

A Model of Transcriptional Regulatory Networks Based on Biases in the Observed Regulation Rules

STEPHEN E. HARRIS,¹ BRUCE K. SAWHILL,² ANDREW WUENSCHÉ,³ AND STUART KAUFFMAN²

¹University of Missouri at Kansas City, 650 25th Street, Kansas City, Missouri 64108; ²The Bios Group, 317 Paseo de Peralta, Santa Fe, New Mexico 87501; and ³Discrete Dynamics Inc., www.dlab.com

Received July 28, 2000; revised December 19, 2001; accepted December 19, 2001

Control rules governing transcription of eukaryotic genes can be modeled as Boolean function, and these rules are strongly biased toward large numbers of “canalizing” inputs. The ensemble of networks with the observed canalizing bias predicts cells are in an ordered regime with convergent flow in transcription state space, a percolating subnetwork of genes fixed on or off an isolated islands of twinkling genes turning on or off, and a near power-law distribution of cascades of gene activity changes following perturbations. The data suggest that a given cell state or type can be represented as an attractor of transcriptional activity or flow over time. © 2002 Wiley Periodicals, Inc.

Key Words: transcriptional regulatory networks, Boolean net models, eukaryotic genes, canalizing functions

1. INTRODUCTION

Transcriptional regulatory networks are complex dynamical systems that present science with a great challenge in deciphering the rules that govern their behavior [1]. One approach is to begin to build computer models that capture some of the properties of these networks and then integrate observed and experimental biases and behavior into the computer models. This approach is well understood outside the general area of biology (i.e., cellular automata and Boolean networks), and with the increased capability of monitoring global gene expression in a variety of biological systems, these new mathematical approaches are needed to serve as a language for gene expression information [10–21].

The present article reports an analysis of data pertaining to certain biases in the observed patterns of transcription regulation of eukaryotic genes and builds a model of gene regulatory networks based on those biases. What we found, in a Boolean idealization model, a small subset of possible switching rules, the canalizing functions, are highly utilized in the observed data. To draw inferences about the implications of the observed biases, a statistical ensemble was used. Representative gene 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” state. In a quantitative measure of order and chaos, the gene regulatory networks are on the order side of this continuum,

but far from completely ordered networks. A number of testable consequences are derived.

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 40,000 or so interacting genes and their products [1,A]. A given cell type may express thousands of those genes at any moment. Based on gene microarray technology [2,3,4,B,C], 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 a cell or tissue. 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]. As well, in a tissue there are likely a variety of cell types that interact and are in different states over time. However, computer models of gene regulatory networks can be constructed and tested with presently known biases in how genes are regulated at the transcription level and as large scale expression profiles accumulate in the future for a variety of systems, the computer models can be tested and refined to reflect the new data and observational biases. It is in that light that we have undertaken this initial probabilistic abstract model building of gene regulatory networks with some of the apparent transcriptional biases that have already been uncovered.

Genetic regulatory networks can be modeled as systems of continuous [10–13] or discrete, on/off variables [14–18]. For computational tractability, we idealize genetic regulatory networks as Boolean networks. A Boolean network consists of binary (on/off) *nodes* (genes), *links* (casual *cis* and *trans* regulatory interactions between genes), and *rules* (relations that specify the next state of a node or gene 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 [19–21]. A binary variable with K inputs has 2^K possible Boolean functions [16]. A K -input gene can represent a transcription factor, hormone, coactivator, corepressor, or a given *cis*-DNA element in a gene in which the transacting transcription factor or modulator is unknown. We define two classes of Boolean functions, parameterized by different types of biases that are not mutually exclusive. The first bias, *canalization* [16], has the property that at

least one of the K inputs 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 $0, 1, 2, \dots, K$ 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].

