IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 3, 2018, accepted September 18, 2018, date of publication September 24, 2018, date of current version October 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871876

Discovering Regulators in Post-Transcriptional Control of the Biological Clock of *Neurospora crassa* Using Variable Topology Ensemble Methods on GPUs

AHMAD AL-OMARI[1], JAMES GRIFFITH[2,3], CRISTIAN CARANICA[4], THIAB TAHA[5], HEINZ-BERND SCHÜTTLER[6], AND JONATHAN ARNOLD[3]

[1]Department of Systems Engineering and Medical Bioinformatics, Yarmouk University, Irbid 21163, Jordan
[2]College of Agricultural and Environmental Sciences, University of Georgia, Athens, GA 30602, USA
[3]Genetics Department, University of Georgia, Athens, GA 30602, USA
[4]Statistics Department, University of Georgia, Athens, GA 30602, USA
[5]Computer Science Department, University of Georgia, Athens, GA 30602, USA
[6]Department of Physics and Astronomy, University of Georgia, Athens, GA 30602, USA

Corresponding author: Ahmad Al-Omari (aomari@yu.edu.jo) and Jonathan Arnold (arnold@uga.edu)

This work was supported in part by the Joint Award MCB-SSB/PHY-POLS-1713746, in part by the NSF Molecular and Cellular BioSciences Award (Systems and Synthetic Biology), and in part by the NSF Physics of Living Systems Award.

ABSTRACT In the previous paper, we reconstructed the entire transcriptional network for all 2418 clock-associated genes in the model filamentous fungus, *Neurospora crassa* (*N. crassa*). Several authors have suggested that there is extensive post-transcriptional control in the genome-wide clock network (IEEE 3: 27, 2015). Here we have successfully reconstructed the entire clock network in *N. crassa* with a variable topology ensemble method (VTENS), assigning each clock-associated gene to the regulation of one or more of five transcription factors as well as to six RNA operons. The resulting network provides a unifying framework to explore the clock's linkage to metabolism through post-transcriptional regulation, in which ∼850 genes are predicted to fall under the regulatory control of an RNA operon. A unique feature of all of the RNA operons inferred is their functional connection to genes connected to the ribosome. We have been successful in distinguishing several hypotheses about regulatory topologies of the clock network through protein profiling of the regulators.

INDEX TERMS Variable topology ensemble methods, RNA operon, *has-1*, *lhp-1*, ribosome biogenesis, circadian rhythms.

I. INTRODUCTION

One of the major challenges of systems biology is the integration of a variety of omics and physiological data to understand complex traits, such as metabolism, development, behavior, and disease [1]. Genetic networks provide the framework to do this and also provide a framework for testing various regulatory mechanisms controlling a complex trait [2]. One of the major challenges in carrying out a program to test the regulatory mechanisms prevalent in a genome is having network reconstruction methods that: (1) recognize the sparsity of omics data scattered over the whole genome coupled with the large number of parameters (*i.e.*, rate constants and initial conditions of molecular species) [3], [4] for specifying a genetic network; (2) scale to the whole genome [5]; and (3) allow the reconstruction of networks of unknown topology (*i.e.*, who regulates whom) [6]–[8]. Microbial systems, such as *Saccharomyces cerevisiae* and *N. crassa*, have laid the foundation for understanding eukaryotic gene regulation [1], [9]–[14]. These systems continue to provide this foundation on a genomic scale [6]. Recently we developed variable topology ensemble (VTENS) methods to address all three of the challenges described above [6]. We applied these methods to the complex trait of the biological clock and reconstructed an entire transcriptional network for the clock involving 2,436 clock-associated genes for the first time [6]. This reconstruction of the clock network enabled testing of the various regulatory hypotheses about the clock's regulatory role in the genome. As an example, this network allowed

us to frame different hypotheses about how metabolism links the biological clock to aging [15].

Al-Omari *et al.* [6], and Hurley *et al.* [16] have argued that most of the regulatory control of the biological clock must be post-transcriptional because there are only five transcriptional regulators that are clock associated and 2,418 targets in the clock network [17]. On average each transcription factor is expected to have ∼40 targets [18], [19]. That means that only about 200 genes are likely to be transcriptional targets of clock-associated genes, which falls short of the 2,418 clock-associated genes by an order of magnitude. One possibility is that some of the circadian genes are not only regulated transcriptionally, but also post-transcriptionally [20], [21]. In *N. crassa*, there is indeed evidence for post-transcriptional RNA degradation rate control by the clock's *FRQ* protein [22], [23] and by the clock-related exosome [24]. In such an ''RNA operon'' mechanism [21], each RNA-binding regulator can again be expected to control no more than 40-220 circadian target genes [25], [26]. There are six RNA-binding proteins that could act as the master regulators in six RNA operons in *N. crassa [17]*. The RNA operon hypothesis combined with transcriptional control begins to narrow the gap between the observed size of the clock network and the explanation provided by the regulatory mechanisms involved.

In this paper, we formulate a regulatory network with both transcriptional control and post-transcriptional control through RNA-binding proteins that affect the mRNA stability of target genes. We then reconstruct the genome-scale genetic network based on these two regulatory mechanisms using variable topology ensemble methods. We finish by using the integration of RNA and protein profiling information on the clock network to test these hypotheses.

A. RNA OPERON MODEL WITH TRANSCRIPTIONAL REGULATION.

Each clock–associated gene, to which we will refer as a putative *clock-controlled gene* (*ccg*), can be regulated by one or more transcriptional regulators as previously described [6]. The master clock regulator, the WHITE-COLLAR COMPLEX (WCC) regulates 5 transcriptional regulators, which in turn together with WCC may regulate all putative *clock-controlled genes*. The clock mechanism itself is described previously with FRQ providing negative feedback to WCC [23], [27], [28] with many ensemble members stably oscillating. In addition, we developed a new component of the model that involves six post-transcriptional regulators that bind to mRNA (Fig 1A). Each of these post-transcriptional regulators under the control of WCC define an RNA operon. These post-transcriptional regulators are hypothesized to be RNA-binding proteins that affect combinatorially the mRNA stability of each target *ccg*. The combined effect of these six RNA-binding proteins occurs through the signal Sp(t) on each *ccg*mRNA (Fig 1B). There are then two components of the model: (1) the regulatory hierarchy (Fig 1A); (2) the model of regulation for each putative *clock-controlled gene* (Fig 1B).

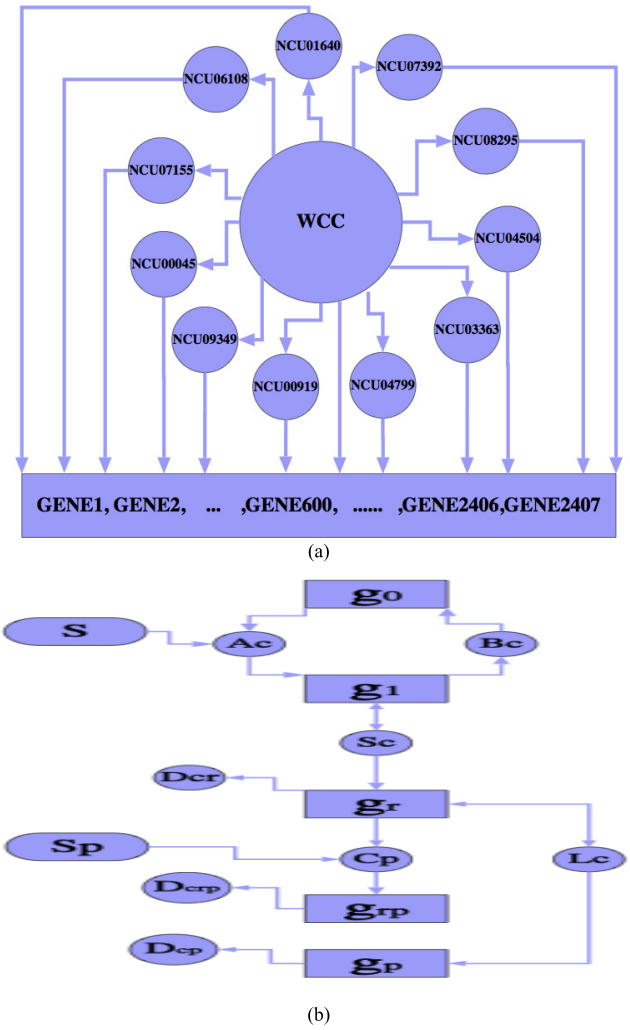


FIGURE 1. **(a) The Supernet, where each of the 2,407 genes is hypothesized to be regulated, potentially, by all of the 12 regulators (6 transcriptional regulators including one repressor and 6 RNA-binding proteins). (b) Putative *clock-controlled genes* are regulated by one or more RNA-binding proteins, which generate a combinatorial signal Sp, which determines the putative *ccg* mRNA (gr) stability. Notation follows that of the GKIN software used to generate this regulatory mechanism [29]. Molecular species (*i.e.*, reactants or products) in the network are represented by boxes, such as those labeled g_0 and g_1. These indicate transcriptionally inactive and active genes respectively; g_r and g_{rp} indicate translationally active and inactive mRNA; and g_p are proteins. Reactions in the network are represented by circles. Arrows pointing to circles identify reactants; arrows leaving circles identify products; and bi-directional arrows identify catalysts. Reactions A_c or B_c are activation and deactivation reactions, respectively. Reactions labeled with S_c, C_p, or Lc, represent transcription, post-transcriptional control, or translation reactions, respectively. Reactions without products, such as D_{crp}, D_{cr}, and D_{cp} are used to indicate decay reactions. The S Box acts as a signal that combines the effects of the five transcriptional regulators and denotes the weighted average of activator protein signals from all five regulators, and the Sp box acts as the signal that combines the effects of the six post transcriptional mRNA regulators, which bind to target mRNAs.**

