. (TSV 118 kb)

Additional file 4: TRN edgelist on the level of operons. An edgelist including regulatory interactions representing the directed multigraph TRN at the level of operons using RegulonDB ECK12 identifiers. (TSV 53.8 kb)

Additional file 5: Additional Figures and Complementary Text. This file includes further details on the RBN model and the alternative forms of control strength discussed in the main text. (PDF 246 kb)

Acknowledgements

The work has been supported by Deutsche Forschungsgemeinschaft (DFG) grant HU-937/6 (digital and analog control) to GM and MTH. MTH also acknowledges support from BMBF (e:med program, grant 01ZX1306D).

Authors' contributions

GM and MTH conceived of the study. PS performed the experiments and generated the gene expression data. MEB and MTH performed the numerical simulations and data analyses. MEB, GM and MTH drafted the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Life Sciences and Chemistry, Jacobs University, Campus Ring 1, 28759 Bremen, Germany. ²Bioinformatics Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, 14195 Berlin, Germany. ³Department of Chromosome Biology, LOEWE-Zentrum für Synthetische Mikrobiologie, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35043 Marburg, Germany.

Received: 5 April 2016 Accepted: 5 August 2016

Published online: 22 August 2016

References

- Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JL, Covert MW. A whole-cell computational model predicts phenotype from genotype. *Cell*. 2012;150(2):389–401.
- Travers AA, Muskhelishvili G, Thompson JMT. DNA information: from digital code to analogue structure. *Phil Trans R Soc A Math Phys Eng Sci*. 2012;370(1969):2960–986.
- Berger M, Farcas A, Geertz M, Zhelyazkova P, Brix K, Travers A, Muskhelishvili G. Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Rep*. 2009;11(1):59–64.
- Geertz M, Travers A, Mehandeziska S, Sobetzko P, Chandra-Janga S, Shimamoto N, Muskhelishvili G. Structural coupling between RNA polymerase composition and DNA supercoiling in coordinating transcription: a global role for the omega subunit? *mBio*. 2011;2(4):00034–110003411.
- Dorman CJ. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat Rev Microbiol*. 2013;11(5):349–55.
- Javer A, Long Z, Nugent E, Grisi M, Siriawetwetchakul K, Dorfman KD, Cicuta P, Cosentino Lagomarsino M. Short-time movement of e. coli chromosomal loci depends on coordinate and subcellular localization. *Nat Commun*. 2013;4:3003.
- Képès F, Jester BC, Lepage T, Rafiei N, Rosu B, Junier I. The layout of a bacterial genome. *FEBS Lett*. 2012;586(15):2043–048.
- Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, Meuli R, Hagmann P. Predicting human resting-state functional connectivity from structural connectivity. *Proc Natl Acad Sci*. 2009;106(6):2035–040. doi:10.1073/pnas.0811168106.
- Deco G, McIntosh AR, Shen K, Hutchison RM, Menon RS, Everling S, Hagmann P, Jirsa VK. Identification of optimal structural connectivity using functional connectivity and neural modeling. *J Neurosci*. 2014;34(23):7910–916.
- Messé A, Hütt MT, König P, Hilgetag CC. A closer look at the apparent correlation of structural and functional connectivity in excitable neural networks. *Sci Rep*. 2015;5:7870. doi:10.1038/srep07870.
- Steuer R, Kurths J, Fiehn O, Weckwerth W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*. 2003;19(8):1019–1026.
- Müller-Linow M, Weckwerth W, Hütt MT. Consistency analysis of metabolic correlation networks. *BMC Systems Biology*. 2007;1(1):44.
- Bansal M, Belcastro V, Ambesi-Impombato A, Di Bernardo D. How to infer gene networks from expression profiles. *Mol Syst Biol*. 2007;3(1):78.
- Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G, et al. Wisdom of crowds for robust gene network inference. *Nat Methods*. 2012;9(8):796–804.
- Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10(8):538–50.
- Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*. 2004;20(18):3594–603.
- Le Novère N. Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet*. 2015;16(3):146–58.
- Strogatz SH. Exploring complex networks. *Nature*. 2001;410(6825):268–76.
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Phys Rep*. 2006;424(4):175–308.
- Arenas A, Diaz-Guilera A, Kurths J, Moreno Y, Zhou C. Synchronization in complex networks. *Phys Rep*. 2008;469(3):93–153.
- Drossel B. Random Boolean networks In: Schuster HG, editor. *Reviews of Nonlinear Dynamics and Complexity*. Weinheim: Wiley VCH; 2008. p. 69. Chap. 3.
- Bornholdt S. Less Is More in Modeling Large Genetic Networks. *Science*. 2005;310(5747):449–51. doi:10.1126/science.1119959.
- Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A*. 2004;101(14):4781–786. doi:10.1073/pnas.0305937101.
- Wang RS, Saadatpour A, Albert R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol*. 2012;9(5):055001.
- Sonnenschein N, Geertz M, Muskhelishvili G, Hütt MT. Analog regulation of metabolic demand. *BMC Syst Biol*. 2011;5(1):40. doi:10.1186/1752-0509-5-40.
- Ouafa ZA, Reverchon S, Lautier T, Muskhelishvili G, Nasser W. The nucleoid-associated proteins h-NS and FIS modulate the DNA supercoiling response of the pel genes, the major virulence factors in the plant pathogen bacterium *dickeya dadantii*. *Nucleic Acids Res*. 2012;40(10):4306–319.
- Sobetzko P, Travers A, Muskhelishvili G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci USA*. 2012;109(2):42–50.
- Beber ME, Fretter C, Jain S, Sonnenschein N, Müller-Hannemann M, Hütt M. Artefacts in statistical analyses of network motifs: general framework and application to metabolic networks. *J R Soc Interface*. 2012;9(77):3426–435.
- Sonnenschein N, Marr C, Hütt M. A topological characterization of medium-dependent essential metabolic reactions. *Metabolites*. 2012;2(4):632–47.
- Sonnenschein N, Golib Dzib JF, Lesne A, Eilebrecht S, Boukroun S, Zennaro MC, Benecke A, Hütt M. A network perspective on metabolic inconsistency. *BMC Syst Biol*. 2012;6(1):41.
- Sobetzko P, Glinkowska M, Travers A, Muskhelishvili G. DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. *Mol Biosyst*. 2013;9(7):1643.
- Marr C, Geertz M, Hütt MT, Muskhelishvili G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol*. 2008;2(1):18.
- Sonnenschein N, Hütt MT, Stoyan H, Stoyan D. Ranges of control in the transcriptional regulation of *escherichia coli*. *BMC Syst Biol*. 2009;3(1):119. doi:10.1186/1752-0509-3-119.

