Biased Eukaryotic Gene Regulation Rules Suggest Genome Behavior Is Near Edge of Chaos

Stephen E. Harris
Bruce Kean Sawhill
Andrew Wuensche
Stuart A. Kauffman

SFI WORKING PAPER: 1997-05-039

SANTA FE INSTITUTE

BIASED EUKARYOTIC GENE REGULATION RULES SUGGEST GENOME BEHAVIOR IS NEAR EDGE OF CHAOS

Stephen E. Harris†, Bruce K. Sawhill*, Andrew Wuensche*, Stuart Kauffman*‡

***Santa Fe Institute**
1399 Hyde Park Rd.
Santa Fe, N.M. 87501

†Dept. of Medicine
Division of Endocrinology
The Univ. of Texas Health Science Center at San Antonio
7703 Floyd Curl Drive
San Antonio. Texas 78284–7877

‡To whom all correspondence should be directed.

Header: Control rules governing activity of eukaryotic genes appear strongly biased towards large numbers of "canalizing" inputs. The ensemble of networks with the observed bias predicts cells are in an ordered regime with convergent flow in transcription state space, a percolating subnetwork of genes fixed on or off and isolated islands of twinkling genes, a near power-law distribution of cascades of gene activity changes following perturbations, and a square-root relationship between the number of genes and the number of cell types.

Section 1 Introduction

The present article reports an analysis of data pertaining to certain biases in the observed patterns of transcription regulation of eukaryotic genes. In the Boolean idealization, a small subset of possible switching rules, the canalizing functions, are highly utilized. To draw inferences about the implications of the observed biases, a statistical ensemble approach was used. Representative networks constructed within the ensemble of networks that satisfy the biases were studied numerically. The consequences indicate that modeled genomic regulatory systems are in a dynamical "ordered" regime, measurably close to a transition to a "chaotic" regime. A number of testable consequences are derived.

Section 2 Transcription State Spaces, Trajectories, Attractors, and Boolean Net Models

A *state space* is a mathematical abstraction used to describe a dynamical system consisting of a number of interacting variables. The human genomic regulatory system consists of 80,000 to 100,000 interacting genes and their products[1]. A given cell type may express thousands of those genes at any moment. Based on chip technology[2, 3, 4], SAGE analysis [5], quantitative PCR [2, 6, 7] or other approaches, it is rapidly becoming feasible to measure the simultaneous transcript abundance of thousands of genes in single cells, cell lines, or developing systems. Each such measurement gives a snapshot of the current

transcription state of the cells. Snapshots at a timed succession of moments can be linked in a movie that exhibits the trajectory over time of the integrated genetic regulatory system through its state space. The state of a cell at an instant is more complex than a mere transcriptional snapshot, for it includes not only the concentrations of all RNA, protein and other molecular species and species complexes, but their spatial locations and relative motions as well [8, 9].

Genetic regulatory networks can be modeled as systems of continuous[10, 11, 12, 13] or discrete, on/off variables,[14, 15, 16, 17, 18]. For computational tractability, we idealize genetic regulatory networks as Boolean networks. A Boolean network consists of binary (on/off) *nodes* (genes), *links* (causal cis and trans regulatory interactions between genes), and *rules* (relations which specify the next state of a node as a function of the states of its previous inputs). The dynamics are simplified by parallel synchronous update of the entire network. A network and its flow in state space are shown in Figure 1[19, 20, 21]. A binary variable with K inputs has 2^{2^K} possible Boolean functions[16]. We define two classes of Boolean functions, parametrized by different types of bias which are not mutually exclusive. The first bias, *canalization*[16], has the property that at least one input has one value, 1 or 0, which alone suffices to guarantee the activity or inactivity of the regulated variable. For the $K = 2$ OR function, the regulated gene is active at the next moment if either or both of its inputs are currently active. Thus, if either input alone is active, each guarantees that the regulated gene is

active at the next moment. Each such input is a canalizing input. A Boolean function with K inputs may have $0, 1, 2, \ldots, K$ canalizing inputs. The second bias is denoted by a parameter $p : 0.5 \leq p \leq 1.0$, representing the bias away from an equal probability of ones and zeroes in the responses of the Boolean function,[16].