The new model component (Fig 1B) describing transcriptional regulation and post-transcriptional regulation shown below is specified by a subsystem of ordinary differential equations (ODEs) for each *ccg*(see Fig 1B):

$$\frac{dg_0}{dt} = B_c g_1 - A_c g_0 S(t) + A_c g_1(t)\,\mu_4 \left[\frac{r_0}{r_4}\right]^m \left[\mathrm{Reg}_4(t)\right]^m$$

$$\frac{dg_1}{dt} = A_c g_0 S(t) - B_c g_1 - A_c g_1(t)\,\mu_4 \left[\frac{r_0}{r_4}\right]^m \left[\mathrm{Reg}_4(t)\right]^m$$

OR

$$\begin{aligned}
\left(\frac{dg_1}{dt} = -\frac{dg_0}{dt}\right)\\
\frac{dg_r}{dt} &= S_c g_1(t) - D_{cr} g_r(t) - C_p S_p(t) g_r(t)\\
\frac{dg_{rp}}{dt} &= C_p S_p(t) g_r(t - D_{crp} g_{rp}(t)\\
\frac{dg_p}{dt} &= L_c g_r(t) - D_{cp} g_p(t) \qquad (1)
\end{aligned}$$

The dynamical variables here are the concentrations of the putative *clock-controlled gene* (g_0 and g_1 in the inactive and active state, respectively), its mRNAs (g_r and g_{rp}), and its protein (g_p). See the (Fig 1B) legend for rate constants A_c, B_c, S_c, D_{cr}, C_p and D_{crp} and their descriptions. The differential equations for the protein level g_p are presented but not used here to fit the mRNA levels observed because the protein levels are usually not observed.

The supernet [6] allows for possible regulation of the putative *ccg* by any one of the six possible transcriptional *ccg*-regulators or six post transcriptional regulators, with **S(t)** acting as a signal that combines the effects of the six transcriptional regulators. The signal S(t) denotes the weighted average of activator protein signals from all six regulators. The signal **Sp(t)** acts to combine the effects of the six post-transcriptional mRNA regulators, which bind to target mRNAs. The hypothesized RNA binding proteins are: NCU08295, NCU04504, NCU03363, NCU04799, NCU00919, NCU09349. These RNA-binding proteins were identified by four separate microarray experiments including: (1) an assay for circadian rhythms in the dark; (2) a light entrainment response; (3) evidence for regulation by WCC through a *wc-1* knockout [17]. The signal for transcriptional regulation is formally defined as:

$$S(t) = \mu_0 \left[\mathrm{Reg}_0(t)\right]^m + r_0^m \sum_{k=1,k\neq 4}^{5} \left(\left[\frac{1}{r_k}\right]^m \mu_k \left[\mathrm{Reg}_k(t)\right]^m\right) \qquad (2)$$

The concentrations $\mathrm{Reg_k}(t)$ ($_k \in [1], [5]$): are the protein concentrations from 4 positive transcriptional activators (NCU07392, NCU01640, NCU06108, NCU07155) and one transcriptional repressor NCU00045), which all **bind to DNA. The protein profiles of the activators were captured using Westerns (see Materials and Methods).**

Notice that $\mathrm{Reg_0}(t)$ for WCC is treated as in the previous work [6] to generate the WCC concentrations because Western data were not available to measure the concentration [WCC] over time. Instead this trajectory was identified from an ensemble generated by a maximally informative sequence of next experiments (MINE) [17] involving four microarray experiments and initialized with published data [22], [30], [31] including Western data on WC-1.

The post-transcriptional signal affecting the *ccg* mRNA stability is defined as:

$$\mathrm{Sp}(t) = \sum_{k=6}^{11} \left(\left[\frac{1}{r_k}\right]^m \mu_k \left[\mathrm{Reg_k}(t)\right]^m\right) \qquad (3)$$

The concentrations $\mathrm{Reg_k}(t)$($_k \in [6], [11]$): are the post-transcriptional regulator concentrations. We used exactly the same approach that we did in the previous paper [6] (See Materials and Methods/Part A to generate the concentrations of these post-transcriptional regulators.

For this model, the weights (binding strengths) μ_k are constrained only by $\mu_k \geq 0$ and $\sum_{k=0,k\neq 4}^{11} \mu_k = 1$, while $\mu_4 \geq 0$ has no upper bound. The reason for no upper bound is that the 4th regulator is a repressor and acts without constraint from the other regulators. The other regulators are relative to WCC. The only constraint on μ_4 is the same one imposed on rate constants in general, such as A_c or D_{cr}. These weights give the relative importance of control by a particular regulator (with the exception of the repressor).

Regulators k=[1,11] are themselves putative *ccg* products, assumed to be regulated and controlled by WCC, *i.e.*, having a fixed $\mu_0 = 1$. For any other, non-regulatory *ccg*, the μ_k, are ensemble Monte Carlo variables, randomly varied to fit the data.

II. MATERIALS AND METHODS

A. DATA

A total of 13 cultures in triplicate were grown for a constant amount of time of 50 h. Samples varied for the amount of time in the dark. The experiments were done in such a way that the total growth time for each sample was kept constant at 50 hours so as not to confound circadian rhythms with growth effects. What varied was the amount of time in the dark. For example, the zero time point was in the dark for 0 hours and in the light, for 50 hours. The 13th time point was in the dark for 48 hours and in the light, 2 h. The average amount of time in the light for cell synchronization was 26 h before transfer to the dark. Total RNA was extracted from each of the 13 cultures grown in the dark for varying amounts of time. The total RNA of these 13 samples was probed with a microarray chip as described previously [17]. RNA profiling data on all 11,000 genes in this D/D experiment are publicly available under Accession 13 [32].

B. PROTEIN EXTRACTION

Protein extraction was carried out with kit YT-015 (Invent Biotechnologies, Inc. Eden Prairie, MN). The kit protocol was modified to the following six step protocol: (1) On dry ice, about 0.1 – 0.13 grams of previously ground up and frozen *N. crassa* mycelia were collected into a 1.7 ml microcentrifuge tube. This material was kept frozen until step #2. (2) The pellet was washed with one ml of water by centrifuging at top speed in a microcentrifuge for 3 min. The supernatant was removed completely. (3) 100 mg protein extraction powder was weighed out and added into the bottom of the tube (avoiding powder on the wall of the tube). This can

be done by weighing out the powder on a piece of folded wax paper and pouring the powder to the bottom of the tube). (4) ∼25 μl of denaturing buffer and 5 μl of 7× Protease Inhibitor Cocktail (''cOmplete Mini, EDTA free'' from Roche) were added to the bottom of the tube. The material in the tube was ground repeatedly with the pestle for about 2-3 min with twisting force. (5) 180 μl protein extraction buffer (Denaturing)/30 μl of 7× PIC used in step 2 above was added, and grinding was continued for about thirty seconds. The tube was capped and vortexed vigorously for 10 seconds. (6) The tube was centrifuged at top speed for 3-4 min. The supernatant was transferred to a fresh tube (this is the extracted total protein). If more protein was desired, steps 3-4 were repeated one more time. Extracted protein was then aliquoted and stored at −20C for future use.

C. ANTIBODY PRODUCTION

Polyclonal antibody production (PolyExpress Gold Package SC1649) was carried out by GenScript USA, Inc. (860 Centennial Ave, Pascataway, NJ 08854) for gene products from NCU00045, NCU01640 (*rpn-4*), NCU06106, NCU07155, and NCU07392 (*adv-1*), all hypothesized to be transcriptional regulators.

D. PROTEIN PROFILING OF TRANSCRIPTIONAL REGULATORS

Western blots were performed to provide protein profiles on four transcriptional regulators.

E. PROTEIN GELS

All gels were pre-cast from 4-20% ''ExpressPlus PAGE Gels'' from ''GenScript.'' A ''Qubit 2.0 Fluorometer'' instrument as detailed previously [33] was used to equalize loadings in each gel. They were run at 140 volts for about 70 minutes in Tris-MOPS (SDS) running Buffer also from ''GenScript.'' Protein Ladders used were either ''BLUEstain'' Protein Ladder from ''Gold Biotechnology, Inc. (St. Louis, MO) or ''Precision Plus Protein ''WesternC'' Standards (from Bio-Rad, Hercules, CA).

F. GEL TRANSFER

Gel transfer was done using the ''eBlot'' electroblotter from ''GenScript'' following their protocol. Each blot was transferred to the provided ''GenScript WestClear'' Nitrocellulose Membrane for 10 minutes. Following the transfer, the Western membrane was processed using the ''GenScript ONE-HOUR Western Advanced Kit.'' The protocol here was modified from theirs and is described in the next section 2.7.