34. Davidich MI, Bornholdt S. Boolean Network Model Predicts Cell Cycle Sequence of Fission Yeast. *PLoS ONE*. 2008;3(2):1672. doi:10.1371/journal.pone.0001672.
35. Samaga R, Klamt S. Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun Signal (CCS)*. 2013;11(1):43. doi:10.1186/1478-811x-11-43.
36. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñoz-Rascado L, García-Sotelo JS, Weiss V, Solano-Lira H, Martínez-Flores I, Medina-Rivera A, Salgado-Orsorio G, Alquicira-Hernández S, Alquicira-Hernández K, López-Fuentes A, Porrón-Sotelo L, Huerta AM, Bonavides-Martínez C, Balderas-Martínez YI, Pannier L, Olvera M, Labastida A, Jiménez-Jacinto V, Vega-Alvarado L, del Moral-Chávez V, Hernández-Alvarez A, Morett E, Collado-Vides J. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res*. 2013;41(D1):203–13.
37. Fairweather NF, Orr E, Holland IB. Inhibition of deoxyribonucleic acid gyrase: effects on nucleic acid synthesis and cell division in *Escherichia coli* K-12. *J Bacteriol*. 1980;142(1):153–61.
38. Khodursky AB, Peter BJ, Schmid MB, DeRisi J, Botstein D, Brown PO, Cozzarelli NR. Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. *Proc Natl Acad Sci USA*. 2000;97(17):9419–24.
39. Smelkova N, Marians KJ. Timely release of both replication forks from *oriC* requires modulation of origin topology. *J Biol Chem*. 2001;276(42):39186–9191. doi:10.1074/jbc.M104411200.
40. Garcia GC, Lesne A, Hütt MT, Hilgetag CC. Building Blocks of Self-Sustained Activity in a Simple Deterministic Model of Excitable Neural Networks. *Front Comput Neurosci*. 2012;6:50. doi:10.3389/fncom.2012.00050.
41. Cabrera JE, Jin DJ. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol Microbiol*. 2003;50(5):1493–1505. doi:10.1046/j.1365-2958.2003.03805.x.
42. Kahramanoglou C, Seshasayee ASN, Prieto AI, Ibberson D, Schmidt S, Zimmermann J, Benes V, Fraser GM, Luscombe NM. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res*. 2011;39(6):2073–091. doi:10.1093/nar/gkq934.
43. Bordes P, Conter A, Morales V, Bouvier J, Kolb A, Gutierrez C. DNA supercoiling contributes to disconnect σ s accumulation from σ s-dependent transcription in *Escherichia coli*. *Mol Microbiol*. 2003;48(2):561–71. doi:10.1046/j.1365-2958.2003.03461.x.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com