Section 3 Regulation of Eukaryotic Genes Appears to be Strongly Biased Towards Canalizing Functions

To characterize possible biases in known regulated eukaryotic genes we analyzed published data for over 100 regulated genes with K =3, 4 or 5 known direct molecular inputs,(see Supplementary Information). We utilized the following criteria: 1. A known piece of regulatory DNA for a given gene was linked to a reporter gene such as beta-galactosidase or the firefly luciferase gene. 2. A functional assay existed for the expression (transcription) of that piece of regulatory DNA in cells, in vitro, or using transgenic approaches where the control and reporter gene were analyzed in whole organisms. 3. The study used mutational or deletional[22] analysis of the important DNA elements binding the candidate transcription factor(s), or used mutant transcription factors or footprint analysis[23] of the important DNA transcription factor interactions. 4. Many or all of the possible combinations of the transcriptional factors or mutant DNA elements (deletion analysis) were tested or at least reasonably inferred from the study.

Real transcription is not binary, as discussed below. A partial justification for the Boolean idealization lies in the common observation of non-additive collective behavior. Thus, if the level of transcription given input 1 alone is 0.1 of the maximum, given 2 alone is 0.15 of the maximum, and given 1 and 2 together is 1.00, we classified the gene as having $K = 2$ inputs governed by the AND function.

The fraction of all possible Boolean functions which are canalizing on $(1, 2, \ldots, K)$ inputs decreases very rapidly as the number of inputs per node,K, increases. This distribution allows us to test whether actual regulated genes are governed by rules drawn at random from the set of possible Boolean functions. Figures 2a,b show the distribution of numbers of canalizing inputs per gene for $K = 3$ and $K = 4$ as observed from the data and compared to what would be expected from random rule selection. A statistically significant bias towards a high number of canalizing inputs per gene (c) among the sampled regulated eukaryotic genes is observed for $K = 3$ and $K = 4$, (p less than 0.005). There is insufficient data to test for significance for the observed $K = 5$ shift, (K =5 data not shown). Eukaryotic genes are also strongly biased towards high values of p , (data not shown).

Because p and c classes overlap, our results might reflect a bias towards high numbers of canalizing inputs alone, towards high p values alone, or both. To discriminate these, we conditioned on p classes, then tested for residual biases

on the number of canalizing inputs per gene (Tables 1a,b) and found strongly significant residual biases towards high numbers of canalizing inputs for $K = 3$ and $K = 4$ genes. Conversely, after conditioning on the number of canalizing inputs per gene, no residual bias towards high p was found. (Tables 2a,b)

We tentatively conclude from these results that observed regulated eukaryotic genes exhibit a strong bias towards high numbers of canalizing inputs per gene, with no residual bias towards high p values. This conclusion is tempered by the following factors: The papers we chose (see Supplementary Information) may be a biased sample of the known data or may have been misanalyzed by us in carrying out the Boolean idealization. Most importantly, genes governed by canalizing inputs may well be more readily studied experimentally than those governed by non-canalizing Boolean functions. Ultimately this important reservation can be assessed by examining randomly chosen regulated transcription units.

Section 4 An Ensemble Approach Suggests Eukaryotic Genomic Systems are Measurably Within the Ordered Regime

The observed bias towards high numbers of canalizing inputs per gene suggests that large model genomic regulatory networks lie in an ordered dynamical regime close to the transition to chaos, rather than lying in the chaotic regime.[16, 17, 18, 14, 24, 25, 26, 27] To test the expected implications of the observed bias, we constructed ensembles of Boolean networks with $K = 3$, 4, 5, or mixed K

inputs per gene, in which each network was constrained to exhibit the observed biases towards high numbers of canalizing inputs per gene. Except for these biases, network architecture and logic was random. The averaged behaviors of ensemble members exhibit the expected consequences of the observed canalizing bias in the absence of further systematic features such as biases in the connection architecture of the network.