G. WESTERN BLOT WAS CARRIED OUT IN SIX STEPS:

(1) The Western blot was placed in 20 mls of Blocking Reagent (10mls of Pretreat Solution A and 10mls of Pretreat Solution B). This was placed on a shaker at room temperature with moderate agitation for a minimum of 15 minutes. (2) The Western was then transferred to a fresh 20 mls of Blocking Reagent with 10 μl of a respective Knock-Out protein sample of the protein antibody of interest. The Knock-Out protein sample was taken from an ''Invent Technology *Minute Protein Extraction*'' aliquot. The membrane was put back on the shaker to block for about 1 hour. (3) The membrane was then rinsed two times with the provided 1× Wash Solution and then placed in a pre-made mixture (and incubated for 40-60 minutes) of 11-15 μl of Mixture 1 and 10mls of the provided WB-2 Solution. Mixture 1 is 10 mls of the provided WB-1 solution and about 0.1-0.5 of the Primary Antibody. This too was put on the shaker with moderate agitation, at RT for 60 minutes. (4) The membrane was then rinsed one time w/15mls of Wash Solution, followed by three washes of 15-20 minutes each in 20 mls of Wash Solution. Each of these washes was done in a fresh container. (5) During the last wash, the Chemiluminescent HRP Substrate was made up and kept in the dark as much as possible. For our Westerns we combined 800 μl of the provided Reagent A with 1,600 μl of the provided Reagent B. We vortexed to mix and stored in a closed drawer until use in the next step. (6) At the end of the last wash, the Western was taken to a dark room with a ''Safe Light'' and processed with the Chem. HRP Substrate made in Step 5. The Western was blotted semi-dry on clean ''Kimwipes'' (Kimberly-Clark) and then placed in a fresh container. The substrate was then used to slowly cover the entire opposite side of the membrane. This allowed a reaction with the Western for 5 minutes. It was the blotted semi-dry again. It was then wrapped in ''Saran Wrap'' to be exposed to chemiluminescence film (''CL-X Posure Film,'' Thermo Scientific). The amount of time exposed to film can usually be determined visually based on the strength of the chemiluminescence in total darkness. It varies from 5 seconds to 20 minutes. The bands on Westerns were quantified using IMAGEJ [34], and the HIS-3 band intensity at each time point was used to normalize the band intensity of other proteins at the same time point.

H. DETERMINING THE PROTEIN CONCENTRATIONS OF THE POST TRANSCRIPTIONAL REGULATORS

We used the numerical methods described previously [6] (See Materials and Methods/ Part A of the previous paper) to find the concentrations of $\mathrm{Reg_k(t)}$($\mathrm{k} \in$ [6], [11]); NCU08295, NCU04504, NCU03363, NCU04799, NCU00919, NCU09349. We briefly recapitulate the prior more detailed description. (1) The trajectory of WCC is derived from earlier ensemble calculations [17]. (2) For a given model proposal in Eqn (1) for each regulator concentration the protein trajectory of the regulator, $[\mathrm{Reg_k(t)}]$ (k∈[6], [11]), was solved in closed form with initial condition $\mathrm{r_k}$. (3) In this solution we matched the initial condition $\mathrm{r_k}$ to the time average of $[\mathrm{Reg_k(t)}]$ (k∈[6, 11]).

I. ENSEMBLE METHOD FOR DISCOVERING A GENOMIC-SCALE NETWORK OF UNKNOWN TOPOLOGY

The method used for this project is detailed in previous work [6] (See Materials and Methods/ Part B of this previous paper) using a larger scale model that includes here extra

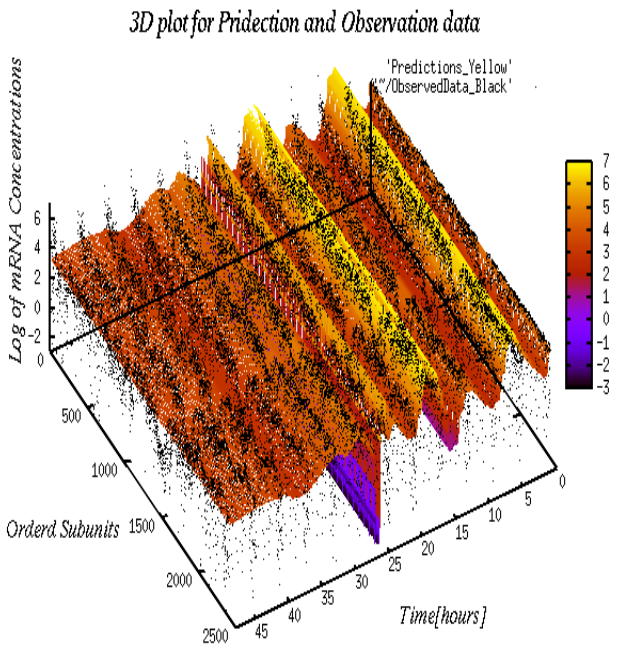


FIGURE 2. The regulatory network with transcriptional control and 6 RNA operons (in red) fits microarray RNA profiling data on 2,418 genes (in black) over a 48 hour window with time points every four hours. The 2,418 genes are arranged in order of the similarity of their RNA profiles. Data are taken from [17]. The model ensemble has 5 transcriptional regulators (including one repressor) and 6 RNA-binding proteins, all acting with a cooperativity of 4. The predictions fit the experimental data within a standard error. The observed data are computed using $\ln(y^{exp.}_{k,m})$-$< \psi_{ck} >$ vs. time (t) and the predictions, computed using $<\ln\left(F_{k,m}(t,\theta,\mu)\right)>$ vs. time. Predictions for the other models and associated plots as above resemble very closely this figure.

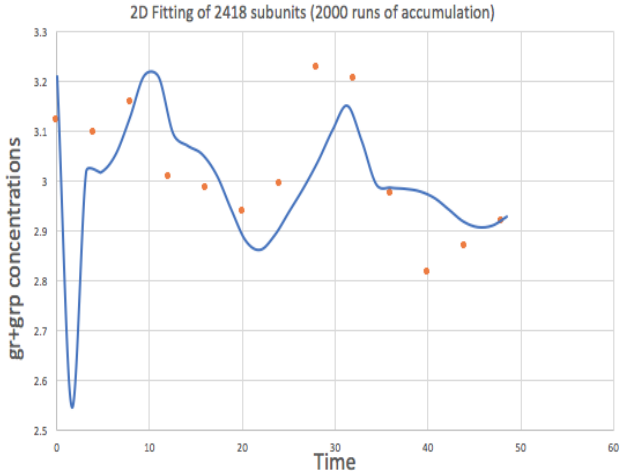


FIGURE 3. An average over 2,418 predicted mRNA profiles (in blue) accurately predicts the observed average over 2,418 mRNA profiles.

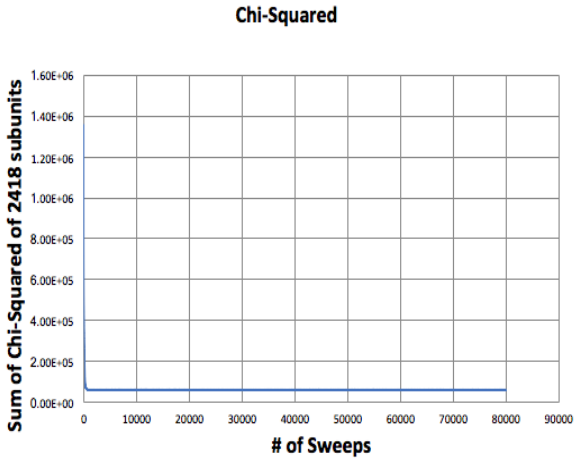


FIGURE 4. The fit (as measured by the chi-squared statistic) of the model ensemble as a function of sweep converges quite quickly under the Variable Topology Ensemble Method. A sweep is a visit on average to each parameter once in the model. There are 53,196 parameters in the model. In this model fitting the topology, the rates, and initial conditions were varied. This equilibration curve is that leading to the model ensemble shown in Fig 2.

post-transcriptional weights μ_k with a total of 22 variables for each gene and total of 53,196 variables for all of the 2,418 putative *clock-controlled genes*. A brief recapitulation of the method is now given. First, a putative *ccg* was chosen at random from among 2,418 putative *ccg*s. Then a fair coin was flipped to decide either to update the rate constants and initial conditions (*i.e.*, $\theta-$variables in Eqn. (1)) or the binding strengths (μ_k) in Eqn.s (2-3). The $\theta-$variables were updated using a Metropolis-Hasting Algorithm with adjustable step-width [6], [35]. The binding strengths were updated with a measure-preserving transformation which also preserves their sum of 1 (with the exception of the repressor). The regulators are assumed to be only regulated by WCC. The measure preserving transformation ensured that the number of parameters remained constant during the Monte Carlo run. The updates were sent to the GPU to solve the large system of ODEs. Then a Metropolis update was performed on the CPU. The Monte Carlo run was continued for ∼40,000 equilibration sweeps followed by ∼40,000 accumulation sweeps to generate the ensemble, where a sweep is a visit to each parameter in the model once on average.

J. AN ENSEMBLE METHOD USING HETEROGENEOUS ODE SOLVERS (ARK AND GQ) ON THE GPU

The system of ODEs described above can be solved using an exact integral solution formula for solving the first order linear ODEs described previously [6]. The subsystem of ordinary differential equations model has two properties in term of its solution:

1) It should be solved nearly 4 billion times during a Monte Carlo run so that we have developed a fast ODE solver using Gauss Quadrature integrator(GQ) with 100 integration points to solve for the 2,418 ODE subsystems in parallel on the GPU. The accuracy of the solution is sufficient for our biological problem [36].