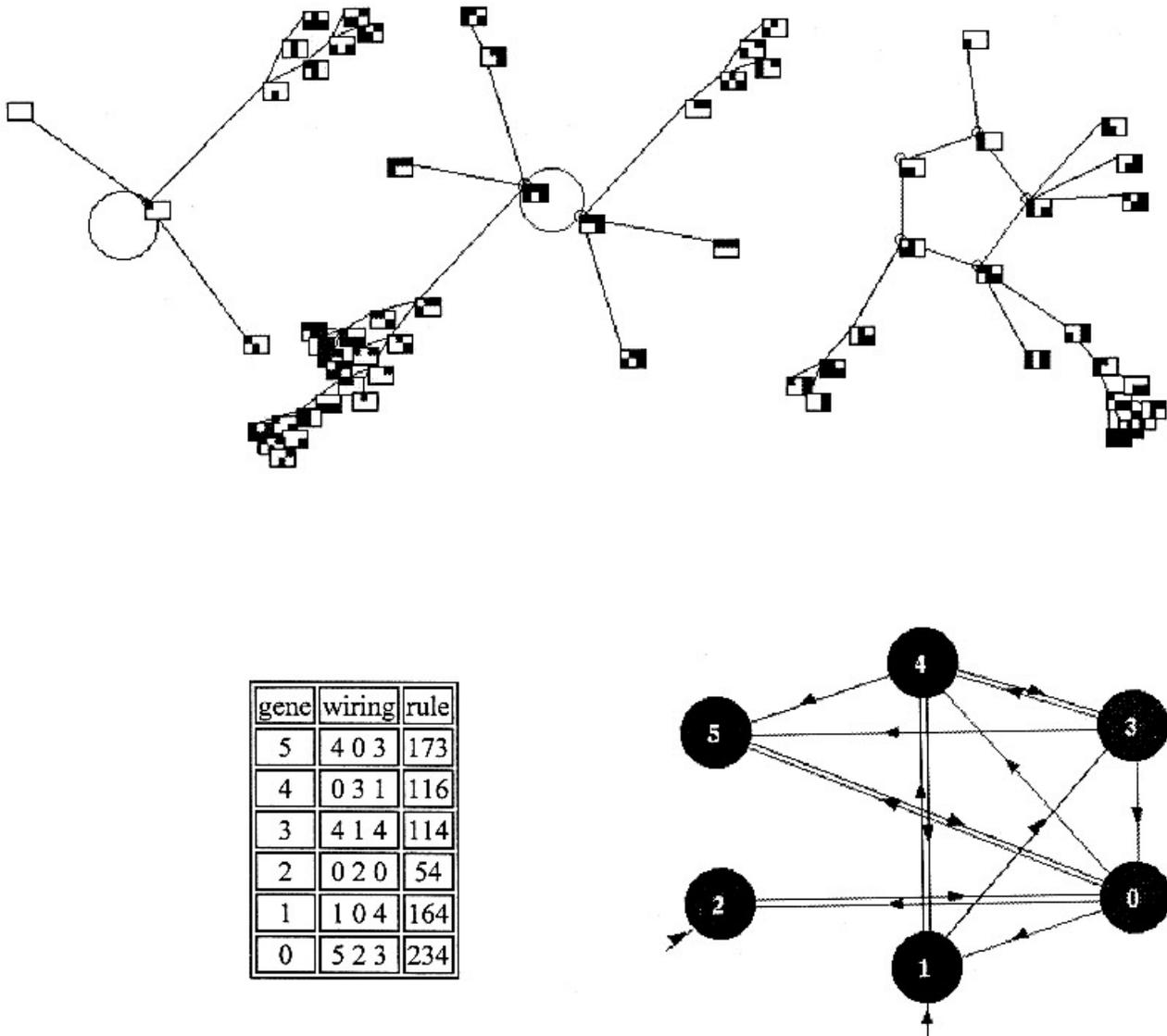
Steady-state mRNA concentrations in a cell for a particular gene are rarely on or off. However, an underlying assumption is that a twofold or greater change, either up or down, in a particular mRNA can have a physiological consequence. Therefore, in our Boolean idealization, the 0 state can represent the twofold or more *decrease* in expression of a gene while the 1 state can represent a twofold or more *increase* in expression of that particular gene with the underlying assumption that the change will affect downstream gene *transcription*. The Boolean model is computationally tractable and yet still reflects the essential dynamics of a gene regulatory network [16]. The model allows us to capture the essential roles of known biases in gene regulatory rules we can derive from known experiments.