A standard measure to test whether a continuous dynamical system is in the chaotic or ordered regime considers the propagation forward in time of nearby points in state space that lie on distinct trajectories. If the trajectories diverge, exhibiting sensitivity to small changes in initial conditions, this is a signature of chaos. If nearby states on different trajectories converge, this is a signature of order. This analysis can be carried over to discrete dynamical systems [24, 25, 26] by sampling randomly in the state space of the system pairs of states at different initial separations and determining whether, averaged over state space, the trajectories of such states tend to converge or diverge at the next time step. The metric of distance in a discrete system is the normalized *Hamming distance* $H(t)$, which counts the fraction of places in which the two states being compared differ. If the normalized Hamming distance increases, $D(t+1) > D(t)$, this is the discrete analog of chaos, if it decreases, $D(t+1) < D(t)$, it signifies order. Previous work [16, 17, 18, 14, 24, 25, 26, 27] shows that networks with $K \leq 2$ inputs lie in the ordered regime, while networks with $K > 2$ inputs are

chaotic, but can be driven into the ordered regime by increasing p or increasing the number of canalizing inputs per gene.

A simple characterization of the overall behavior of a network is provided by the Derrida plot,[26] Figure 3. The averaged Derrida curves of members of the ensembles of networks matching the observed high numbers of canalizing inputs per gene are shown in Figure 3. In all cases, the Derrida curve indicates that the generic behavior of networks in each ensemble lies modestly in the ordered regime, not too far from the transition to chaos. Mixed networks (a distribution of K values, with appropriate sub-distributions of c values) give results which are again in the ordered regime.

In the $K-c$ plane, numerical work has demonstrated that a decreasing fraction of inputs need to be canalizing to be at the phase transition or in the ordered regime as K increases. As K increases the system will pass from the chaotic into the ordered regime if, on average, about 2.6 or more of the inputs per gene are canalizing. Thus, for genetic networks to lie at a given position slightly in the ordered regime as K increases, the average number of canalizing inputs per gene needs to decrease. Interestingly, the observed fraction of canalizing inputs per gene does decrease as K increases as required. While the data are too scant for the trend to be statistically significant, this tentative observation is consistent with the hypothesis that natural selection has tuned the fraction of canalizing inputs per gene for each K class such that networks are slightly within the ordered regime.

The resulting similarity of the Derrida curves for eukaryotic genes with K = 3, 4, and 5 known inputs, Figure 3, suggests again that natural selection has tuned each K class.

Since networks can be driven into the ordered regime for a given value of $K > 2$ by tuning p or the number of canalizing inputs per gene, it is very interesting that the observed rule biases can be accounted for by a bias in favor of high numbers of canalizing inputs, with no residual bias towards high p values. A bias towards canalizing inputs may reflect chemical simplicity, selection, or other factors.

Figure 3 constitutes evidence that eukaryotic cells lie in the ordered regime. Furthermore, the Derrida test should be experimentally feasible by use of a cell population in which a modest number of randomly chosen genes have exogenously controllable promoters introduced upstream. Then initial perturbations of one or several promoter activities can be tried, the corresponding initial unperturbed and perturbed transcription states assessed by matrix, SAGE, or other techniques, and whether these transcription states converge closer over a short time interval can be directly tested. Indeed, if tried for many cell types, and choices of randomly perturbed gene transcription, this would directly test whether convergence - hence homeostasis - is a global (averaged) property in eukaryotic transcription state spaces[16].