2) The ODE solutions have a sudden change (stiff solution) at the early time points. Thus, we need a special solver with high accuracy and error control for this stiff solution, where the GQ solver fails for such a case, so we developed an Adaptive Runge Kutta(ARK) solver that runs on the GPU

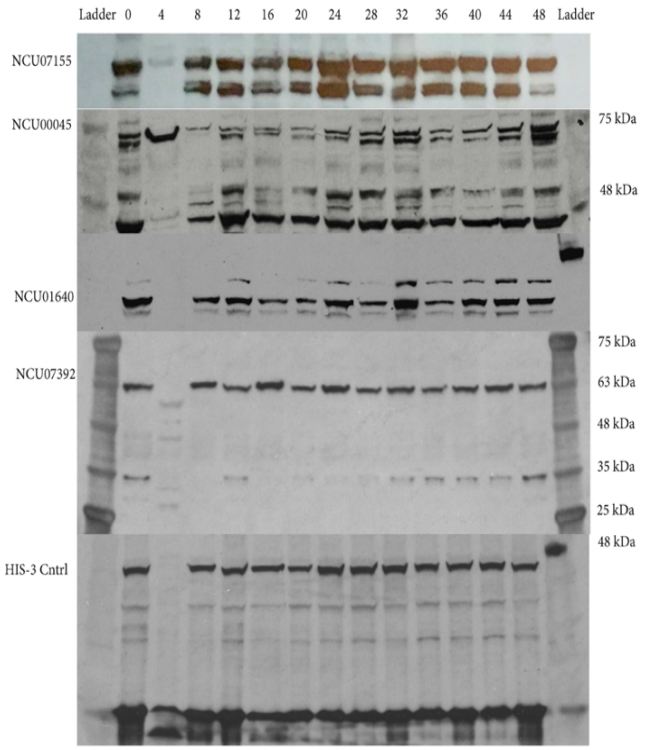


FIGURE 5. **Western Blots of 4 regulators capture temporal profiles of regulatory proteins. The level of each protein is shown in different lanes for 0, 4, . . . , 48 hours. Protein ladders are in the margins of each gel in kiloDaltons (kDa). A full Western gel is shown for the protein of NCU07392 (*adv-1*) [38]. Band intensities were quantified using IMAGEJ [34]. The HIS-3 protein from *N. crassa* is shown as a control (Cntrl).**

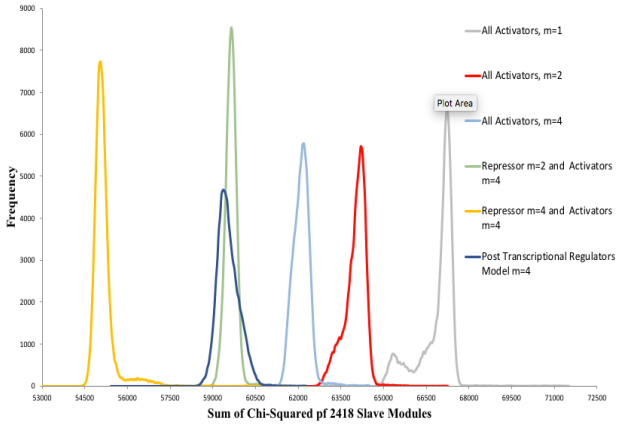


FIGURE 6. **There is strong support for a regulatory network in which there are 4 positive activators and one repressor in the transcriptional component of the network, all acting as tetramers because the histogram of chi-squared statistics of the best fitting model ensemble (in yellow) is non-overlapping with those of the remaining model ensembles and significantly different from those by Table 1. Western profiles of each transcriptional activator were utilized in the fitting.**

in a parallel fashion to find the stiff part [5] of the solution. In essence, using this ARK solver in parallel alone over the whole interval [0 48] would take years of simulation while using just GQ would not give an accurate solution due to the stiffness problem, which occurs at the beginning of the solution. As a consequence, we developed a heterogeneous solver consisting of both solvers to maintain the required speed and accuracy.

A K40 GPU (NVIDIA, Inc., Santa Clara, CA) was used for the equilibration phase, and two K40 GPUs were used for the accumulation phase with 2,418 threads on each GPU divided into a thread block of size 32 threads to solve for the system of ODEs described above, where each system had the same mathematical form and represented different putative *clock-controlled genes* with a different set of parameters. Increasing the number of thread blocks per streaming multiprocessor on the GPU and decreasing the number of threads per block (32 threads per block), provided more independent warps from other thread blocks when one warp was stalled. To achieve the best possible speed, the algorithm used registers on the GPU, which constitute the fastest memory holder, to define our variables. Nonetheless, to avoid the restriction that could happen from the excessive use of registers, all constant data, for instance, the interpolation files and all of the ARK's and GQ's constants, were defined in constant memory on the GPU. This code organization showed the best performance among other code organizations that were tested. The described ODE solver used ARK during the stiff part of the ODE solution interval and used GQ for the rest of the ODE solution to achieve the required speed and accuracy. The software +_data in this paper are deposited in sourceforge.net under at the link: https://sourceforge.net/projects/vtensarkclock1/files/PostTranscriptionalCode/?upload_just_completed=true

An interactive visualization of the network in Fig 8 is provided in a supplement S1 Fig to this publication in Cytoscape [37].

III. RESULTS AND DISCUSSION

A. COMPARISON OF THE REGULATORY MODEL TO PROFILING EXPERIMENTS

The first step in the analysis was to examine how the new regulatory model ensemble explains microarray data obtained every 4 hours over a 48 hour window in which *N. crassa* cultures were shifted to dark conditions (D/D) (Materials and Methods). Predictions were derived as a model average across an ensemble of models (Fig 1) using a Variable Topology Ensemble Method (VTENS) (See Materials and Methods). There were two components to the model, the regulatory topology (Fig 1A) and how each putative *clock-controlled gene* is regulated (Fig 1B). Both components were identified by a recently developed novel ensemble method VTENS using Markov Chain Monte Carlo [6]. The temporal profiles on mRNAs of 2,418 genes known to be circadian were compared directly with those predicted by the model ensemble in Fig 1. The network of 2,418 putative *clock-controlled genes* fits to the experimental data using the ensemble method very well and is assessed quantitatively in the next section using a chi-squared goodness of fit statistic. In the surface in Fig. 2 two peaks and two valleys can be seen corresponding to

TABLE 1. 20 replicate VTENS runs under 5 different regulatory hypotheses support the hypothesis that 4 transcriptional activators and 1 repressor is the "best" hypothesis. The presence/absence of a repressor and the Hill coefficients of regulators are varied.

Model (Replicate)	Repressor(m=)and 4 activators(m=4)	Repressor(m=2)and 4 activators(m=4)	All activators (m = 4)	All activators (m = 2)	All activators (m = 1)
1	60202	62175	63818	65770	68808
2	58391	60271	62417	64345	67302
3	55944	60469	63587	65774	69249
4	55757	59982	62033	64063	66849
5	56309	60136	62268	64175	66975
6	60202	61098	64106	66034	69053
7	56407	60297	63462	65547	68924
8	59163	62487	63771	65704	68931
9	59163	60412	63559	65526	68934
10	56154	60369	62492	64813	68149
11	55565	60347	63419	65542	68880
12	55714	60389	63494	65618	69302
13	56297	60407	62581	64471	66823
14	55493	60213	63351	65393	68804
15	56949	60243	62399	64279	67204
16	56193	60849	63176	65023	68002
17	56419	60073	62153	63919	66747
18	56208	60125	66298	68550	72113
19	56283	60111	64398	66364	69306
20	57219	61918	64812	67261	70871
Average	**56827**	**60619**	63378	65409	68561

TABLE 2. One-way analysis of variance on average χ^2_{ave} in Table 1 across 5 model ensembles of regulation of a circadian network supports the hypothesis that the model ensembles in Table 1 are significantly different in fit to the profiling data. The 5 model ensembles are listed in Table 1.

Source of Variation	Degrees of Freedom	Sum Of Squares	Mean Squares	F
Between Models	4	1570505318.54	392626329.63	274.75 ($P < .0001$)
Between replicates within Models	95	135756329.10	1429013.99	

the circadian rhythms in mRNAs over 50 hours. An interactive display of the network is available in Cytoscape as a supplement [37].

By collapsing the surface in Fig 2 down into the concentration and time dimensions, the fit of the model ensemble to RNA profiling data can be more simply examined. The observed average of the 2,418 putative *ccg* mRNA profiles are compared with those predicted by an average over all 2,418 predicted profiles in Fig 3. The resulting agreement is comparable with that reported for the transcriptional network alone (See Fig 8 in the previous paper) [6]. We conclude that an ensemble of genetic networks predicts the mRNA levels overall of 2,418 putative *clock-controlled genes* (model used here where m=4, four positive regulators, a repressor and, six post- transcriptional regulators).