3. REGULATION OF EUKARYOTIC GENES APPEARS TO BE STRONGLY BIASED TOWARD CANALIZING FUNCTIONS

To characterize possible biases in known regulated eukaryotic genes we analyzed published data for over 150 regulated transcriptional systems with $K = 3, 4$, or 5 known direct molecular inputs, and a few systems in which $K = 7, 8$, or 9 inputs could be defined (see Appendix A) We used the following criteria:

1. A known piece of regulatory DNA for a given gene was linked to a reporter gene such as β -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 transcription factors or mutant DNA elements (deletion analysis) were tested or at least reasonably inferred from the study. Transcription is not binary, as discussed. A partial justification for the Boolean idealization lies in the common observation of nonadditive collective behavior. Thus, if the level of transcription given input 1 alone is

FIGURE 1



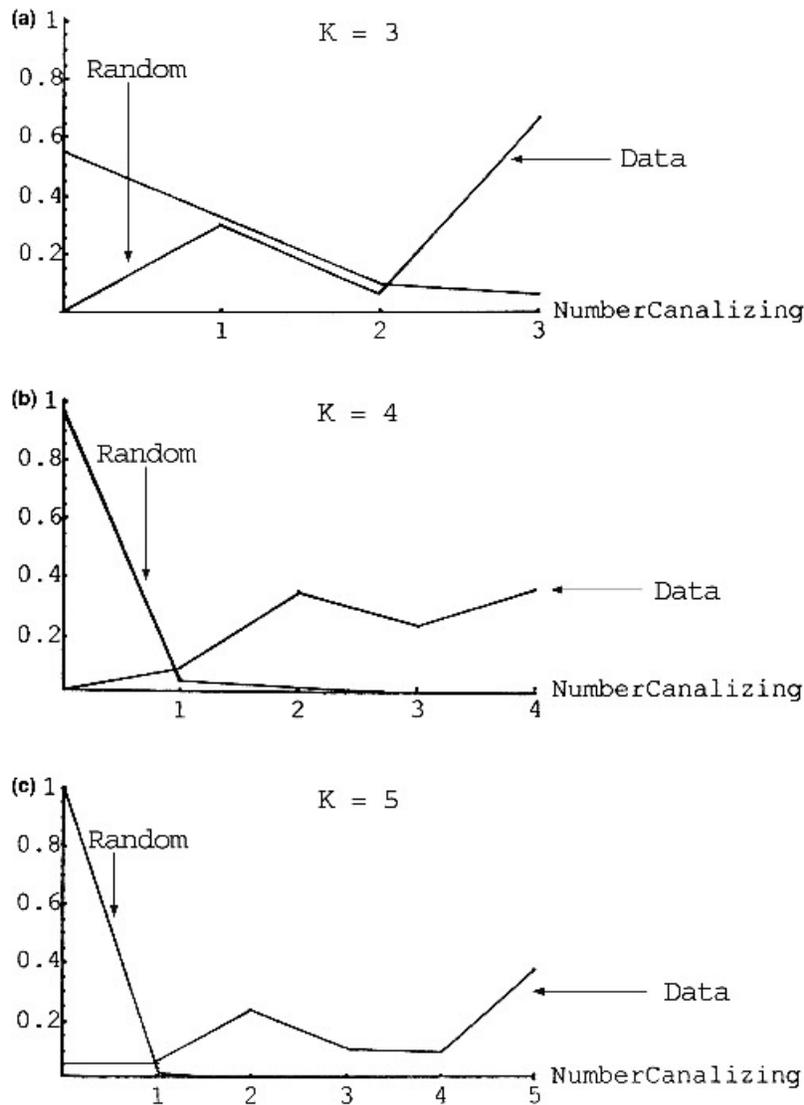
A graphical depiction of the complete state-space of a sample Boolean net of six genes and $K = 3$, assigned at a random. There are three basins of attraction, with attractor periods of 1 (fixed point), 2, and 5. The $2^6 = 64$ states of the six genes are shown within 3×2 rectangles, where active genes are colored, inactive genes are white. Flows proceed inwards, then clockwise around attractor cycles. The actual circuitry (wiring) between the six genes (numbered 0 to 5) is shown on the far right, where self-links are short stubs, and in the table on the right, which also shows the logic functions (rules) for each gene, according to Wolfram's convention [Wolfram 1983]. The computations and graphics were made with DDLab [ref].