Section 5 Additional Predicted Properties

Percolating Frozen Components, "Twinkling Islands", and Mutual Information Measures

We used the ensemble approach to predict a variety of additional properties of genetic networks with the observed strong bias towards high numbers of canalizing inputs per gene. All the properties we discuss are correlated features of the ordered regime, and have testable consequences. Our analysis involved running 1000 or more simulations of randomly wired networks, but with various values for K inputs and N genes, and rule biases for the K inputs. Statistical properties of the gene network simulations are then gleaned from their global behavior.

The first among these is the formation of a connected frozen component of genes in fixed active or fixed inactive states, leaving behind functionally isolated islands of genes twinkling on and off in complex patterns. This is a global property of these networks independent of the specific wiring but dependent on the K inputs and rule biases.

If a model genetic network is initiated at an arbitrary state far from an attractor state cycle, it flows along a "transient" trajectory to its corresponding attractor. For nets in the ordered regime with the observed canalizing bias, almost all of the nodes turn on and off in complex patterns initially. As the transients progress toward the attractor, many of the nodes settle into fixed active or fixed inactive states. Ultimately these frozen nodes form a large connected (or "percolating"

) cluster whose size scales in proportion to the number of nodes in the entire network. Near or on the attractor the frozen component creates functionally isolated “islands” of coupled genes switching on and off in complex twinkling patterns, Figure 4.

The functional isolation is due to the fact that changes of gene activities within one twinkling island cannot propagate changes of gene activities through the percolating frozen component to another twinkling island. Hence once the frozen component forms, the islands are cut off from one another. By contrast, in the chaotic regime where $K > 2$ and random rule selection is used, small frozen islands may form, but do not create a percolating frozen cluster. Instead, the switching or unfrozen nodes form a percolating twinkling "sea" whose size scales in proportion to the size of the network.

The phase transition from chaotic to ordered behavior as measured by the Derrida curve as network parameters such as the fraction of canalizing inputs increases appears to be associated with a transition from a percolating twinkling sea to isolated twinkling islands, Figure 4.

The predicted occurrence of isolated twinkling islands in the behavior of the real eukaryotic genome, if confirmed experimentally, would be of fundamental importance: First, since each such island typically has more than one attractor itself, such islands may represent the basic decision taking circuitry of the genome.

A cell type is then comprised of a kind of combinatorial epigenetic code[17, 18, 16], consisting of a specific choice of one of the possible attractors for each of the different isolated twinkling islands. Second, it should be experimentally feasible using measurements of cell transcription states of cell type populations at timed intervals to discover which genes are members of each isolated island, for genes in the same island should twinkle in a correlated way, while those in different islands should be uncorrelated.

A straightforward approach to identifying genes within one twinkling island is the "mutual information" measure. [28, 29] The mutual information between two genes A and B, is given by the following:$MI(A,B) = H(A)+H(B)-H(A,B)$ where H is the entropy of the state sequence visited as the net traverses its attractor.

If either or both of gene A or B are frozen active or frozen inactive, the mutual information measure is 0. If genes A and B are twinkling on and off in an entirely uncorrelated way, then their mutual entropy equals the sum of their entropies, and the mutual information is again 0, with statistical fluctuations. But if A and B are twinkling in a correlated way, the mutual information is positive. The obvious and testable hypothesis is that genes in the same island should exhibit positive mutual information, whereas genes in different islands should not.

Figure 5a,b confirms this intuition. For model genes within one isolated

island there is a strong positive signal decreasing roughly exponentially as mutual information increases from 0. For model genes in different isolated islands, the signal is sporadic and low. This suggests that it may be experimentally feasible to discover which genes are members of each twinkling island, hence count the number of such islands, the size distribution of islands, and identify the specific genes within each island. Important caveats to this hope are that these numerical studies are based on synchronous Boolean networks. Extension to more realistic asynchronous and continuous models is needed. In addition, experimental observation of fluctuating (unfrozen) gene activities may often be difficult.