B. MEASURING THE FIT OF THE MODEL ENSEMBLE TO THE RNA PROFILING DATA

An ensemble of models fitting this RNA profiling data was constructed by a particular form of Markov Chain Monte Carlo called VTENS by allowing the model parameters (Fig 1) to vary over millions of proposals. The progress in the fit of the model under this random walk is displayed and converged quite rapidly (Fig 4). There were 31,000 data points with an average (across models) of $\chi^2 = 60,597$. The average contribution to the χ^2 per data point is 1.83, which is better than the transcriptional network hypothesis alone [6]. This is surprising given that 4 of the transcriptional regulators had measured protein profiles via Westerns rather than inferred protein profiles in earlier work [6]. The present circumstances present a stronger test that the model is correct

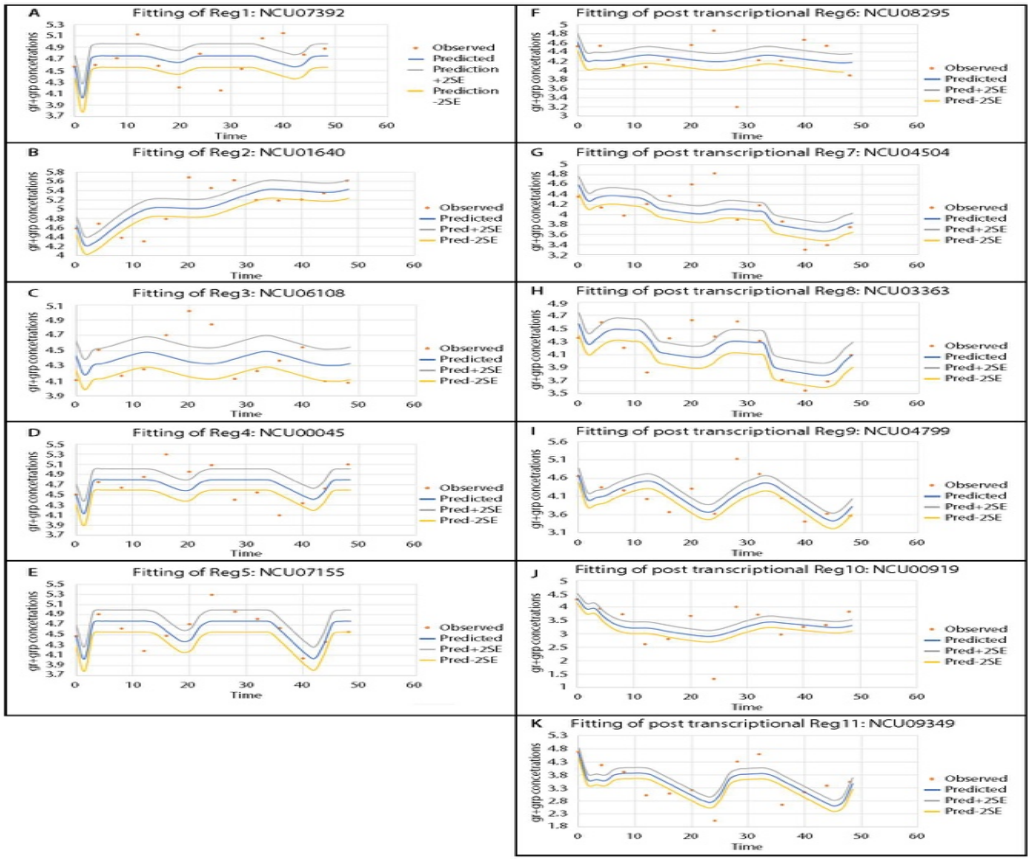


FIGURE 7. The best fitting model ensemble reasonably well predicts the combined RNA + Western profiles of all 11 regulators with some scatter. The best fitting model ensemble has one repressor and 11 activators all with Hill cooperativities of 4. The blue predicted curve represents an average over the model ensemble. The grey and yellow curves form a pointwise 95% confidence band about the average trajectory over the ensemble as computed from 2000 models in an accumulation run involving 40,000 sweeps for the best model.

by incorporating the protein profiling data into the model assessment.

C. UNCOVERING THE REGULATORY TOPOLOGY BY PROTEIN PROFILING OF THE TRANSCRIPTIONAL REGULATORS

In previous work a variety of hypotheses arose about the topology of the regulation. We speculated that one of the transcriptional regulators could be a repressor while the remaining transcriptional regulators could be activators. In addition, the Hill cooperativity of these regulators needed exploration. Varied hypotheses about the regulation were considered, in which the presence or absence of a transcriptional repressor varied and in which the Hill cooperativity of the transcriptional regulators was allowed to vary. While the earlier results suggested (Fig 11 in [6]) one favored hypothesis (4 activators and 1 repressor each with Hill Coefficient of 4), among the array of hypotheses about the regulatory network considered, no single hypothesis could be selected with high probability to provide a better fit to the profiling data. At that time there were no Western data on the transcriptional regulators available.

The Western profiles on all but one of the transcriptional regulators is shown in Fig 5. The time point at 4 h was problematic after 5 independent protein extractions. As a consequence, in addition to loading samples with equal protein concentrations (using the Qubit method as described in the Materials and Methods), we normalized the band intensity at each time point relative to the HIS-3 band intensity at the same time point in Fig 5. An ensemble of models was identified under each hypothesis shown in Fig 6, using the four Western profiles of the transcriptional regulators (See Materials and Methods). As can be seen, the hypothesis with 4 activators and 1 repressor all with Hill Coefficient of 4 is now strongly supported - the histogram of chi-squared goodness of fit statistics across the ensemble of the best fitting model (in yellow) is non-overlapping with the histograms of other models and has a lower chi-squared goodness of fit statistic. The limitation on this computational experiment is that with a different Monte Carlo run a slightly different distribution for the χ^2 can arise, a distribution that is shifted to the right (worse fit) or left (better fit). To overcome this limitation we repeated the above experiment 20 times.

To assess the significance of the differences among different model ensembles in Fig 6, each VTENS run was replicated 20 times (Table 1). As can be seen, the model with one repressor and four activators and a cooperativity of 4 (what we will call the best model) remains the best model

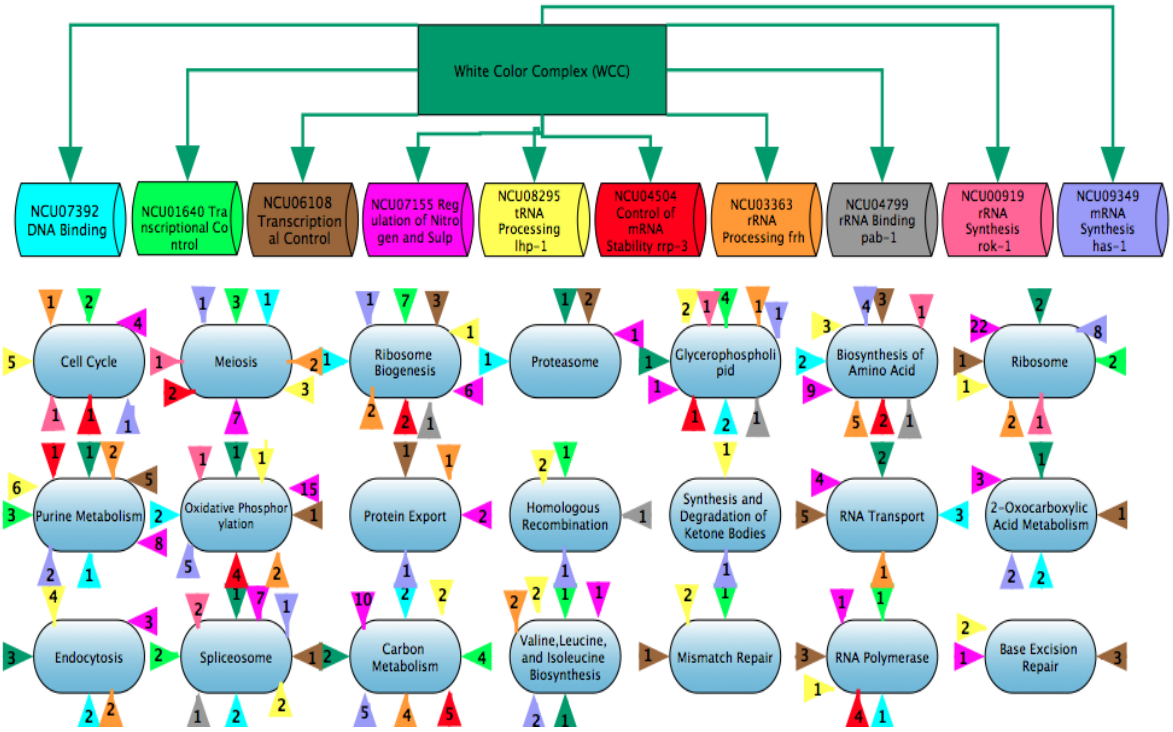


FIGURE 8. RNA operons are pervasive in their effects across functional categories in the transcriptome. A regulatory genetic network for the five transcriptional regulators (WCC, NCU07392 (*adv-1*) [38], NCU01640 (*rpn-4*) [39], NCU06108, NCU07155), the six post-transcriptional regulators (NCU08295, NCU04504, NCU0336, NCU04799, NCU00919, NCU09349) and the putative *clock-controlled genes*. The triangle is used to denote the regulator controlling a particular functional category for genes and their encoded products. The number on the triangle indicates how many annotated genes that are regulated by a particular regulator and participating in a particular pathway or function (small blue ovals).

ensemble across all replicates. We conclude the use of the Western profiles allows us to strongly discriminate between different hypotheses about the regulatory network. An Analysis of Variance on the χ^2 averages in Table 2 revealed that the hypothesis with 4 activators and 1 repressor and a Hill Coefficient of 4 was a significant improvement over the other regulatory models.

D. DOES THE "BEST MODEL" WITH 6 RNA OPERONS + TRANSCRIPTIONAL NETWORK OFFER A SIGNIFICANT IMPROVEMENT OVER THE BEST TRANSCRIPTIONAL NETWORK WITHOUT RNA OPERONS?