0.1 of the maximum, given 2 alone is 0.25 of the maximum, and given 1 and 2 together is 1.00, we classified the gene as having $K = 2$ inputs and of being governed by the "and" function.

The fraction of all possible Boolean functions which are canalizing on $\{1, 2, \dots, K\}$ inputs decreases very rapidly as the number of inputs per node, K , increases. This distribution allows us to test whether actual regulated

genes, modeled as Boolean rules, are governed by rules drawn at random from the set of possible Boolean functions. Figure 2a, b, and c, shows the distribution of numbers of canalizing inputs per gene for $K = 3, 4$, and 5, as observed from the data and compared with what would be expected from random rule selection. A statistically significant bias toward a high number of canalizing inputs per gene, c , among the sampled regulated eukaryotic genes is observed for $K = 3, 4$, and 5 (statistical significance < 0.01). Eukaryotic

FIGURE 2



The upward trending lines are the data, the downward lines represent the distributions of canalyzing from random Boolean functions of K variables.

genes are also strongly biased toward high values of p (data not shown). Because of the overlap between the high p and the canalyzing classes of Boolean functions, our results might reflect a bias toward high numbers of canalyzing inputs alone, toward high p values alone, or both. To discriminate these, after conditioning on p classes, we tested for residual biases on the number of canalyzing inputs per gene (see Appendix B, B1a, B2a, and B3a) and found strong residual biases toward high numbers of canalyzing inputs for $K=3$, $K=4$, and $K=5$ genes. Conversely, no obvious residual bias toward high p was found after conditioning on the number of canalyzing inputs per gene (see Appendix B, Tables B1b, B2b, and B3b). Note that of the 17 cases in Table B1 for $K=3$, 4, and 5, that examine, after controlling for p

values, possible shifts toward high canalyzing inputs per gene, all 17/17 cases show such shifts. Of these cases, 13 are statistically significant. By contrast, the 8 cases of possible shifts, after controlling for c values, toward high p values, show 4/8 cases of shifts to lower p values, 2/8 cases of shifts to higher values, and 2/8 cases of no apparent shifts. Of these, two of the shifts to lower p values were significant statistically, whereas only one of the shifts to higher p values was significant statistically.

We tentatively conclude from these results that observed regulated eukaryotic genes with $K=3$, 4, and 5 inputs show a strong bias toward high numbers of canalyzing inputs per gene, with no residual bias toward

high p values. This conclusion is tempered by the following factors: We may have misread or misanalyzed the published data. 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.

4. AN ENSEMBLE APPROACH SUGGESTS EUKARYOTIC GENOMIC SYSTEMS ARE MEASURABLY WITHIN THE ORDERED REGIME

The observed bias toward high numbers of canalizing inputs per gene suggests that large model transcriptional regulatory networks lie in an ordered dynamical regime not far from the transition to the chaotic region of a dynamical system [14,16–18,24–27]. To test the expected implications of the observed bias, we constructed ensembles of Boolean networks with $K = 3, 4$, and 5 , or mixed K inputs per gene, in which each network was constrained to exhibit the observed biases toward high numbers of canalizing inputs per gene as found from analyses and modeling of the data from laboratory experiments. Except for these biases, network architecture and logic were 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, although recent findings have shown “small-world” structure in genetic regulatory networks [39,40].