The predicted scaling behavior for the size distribution and mean number of islands in a single network as a function of the number of genes in the network shows that the size distribution obeys a power law while the mean number of islands increases as a logarithmic function of the size of the network, (data not shown). Extrapolating to the human genome with 80,000 to 100,000 genes, only about eight to ten islands are predicted. If these ensemble based predictions are correct, this is encouraging, for the number of islands is only modest and should be rather easily discovered using mutual information measures. Our predicted scaling relations may be sensitive to our assumption of random network connections.

Cascades of Changes in Gene Activities

If a signal (hormone, growth/differentiation factor, etc.) is added to a cell

population, typically an avalanche of changes in gene activities cascades from one or two initial genes directly affected by the signal to dozens or even a few hundred other genes. Such cascades are the concept of a "genetic pathway". "Cross talk" between cascades are the mutual interactions of avalanches started at more or less the same time from different initial genes in a given cell. In terms of the state space picture, an avalanche is an alteration in gene activity patterns due to perturbing the cell from its initial (transcriptional) state to a nearby state. Such a perturbation may leave the system on a transient leading to the same attractor, or to some other attractor. One example would be the choice of differentiation of a mesenchyme cell to either a mature bone cell or to a fat cell, depending on the initial signals.

We can define a gene as "damaged"[16, 25, 27] by a perturbation such as transient exposure to a signal if its on/off behavior is ever different from what it would have been if unperturbed. Once a gene is damaged, it remains damaged even if thereafter its behavior is "normal" or continues to show successive differences with the unperturbed state. The definition of damage allows us to define the size of an avalanche of damage induced by a perturbation such as addition of a signal. We study this computationally by flipping the state of a single node in one copy of a network and monitoring the spread of the difference pattern created by propagating the perturbed and unperturbed networks forward in time. Results, Figure 6, for networks incorporating observed canalizing biases show that

the distribution of avalanche sizes follows a near power-law distribution, truncated with a finite size cutoff which appears to scale as a function of $\sqrt{N}$. Thus, for the human genome with an estimated 80,000 genes [1], maximum avalanches once the frozen component has formed should involve about 800 genes. The predicted size distribution of avalanches of changes of gene activities is directly testable, and, if carried out and confirmed would constitute further evidence that the genomic system lies in the ordered regime. Furthermore, once the frozen structure is in place, any such avalanche should be confined to within one isolated twinkling island. Therefore, study of avalanches should offer an independent experimental means, in addition to mutual information measures, to discover which genes are members of the same functionally isolated island of genes.

The size distribution of avalanches also allows a means to test whether the zygote is initiated on a transient far from any attractor, or is already on an attractor. If the former, then the frozen component has not yet formed, and avalanches of damage should be typical of the chaotic regime, with a power-law distribution of small avalanches and a large number of vast avalanches affecting tens of thousands of genes. If the frozen component is already in place in the zygote, then the largest avalanches should scale as a square root function of the number of genes.

Attractors as Cell Types, Scaling Properties

A tentative interpretation of genetic network models says that cell types

correspond to attractors[16, 14, 17, 18, 15]. If so, then any scaling relation between numbers of attractors and network size, as a function of position in the ordered or chaotic regime, becomes a testable prediction of the theory. We carried out numerical analysis of the scaling behavior for the number of attractors as a function of network size for networks with $K = 3$, 4, and 5 inputs tuned to the observed canalizing bias and found that the number of attractors increases as a square root function of the number of genes, (data not shown). This scaling behavior for $K = 3$, $K = 4$, and $K = 5$ is the same and persists in a relatively broad region around the order/chaos diagonal as defined by the Derrida curve. Similar scaling behavior has been observed computationally and analytically on the $K - p$ boundary between order and chaos[30, 31, 32, 33], though this transition shows conventional phase transition behavior and becomes sharper for larger nets.