There are two possibilities for the structure of the clock regulatory network. Smith *et al.* [39] identified 22 transcription factors that were targets of WCC by ChIP-SEQ in a light response only experiment. It is thus possible that the clock network is a large flat transcriptional network. An alternative hypothesis is that there is substantial post-transcriptional control [6], [16]. We calculated the average and standard error of the chi-squared goodness of fit values from the histogram of chi-squared values under the hypothesis of the best transcriptional network with 6 transcriptional regulators for comparison with the histogram of chi-squared values under the hypothesis of the best post-transcriptional network with 12 regulators. Each histogram contained at least 40,000 chi-squared values for each ensemble. Under the transcriptional network hypothesis there were four activators, one repressor, and WCC. Under the post-transcriptional hypothesis there were 12 regulators. The mean chi-squared under the former was 56,192 +/− 109, and under the latter, 60,597 +/− 76. A KS test of the two distributions of chi-squared values was significantly different at $P < 0.0001$, but this outcome is likely to change as Western profiles of regulators are added, such as CSP-1 [40]. The maximum difference between the cumulative distributions in the KS-test was 0.1257875 [41].

A final test of goodness of fit was made by comparing the observed profiles of the regulators versus the predictions of the best fitting model ensemble (Fig 7). As can be seen the combined RNA profiles and Western profiles was reasonably well predicted with some scatter by the best model ensemble with one repressor and 11 activators and a Hill cooperativity of 4. The oscillations in the regulators under the best ensemble are also apparent.

E. PREDICTING THE TARGET PROFILES AND TARGET GENES OF RNA-BINDING PROTEINS.

In Table 3 the binding strengths of each RNA-binding protein were used to characterize their targets in the transcriptome. All of the RNA-binding proteins examined (Table 3)

TABLE 3. Regulator binding strength and target gene functions. The strength of regulator binding is computed by asking: what is the average of μ's across the 40,000 accumulation sweeps that is assigned to a group of genes that have the same function or pathway?

Function\Regulators	WCC	NCU07392	NCU01640	NCU06108	NCU07155	NCU08295	NCU04504	NCU03363	NCU04799	NCU00919	NCU09349
Ribosome	0.4914	0.4582	0.0000	0.4657	0.4772	0.4340	0.0000	0.4675	0.0000	0.4832	0.5584
Ribosome biogenesis in eukaryotes	0.0000	0.4541	0.4630	0.5493	0.5205	0.4031	0.4258	0.4564	0.4608	0.0000	0.4962
Meiosis - yeast	0.0000	0.4487	0.4785	0.0000	0.4596	0.4204	0.5392	0.4679	0.0000	0.4470	0.4730
Proteasome	0.4989	0.0000	0.4132	0.5052	0.4807	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Carbon metabolism	0.7028	0.4923	0.3844	0.0000	0.4978	0.4166	0.5378	0.4612	0.0000	0.0000	0.5388
Valine, leucine and isoleucine biosynthes	0.4197	0.4649	0.0000	0.0000	0.5215	0.3883	0.0000	0.4441	0.0000	0.0000	0.5256
Base excision repair	0.0000	0.0000	0.0000	0.5130	0.4802	0.3781	0.0000	0.0000	0.0000	0.0000	0.0000
Cell cycle - yeast	0.0000	0.4594	0.0000	0.0000	0.4228	0.4304	0.4711	0.4371	0.0000	0.4470	0.4730
Glycerophospholipid metabolism	0.4874	0.4665	0.5435	0.0000	0.3931	0.3864	0.5169	0.4876	0.4691	0.4505	0.4696
Synthesis and degradation of ketone bod	0.0000	0.0000	0.0000	0.0000	0.0000	0.3738	0.0000	0.0000	0.0000	0.0000	0.3681
Biosynthesis of amino acids	0.0000	0.0000	0.3844	0.4609	0.5032	0.4211	0.5061	0.4585	0.5157	0.4701	0.5452
Purine metabolism	0.4973	0.4837	0.3670	0.7374	0.4764	0.4204	0.5257	0.4215	0.0000	0.0000	0.4641
Oxidative phosphorylation	0.4914	0.0000	0.4529	0.4910	0.4839	0.4220	0.4703	0.4654	0.0000	0.4476	0.6401
Protein export	0.0000	0.0000	0.0000	0.5326	0.4471	0.0000	0.0000	0.4714	0.0000	0.0000	0.5047
Homologous recombination	0.0000	0.5743	0.0000	0.0000	0.0000	0.3698	0.0000	0.0000	0.3673	0.0000	0.5097
RNA transport	0.7477	0.0000	0.4735	0.5753	0.5070	0.0000	0.4507	0.4775	0.0000	0.0000	0.0000
2-Oxocarboxylic acid metabolism	0.4914	0.0000	0.4529	0.4661	0.5782	0.0000	0.0000	0.0000	0.0000	0.0000	0.4646
Endocytosis	0.4211	0.0000	0.4333	0.0000	0.4657	0.3836	0.0000	0.3965	0.0000	0.0000	0.0000
Spliceosome	0.5024	0.4632	0.3828	0.3843	0.5428	0.4161	0.0000	0.0000	0.3760	0.0000	0.4555
RNA polymerase	0.0000	0.5224	0.3670	0.6668	0.4669	0.4639	0.0000	0.0000	0.0000	0.0000	0.0000
Mismatch repair	0.0000	0.5170	0.0000	0.4281	0.0000	0.4068	0.0000	0.0000	0.0000	0.0000	0.0000

targeted genes connected with the ribosome or ribosome biogenesis, amino acid biosynthesis, and glycerophospholipid metabolism. Five of the six RNA-binding proteins targeted meiosis and oxidative phosphorylation. The fact that *lhp-1* (NCU8295) and *has-1* (NCU9349) encode proteins that target ribosome related and ribosome biogenesis functions comes as no surprise [25], [42]. The Has1p protein in *S. cerevisiae* is the only DEAD box RNA helicase required for 40S ribosome subunit biogenesis. It also plays a role in 60S ribosome subunit biogenesis [43]. In *S. cerevisiae* Lhp-1p targets specific mRNA populations as well as components of rRNA processing, such as the U3 small nucleolar RNA (snoRNA) associating with the 90S pre-ribosome [25]. These two RNA-binding proteins are also distinct in that they have unique putative targets. The *N. crassa* protein LHP-1 is predicted to target uniquely RNA polymerase functions and mismatch repair, and the protein HAS-1 is predicted to target uniquely homologous recombination and 2-oxycarbolic acid metabolism. As shown in Fig 8, the six inferred RNA operons are pervasive in their functional effects. Only two functional categories in Fig 8 escape the influence of RNA operons, the proteasome functions and endocytosis. A full interactive version of Figure 8 is available in Cytoscape as a supplementary S1 Fig to this publication [37]. A zoomed-out view of the regulatory network is provided by Cytoscape in Fig 9.

What is also interesting is that the total number of targets predicted in the RNA operons is 850 mRNAs which contrasts to the 1568 genes under transcriptional control. This means each RNA operon in *N crassa* has on average 142 target genes. The range reported for the size of RNA operons in *S. cerevisiae* is 40-220 genes [25], [26]. The introduction of the RNA operon hypothesis substantially closes the gap between the 2,436 putative *ccg*s and those that can be reasonably considered as part of a transcriptional network.

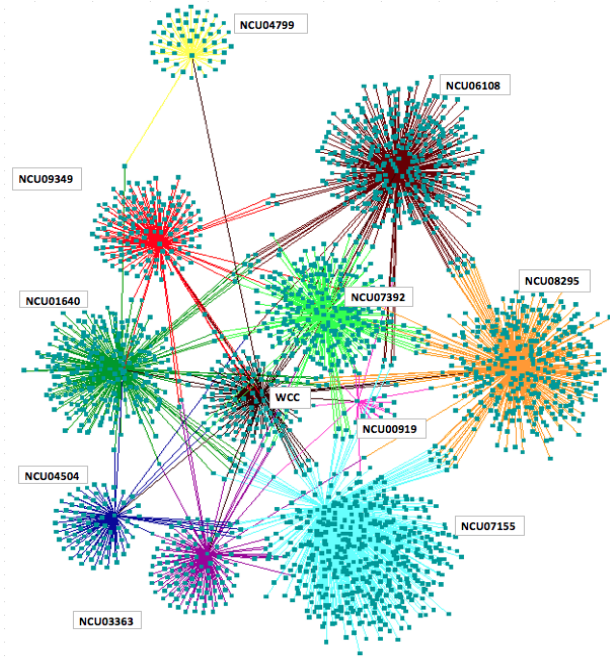


FIGURE 9. The whole regulatory clock network of 2,418 genes was reconstructed down to the single gene level in Cytoscape. The different colored edges emanate from different transcriptional and post-transcriptional regulators. The different clusters are labeled by their regulator.

IV. CONCLUSION

With the introduction of RNA operons into the transcriptional network for the clock we have substantially reduced the gap between the known size of the clock network and that part explainable by transcriptional control. Over 800 genes are predicted to be part of 6 RNA operons in the clock network. Two of these identified RNA operons have functional characteristics similar to those in *S. cerevisiae*. The relevant putative RNA-binding proteins are encoded by *has-1* and *lhp-1*, both

of which are involved in ribosome biogenesis [25], [42]. This mechanism of regulation may also be operational in the clocks of other model systems. There is evidence of RNA-binding proteins in mouse, such as Brf1 and Ccr4 [44], and in Arabidopsis, such as *dst1 [45]*, that are circadian in expression. Some clock-output pathways have been identified as targets of a RNA-binding protein LARK in the fly [46]. More recently, RNA-binding proteins, Cirbp and Rbm3, have been found to be cold-induced and targeted clock-associated genes [27]. The RNA operon hypothesis may be a general mechanism by which the clock extends its influence in the transcriptome [21].