In order to test for the number of attractors in a network, we carried out numerical simulations in which the network was initialized with a succession of random initial states and state cycle attractors were encountered and discriminated. In order to test that we had "saturated" the state cycle attractors, we implemented a series of searches in which search was stopped if 4, 20, 100, 500, and 2500 successive initial states lay on trajectories that revealed no new state cycle.

These scaling results, plus the interpretation of an attractor as a cell type, predicts that humans should have about $\sqrt{80000} \cong 283$ cell types. By comparison, Alberts et al estimate 265 cell types for humans, [8]. By similar measures the

number of cell types appears to increase at a rate between a square root and a linear function of the rough number of genes[16]. Our theoretical predictions may be sensitive to the assumption of random connections between genes. Experimental confirmation of these predictions requires assessment of the actual asymptotic dynamical behaviors of genetic networks, and an accurate count of the number of effective genes and other relevant variables in eukaryotic cells.

Gene Expression Overlaps Between Cell Types Cluster

We define the "skeleton" of an attractor to characterize a gene as fixed off, 0, fixed on, 1, and transiently switching, 2. We then measured the overlap between different attractors as the normalized Hamming distance between skeletons, *i.e.* the fraction of genes that are in different "states", 0, 1, or 2, on the skeletons. A typical distance matrix for a network of 1000 genes with $K = 3$ and the observed canalizing bias show that skeletons are within 10% of one another, and may form a hierarchy of distances.

Our results parallel known features of gene expression overlaps between eukaryotic cell types: First, the existence of a percolating frozen component common to all attractors, predicts that all cell types share a common core of genes in the same fixed activities, fixed on or fixed off. This prediction appears to fit older data on the large overlap of gene transcription at the nuclear level in all the different cell types in a given higher eukaryote based on RoT data[34], and

should soon be reexamined with chip, SAGE or quantitative PCR techniques.

Second, in addition to a common core of expressed genes, the typical distribution of differences in gene expression patterns between different cell types is on the order of a few percent, [16]. In general, model and real cell types differ in a few to 10% of the expressed genes, [16]. One example of this property might be that cartilage and bone cell types may have a common skeleton but the detailed structure of their attractors may be quite different. These predicted overlap distributions are directly testable by display chips, SAGE, quantitative PCR, or other means to test the transcriptional state of thousands of genes in different defined cell types.

Section 6 Discussion

Cell and molecular biology is now entering the era in which study of the integrated behavior of genomic regulatory systems, including genes, RNA, proteins, protein modifications, and cell signaling pathways, is emerging as the next major task, [35]. Given the complexity of the cellular system, theory and experiment will increasingly need to be integrated. At least three theory based approaches compliment one another: First, construction of detailed kinetic models of portions of the total "circuitry", [36]; second, reverse engineering by inferences from the temporal patterns of transcription, translation and other molecular species' activities to hypotheses about the circuitry and logic connecting the components[37,

11]; third, use of the ensemble approach to deduce the expected structure and behavior of genomic networks based on any known constraints, testing those predictions, finding new constraints, hence the next improved ensemble. [16, 14, 17, 18, 24, 25, 26, 31, 30, 32, 15, 27, 38]

The three approaches have complementary strengths and weaknesses. Detailed circuit models must deal with the fact that many components of the circuit may not yet be known. Reverse engineering may often lead to many candidate circuits that might account for the temporal patterns observed. An ensemble approach has the strength of predicting properties that are insensitive to many details of network structure and logic, but the weakness that only statistical predictions are made.

The present study is based on the ensemble approach. We have shown evidence suggesting a marked local constraint: Observed regulated eukaryotic genes exhibit a strong bias, in the Boolean idealization, towards canalizing Boolean functions. The most important hesitation with respect to this conclusion is the fact that genes regulated by canalizing functions may be more readily studied. Ultimately, this bias must be assessed using randomly chosen transcription units and their control rules.

We have idealized gene activities as binary variables. More accurate descriptions of gene activities might include continuous or stochastic differential equa-

tions. Reasonable grounds exist to believe that the broad properties of Boolean networks recur in a homologous class of continuous, nonlinear network models. In particular, Glass and his colleagues [10, 11, 12, 13] have studied nonlinear and piecewise linear differential equation network models. Recently, preliminary evidence for the phase transition between order and chaos seen in Boolean networks has been found along the p-K boundary in piecewise linear systems, (Glass, personal communication). Nevertheless, extension of our ensemble studies to nonlinear and stochastic network models are required to establish the robustness of our results.

Our numerical study of Boolean networks with observed canalizing biases revealed a number of robust properties of model genomic regulatory systems.

1) Such systems lie in the ordered regime with slightly convergent flow along neighboring trajectories in state space. Convergence following perturbation in the transcription state of cells is testable.

2) A percolating frozen sub-network arises in which genes are in fixed active or inactive states on all attractors - model cell types.

3) The percolating frozen sub-network leaves behind one or more functionally isolated twinkling islands of genes unable to communicate with one another through the frozen component. Members of each island are discoverable with current experimental techniques.

4) The power-law size distribution of such twinkling islands and near power law distribution of avalanches of damage are predicted and testable.

5) The possibility that in the zygote the frozen component is not yet formed can be tested by the occurrence of very large avalanches of damage following perturbation of single gene activities.

6) The predicted square root scaling relation between the numbers of cell types and the number of genes, the overlaps in gene activity patterns in cell types, and a combinatorial epigenetic code for the alternative cell types of a higher eukaryote are also open to test.

The above predictions demonstrate that an ensemble approach, while limited to statistical predictions, may yield important insight into the integrated behavior of genomic systems. Where predictions fail, the new data can be used to demonstrate further biases in construction, hence the next improved ensemble.

Section 7 Acknowledgments

We thank Andreas Wagner for critical discussion. Graphical simulations were generated using *DDLab*^{TM} software, courtesy of A. W.[19, 20, 21] This research was partially funded by NIH GM49619–04 and The Coleman Foundation 96–05–3003.

Bibliography

[1] G.D. Schuler, M.S. Boguski, and L.D. et al Stewart. A gene map of the human genome. *Science 274:540*, 1996.

[2] J. Eberwine, H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel, and P. Coleman. Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. (USA) 89:3010*, 1992.

[3] D. Lockhart, H. Dong, M. Byrne, M. Follettie, M. Gallo, M. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, and E. Brown. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotech. 14:1675*, 1996.

[4] D. Schena, R. Shalon, R.W. Davis, and P.O. Brown. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science 270:467*, 1995.

[5] V.E. Velculescu, L. Zhang, B. Vogelstein, and K.W. Kinzler. Serial analysis of gene expression. *Science 270:484*, 1995.

[6] N.J. Berry. *Quantification of viral DNA by a nested PCR radiometric incorporation assay. In: PCR: Essential Techniques*. John Wiley and Sons (Burke, J., Ed.). pp. 22-28, 1996.

[7] R. Somogyi, X. Wen, W. Ma, and J.L. Barker. Developmental kinetics of gad family mrnas parallel neurogenesis in the rat spinal cord. *J. of Neuroscience*

15:2575, 1995.

[8] A. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. *Molecular Biology of the Cell*. Garland, New York, 1983.

[9] G. Orphanides, T. Lagrange, and D. Reinberg. The general transcription factors of rna polymerase ii. *Genes and Dev. 10:2657*, 1996.

[10]L. Glass and S.A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol. 39:103*, 1973.

[11]L. Glass and R.E. Young. Structure and dynamics of neural network oscillators. *Brain Research 179:207*, 1979.

[12]John E. Lewis and L. Glass. Steady states, limit cycles, and chaos in models of complex biological networks. *Internat. J. of Bifurcation and Chaos 2:477*, 1991.

[13]T. Mestl, C. Lemay, and L. Glass. Chaos in high-dimensional neural and gene networks. *Physica D1385:1*, 1996.