In our previous work we established for the first time that the clock controlled ribosome biogenesis in *N. crassa* [17]. This result was reconfirmed five years later in another model system, the mouse [47]. What is striking about the RNA operons identified by variable topology ensemble methods (VTENS) is that all six RNA operons have a connection to the ribosome or its biogenesis. In particular, the *has-1* and *lhp-1* target genes are connected with ribosome biogenesis (Fig 8). This raises the possibility of another post-transcriptional mechanism for gene regulation operating in the clock network. The master clock regulator WCC has the ability to modulate the translation rate of *any* protein through its control of ribosome biogenesis. In particular, there are over 100 transcription factors that could be affected [48]. This additional mechanism could easily explain how 2,436 genes are part of the clock network, and this additional mechanism can be addressed by modifying Fig 1B to account for this mechanism and fitted by the VTENS methods introduced here. Other mechanisms are also being considered, such as transcription factors that are targets of those transcription factors or RNA binding proteins in Fig 8. These other mechanisms can be tested with a similar framework to Fig 1 and be evaluated by ensemble methods.

The construction of this larger regulatory network with post-transcriptional control involved the development of several new approaches to genetic network reconstruction. The key idea to reconstruction of the regulatory network is the supernet in Fig 1A. Each regulator is considered to be a candidate regulator of each target gene. Each regulator has a binding strength to each target. Ensemble methods are then used to identify the binding strength of each regulator to a target. The supernet is the key to the implementation of a variable topology ensemble method of network identification (VTENS) and resolving the character of regulation in the clock network.

A second innovation needed for implementing VTENS on a regulatory network with post-transcriptional control was the use of a heterogeneous solver for specifying the dynamics of the genetic network in Fig 1B. While the fast Gaussian Quadrature (GQ) solver was sufficient in previous work for a transcriptional network [6], GQ was not sufficiently accurate here for VTENS using the RNA operons in Fig 1B. The GQ solver had troubles near the zero time point, which can be overcome by the use of the ARK solver near time 0 [5].

The VTENS method is broadly applicable to microarray or RNA-seq data or a mixture. During the ensemble fitting the observed trajectories are assigned to different scale factor classes to render profiling data from different methodologies onto a common scale provided by the model. The scales are automatically identified during the fitting process. Each RNA or protein profile can be placed into its own scale-factor class for data integration as in earlier ensemble methods [49]. The VTENS methodology is scalable to genomes, such as the human genome, involving tens of thousands of genes in a genetic network(see Fig 4 in [6]). When combined with new ensemble tools of model-guided discovery, such as MINE [50], these tools will provide for the simultaneous choice of experiments and identification of genome-wide wide regulatory mechanisms.

The use of VTENS allowed us to profile the function of the RNA-binding proteins targeted (Table 3). There are now strong predictions of what the regulators are regulating (Fig 8, Fig 9, or interactive graphic supplement S1 Fig). Some of the predictions are consistent with the known targets of homologous RNA-binding proteins in *S. cerevisiae*. The RNA binding proteins are likely to be pervasive in their effects in the clock network (Fig 8). We now have a path forward for reconstructing on a genomic scale the regulatory mechanisms at work in the clock network in particular and regulatory networks for other complex traits in general. The tools here will provide a new vista on the genomics of fungi [51].

ACKNOWLEDGEMENTS

The authors would like to thank the careful editing by Professor Sidney Kushner.

S1. Fig. This interactive graphic network.cys is provided for loading into Cytoscape to render the interactive graphic of the clock network.

REFERENCES

[1] T. Ideker *et al.*, "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network," *Science*, vol. 292, no. 5518, pp. 929–934, 2001.
[2] E. O. Voit, "The best models of metabolism," *Wiley, Interdiscipl. Rev., Syst. Biol. Med.*, vol. 9, no. 6, p. e1391, Nov. 2017.
[3] D. Battogtokh, D. K. Asch, M. E. Case, J. Arnold, and H.-B. Schuttler, "An ensemble method for identifying regulatory circuits with special reference to the *qa* gene cluster of *Neurospora crassa*," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 26, pp. 16904–16909, Dec. 2002.
[4] E. Zamora-Sillero, M. Hafner, A. Ibig, J. Stelling, and A. Wagner, "Efficient characterization of high-dimensional parameter spaces for systems biology," *BMC Syst. Biol.*, vol. 5, no. 1, p. 142, Sep. 2011.
[5] A. AL-Omari, J. Arnold, T. Taha, and H.-B. Schuttler, "Solving large nonlinear systems of first-order ordinary differential equations with hierarchical structure using multi-GPGPUs and an adaptive Runge Kutta ODE solver," *IEEE Access*, vol. 1, pp. 770–777, 2013.
[6] A. Al-Omari, J. Griffith, M. Judge, T. Taha, J. Arnold, and H. Schüttler, "Discovering regulatory network topologies using ensemble methods on GPGPUs with special reference to the biological clock of *Neurospora crassa*," *IEEE Access*, vol. 3, pp. 27–42, 2015.
[7] C. Oguz, L. T. Watson, W. T. Baumann, and J. J. Tyson, "Predicting network modules of cell cycle regulators using relative protein abundance statistics," *BMC Syst. Biol.*, vol. 11, p. 30, Feb. 2017.
[8] M. Sunnåker *et al.*, "Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key Msn2 control mechanism," *Sci. Signaling*, vol. 6, no. 277, p. ra41, 2013, doi: 10.1126/scisignal.2003621.

[9] M. Schweizer, M. E. Case, C. C. Dykstra, N. H. Giles, and S. R. Kushner, "Identification and characterization of recombinant plasmids carrying the complete *qa* gene cluster from *Neurospora crassa* including the *qa*-1$^+$ regulatory gene," *Proc. Nat. Acad. Sci. USA*, vol. 78, no. 8, pp. 5086–5090, 1981.

[10] M. Johnston, "A model fungal gene regulatory mechanism: The GAL genes of Saccharomyces cerevisiae," *Microbiol. Rev.*, vol. 51, no. 4, p. 458, 1987.

[11] N. H. Giles *et al.*, "Gene organization and regulation in the *qa* (quinic acid) gene cluster of *Neurospora crassa*," *Microbiol. Rev.*, vol. 49, no. 3, p. 338, 1985.

[12] X. Tang *et al.*, "Systems biology of the *qa* gene cluster in *Neurospora crassa*," *PLoS ONE*, vol. 6, no. 6, p. e20671, Jun. 2011.

[13] J. A. Hautala, C. L. Bassett, N. H. Giles, and S. R. Kushner, "Increased expression of a eukaryotic gene in Escherichia coli through stabilization of its messenger RNA," *Proc. Nat. Acad. Sci. USA*, vol. 76, no. 11, pp. 5774–5778, 1979.

[14] N. K. Alton, J. A. Hautala, N. H. Giles, S. R. Kushner, and D. Vapnek, "Transcription and translation in *E. coli* of hybrid plasmids containing the catabolic dehydroquinase gene from *Neurospora crassa*," *Gene*, vol. 4, no. 3, pp. 241–259, Nov. 1978.

[15] M. Judge, J. Griffith, and J. Arnold, "Aging and the biological clock," in *Circadian Rhythms and Their Impact on Aging*. Berlin, Germany: Springer, 2017, pp. 211–234.

[16] J. M. Hurley *et al.*, "Analysis of clock-regulated genes in *Neurospora* reveals widespread posttranscriptional control of metabolic potential," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 48, pp. 16995–17002, 2014.

[17] W. Dong *et al.*, "Systems biology of the clock in *Neurospora crassa*," *PLoS One*, vol. 3, no. 8, p. e3105, 2008.

[18] C. T. Harbison *et al.*, "Transcriptional regulatory code of a eukaryotic genome," *Nature*, vol. 431, no. 7004, pp. 99–104, Sep. 2004.

[19] T. I. Lee *et al.*, "Transcriptional regulatory networks in *Saccharomyces cerevisiae*," *Science*, vol. 298, no. 5594, pp. 799–804, Oct. 2002.

[20] J. D. Keene and S. A. Tenenbaum, "Eukaryotic mRNPs may represent posttranscriptional operons," *Mol. Cell*, vol. 9, no. 6, pp. 1161–1167, 2002.

[21] J. D. Keene, "Biological clocks and the coordination theory of RNA operons and regulons," in *Proc. Cold Spring Harbor Symp. Quant. Biol.*, vol. 72, 2007, pp. 157–165.

[22] K. Lee, J. J. Loros, and J. C. Dunlap, "Interconnected feedback loops in the *Neurospora* circadian system," *Science*, vol. 289, no. 5476, pp. 10–107, Jul. 2000.

[23] Y. Yu *et al.*, "A genetic network for the clock of *Neurospora crassa*," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 8, pp. 2809–2814, 2007.

[24] J. Guo, P. Cheng, H. Yuan, and Y. Liu, "The exosome regulates circadian gene expression in a posttranscriptional negative feedback loop," *Cell*, vol. 138, no. 6, pp. 1236–1246, Sep. 2009.

[25] M. Inada and C. Guthrie, "Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 2, pp. 434–439, Jan. 2004.

[26] A. P. Gerber, D. Herschlag, and P. O. Brown, "Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast," *PLoS Biol*, vol. 2, no. 3, p. E79, Mar. 2004.

[27] Y. Liu *et al.*, "Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation," *Sci. Rep.*, vol. 3, Jun. 2013, Art. no. 2054.

[28] C. I. Hong, P. Ruoff, J. J. Loros, and J. C. Dunlap, "Closing the circadian negative feedback loop: FRQ-dependent clearance of WC-1 from the nucleus," *Genes Develop.*, vol. 22, no. 22, pp. 3196–3204, 2008.

[29] J. Arnold, T. R. Taha, and L. Deligiannidis, "GKIN: A tool for drawing genetic networks," *Netw. Biol.*, vol. 2, no. 1, p. 26, 2012.

[30] M. Görl, M. Merrow, B. Huttner, J. Johnson, T. Roenneberg, and M. Brunner, "A PEST-like element in FREQUENCY determines the length of the circadian period in *Neurospora crassa*," *EMBO J.*, vol. 20, no. 24, pp. 7074–7084, Dec. 2001.

[31] N. Y. Garceau, Y. Liu, J. J. Loros, and J. C. Dunlap, "Alternative initiation of translation and time-specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY," *Cell*, vol. 89, no. 3, pp. 469–476, May 1997.

[32] Z. Zhang and J. P. Townsend, "The filamentous fungal gene expression database (FFGED)," *Fungal Genet. Biol.*, vol. 47, no. 3, pp. 199–204, Mar. 2010.

[33] C. Caranica *et al.*, "Ensemble methods for stochastic networks with special reference to the biological clock of *Neurospora crassa*," *PLOS ONE*, vol. 13, no. 5, p. e0196435, 2018.

[34] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, no. 7, pp. 671–675, 2012.

[35] J. Dai *et al.*, "Exact sample size needed to detect dependence in $2 \times 2 \times 2$ tables," *Biometrics*, vol. 63, no. 4, pp. 1245–1252, 2007.

[36] G. Huang, "Parmaeter estimation of chemical reaction networks: The super-ensemble approach," M.S. thesis, Dept. Phys. Astron., Univ. Georgia, Athens, GA, USA, 2007.

[37] P. Shannon *et al.*, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, 2003, doi: 10.1101/gr.1239303.

[38] R. Dekhang *et al.*, "The *Neurospora* transcription factor ADV-1 transduces light signals and temporal information to control rhythmic expression of genes involved in cell-fusion," *G3, Genes|Genomes|Genet.*, vol. 7, no. 1, p. 129, 2017, doi: 10.1534/g3.116.034298.

[39] K. M. Smith *et al.*, "Transcription factors in light and circadian clock signaling networks revealed by genomewide mapping of direct targets for *Neurospora* white collar complex," *Eukaryotic Cell*, vol. 9, no. 10, pp. 1549–1556, 2010.

[40] G. Sancar, C. Sancar, and M. Brunner, "Metabolic compensation of the Neurospora clock by a glucose-dependent feedback of the circadian repressor CSP1 on the core oscillator," *Genes Develop.*, vol. 26, no. 21, pp. 2435–2442, 2012.

[41] M. Kendall and A. Stuart, *The Advanced Theory of Statistics: Inference and Relationship*, vol. 2. New York, NY, USA: Macmillan, 1979, p. 530.

[42] X.-H. Liang and M. J. Fournier, "The helicase Has1p is required for snoRNA release from pre-rRNA," *Mol. Cell Biol.*, vol. 26, no. 20, pp. 7437–7450, Oct. 2006.

[43] B. Emery, J. De La Cruz, S. Rocak, O. Deloche, and P. Linder, "Has1p, a member of the DEAD-box family, is required for 40S ribosomal subunit biogenesis in *Saccharomyces cerevisiae*," *Mol. Microbiol.*, vol. 52, no. 1, pp. 141–158, Apr. 2004.

[44] S. Panda *et al.*, "Coordinated transcription of key pathways in the mouse by the circadian clock," *Cell*, vol. 109, no. 3, pp. 307–320, May 2002.

[45] P. Lidder, R. A. Gutierrez, P. A. Salome, C. R. McClung, and P. J. Green, "Circadian control of messenger RNA stability. Association with a sequence-specific messenger RNA decay pathway," *Plant Physiol.*, vol. 138, no. 4, pp. 2374–2385, Aug. 2005.

[46] Y. Huang, G. Genova, M. Roberts, and F. R. Jackson, "The LARK RNA-binding protein selectively regulates the circadian eclosion rhythm by controlling E74 protein expression," *PLOS ONE*, vol. 2, no. 10, p. e1107, 2007.

[47] C. Jouffe *et al.*, "The circadian clock coordinates ribosome biogenesis," *PLoS Biol.*, vol. 11, no. 1, p. e1001455, 2013.

[48] H. V. Colot *et al.*, "A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 27, pp. 10352–10357, Oct. 2006.

[49] Y. Yu *et al.*, "A genetic network for the clock of *Neurospora crassa*," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 8, pp. 2809–2814, Feb. 2007.

[50] R. L. McGee and G. T. Buzzard, "Maximally informative next experiments for nonlinear models," *Math. Biosci.*, vol. 302, pp. 1–8, Aug. 2018.

[51] J. W. Bennett and J. Arnold, "Genomics for fungi," in *Biology of the Fungal Cell*, R. J. Howard and N. A. R. Gow, Eds. Berlin, Germany: Springer, 2001, pp. 267–297.

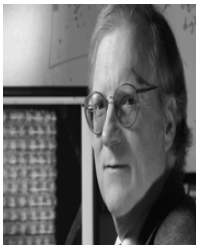
AHMAD AL-OMARI received the B.Sc. degree in electrical and computer engineering from Yarmouk University, Irbid, Jordan, in 2004, and the Ph.D. degree in bioinformatics from the University of Georgia, Athens, USA, in 2015. In 2015, he joined the Department of Biomedical Systems and Bioinformatics Engineering, College of Engineering, Yarmouk university, where he currently serves as a Professor of bioinformatics. He is considered as the first Bioinformatician in Jordan. He has many published papers and posters in bioinformatics, numerical methods, and parallel algorithms on the GPGPUs. He received grants from different institutes to pursue his research. His research interests include parallel computation, systems biology, numerical analysis, biological circuits, and gene networks.

JAMES GRIFFITH received the B.S. degree in animal and dairy science from the University of Georgia in 1984. Following several years in areas of bovine genetics, he joined the Genetics Department, University of Georgia, in 1989, as a Research Technician, where he is currently the Lab Manager and a Research Professional for Dr. J. Arnold. He has published over 20 publications in fungal genomics and systems biology. He was heavily involved in the physical mapping of several fungi (*Aspergillus nidulans*, *Aspergillus flavus*, and *Neurospora crassa*), and for the last seven years with longevity and the biological clock of *Neurospora crassa*.

CRISTIAN CARANICA received the Ph.D. degree in mathematics from Louisiana State University in 2009. He is currently pursuing the Ph.D. degree in statistics under the supervision of Prof. J. Arnold and H.-B. Schuttler with the University of Georgia, Atlanta, GA, USA, with a focus on systems biology and stochastic genetic networks. He is currently a Research Assistant with the University of Georgia under NSF support developing ensemble methods for stochastic networks. His current research interests include parallel computation, systems biology, stochastic biological circuits and gene networks, and inference about genetic networks with ensemble methods.

THIAB TAHA received the Ph.D. degree from Clarkson University in 1982. He was a Fulbright Scholar from 1995 to 1996. In 1982, he joined the Computer Science Department, University of Georgia (UGA), where he is currently the Department Head. He is the currently the Director of the CUDA Teaching and Research Centers, UGA. He is also the Vice President of IMACS. He has authored over 60 research papers. His research interests include scientific and distributed computing, and software development for solving problems in nonlinear waves, optical fiber communication systems, and biochemical reaction networks and related topics. He received the M. G. Michael Award for research in sciences from UGA in 1985. He is the Co-Editor and the Chief of the *Applied Numerical Mathematics* Journal. He is also a Senior Editor of the *Mathematics and Computers in Simulation* Journal.

HEINZ-BERND SCHÜTTLER received the Diploma degree in physics from the Technische Universität München, Germany, in 1981, and the Ph.D. degree in physics from the University of California at Los Angeles, Los Angeles, in 1984. He held post-doctoral positions at the University of California at Santa Barbara, Santa Barbara, and the Argonne National Laboratory. He joined the Faculty of the Department of Physics and Astronomy, University of Georgia, in 1987, where he is currently serves as a Professor of physics. His research interests include applications of computational statistical mechanics in biology and condensed matter physics. His recent notable contributions in computational biology include the development of the ensemble network simulation (ENS) method for the reconstruction of biological circuits and ENS-based methods for maximally informative next experiment design.



JONATHAN ARNOLD received the B.S. degree in mathematics and the M.Phil. and Ph.D. degrees in statistics from Yale University in 1975, 1978, and 1982, respectively. In 1982, he joined the Departments of Statistics and Genetics, College of Arts and Sciences, University of Georgia, where he was promoted to a Professor of Genetics, Statistics, and Physics and Astronomy. He has authored or co-authored over 140 papers in journals, book chapters, and conference proceedings in the subjects of statistical genetics, population genetics, computational biology, fungal genomics, and systems biology. He has served on the NSF Computational Biology Panel, the DOE Genome Panel, the NSF Systems and Synthetic Biology Panel. He was a Charter Member of the NIH Genetic Variation and Evolution Study Section. His research interests include the development and identification of genetic networks of fundamental processes in the model system *Neurospora crassa*, Computing Life. He was elected as an AAAS Fellow in 2011.