[14]S.A. Kauffman. Metabolic stability and epigenesis in randomly connected nets. *J. Theor. Biol. 22: 437*, 1969.

[15]S.A. Kauffman. Gene regulation networks: A theory for their global structure and behavior. *Current Topics in Dev. Biol. 6: 145*, 1971.

[16]S.A. Kauffman. *Origins of Order*. Oxford University Press, Oxford, 1993.

[17]S.A. Kauffman. The large-scale structure and dynamics of gene control circuits: An ensemble approach. *J. Theor. Biol. 44: 167*, 1974.

[18]S.A. Kauffman. Emergent properties in random complex automata. *Physica D10: 145*, 1984.

[19]A. Wuensche and M. Lesser. *The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata, Santa Fe Institute Studies in the Sciences of Complexity, Reference Vol. I.* Addison-Wesley, Reading, Mass., 1992.

[20]A. Wuensche. *The Ghost in the Machine; Basin of Attraction Fields of Random Boolean Networks, in Artificial Life III, Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Vol. XVII, C.G.Langton, ed.* Addison-Wesley, Reading, Mass., 1994.

[21]A. Wuensche. *Discrete Dynamics Lab (DDLab).* http://www.santafe.edu/ wuensch/ddlab.html, 1995.

[22]J. Tsien, D. Feng Chen, D. Gerber, C. Tom, E. Mercer, D. Anderson, M. Mayford, E. Kandel, , and S. Tonegawa. Subregion-and cell type-restricted gene knockout in mouse brain. *Cell 87:1317*, 1996.

[23]J. Kadonga, K. Carner, Masiarz. S., and R. Tijian. Isolation of cdna encoding transcription factor sp1 and functional analysis of the dna-binding domain. *Cell 51:1079*, 1987.

[24]B. Derrida and Y. Pomeau. Random networks of automata: A simple annealed approximation. *Europhys. Lett. 1: 45*, 1986.

[25]B. Derrida and D. Stauffer. Phase transitions in two-dimensional kauffman cell automata. *Europhys. Lett. 2: 739*, 1986.

[26]B. Derrida and G. Weisbuch. Evolution of overlaps between configurations in random boolean networks. *J. Physique 47: 1297*, 1986.

[27]G. Weisbuch and D. Stauffer. Phase transition in cellular random boolean nets. *J. Physique 49: 11*, 1987.

[28]Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[29]Claude Shannon and Eric Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Chicago and Urbana, 1949.

[30]U. Bastolla and G. Parisi. Closing probabilities in the kauffman model: an annealed computation. *Physica D98:1*, 1996.

[31]U. Bastolla and G. Parisi. The critical line of kauffman networks: a numerical study. *submitted to J. Theor. Biology*, 1996.

[32]A. Bhattacharjya and S. Liang. Power-law distributions in some random boolean networks. *Phys. Rev. Lett.77:1644*, 1996.

[33]A. Bhattacharjya and S. Liang. Median attractor and transients in random boolean nets. *Physica D95:29*, 1996.

[34]J.J. Monahan, S.E. Harris, and B.W. O'Malley. *Analysis of cellular messenger RnA using complementary DnA probes. In: Receptors and Hormone Action*. Academic Press, New York., Vol. I. pp. 297-329, 1977.

[35]E.S. Lander. The new genomics: Global view of biology. *Science 274:536*, 1996.

[36]H.H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA 94:814*, 1997.

[37]R. Somogyi, S. Fuhrman, M. Askenazi, and A. Wuensche. *The Gene Expression Matrix: Towards the Extraction of Genetic Network Architectures, in Proc. of the Second World Congress of Nonlinear Analysis*. Elsevier Science, 1996.

[38]E. Bienenstock, F. Fogelman-Soulie, and eds. Weisbuch, G. *Disordered Systems and Biological Organization. Series F: Computer and Systems Sciences, Vol. 20*. Springer, New York, 1986.