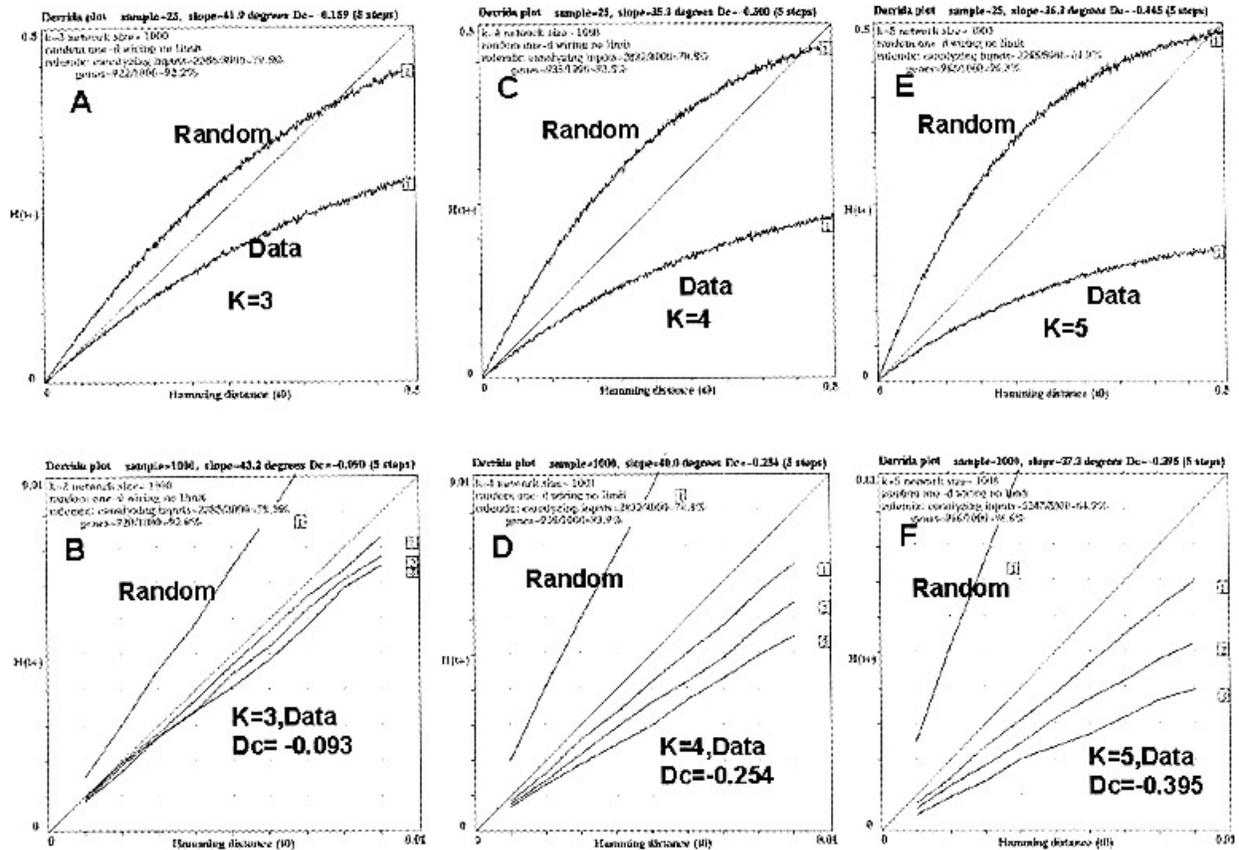
A standard measure to test whether a continuous dynamical system is in the chaotic or ordered regime considers the propagation forward in time of nearby points in state space that lie on distinct trajectories. If the trajectories diverge, exhibiting sensitivity to small changes in initial conditions, 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 [24–26] 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 [14,16–18,24–27] shows that networks with $K < 2$ inputs lie in the ordered regime, whereas 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], and calculation of the Derrida Coefficient, C_D , which is the log of the slope

of the Derrida plot at the origin. The log of the slope at the origin describes the behavior of small perturbations to a network and is the discrete system homologue of the Lyapunov exponent for continuous systems. The scale of the C_D was derived by first building $K = 3, 4$, and 5 networks with all K inputs being canalizing (i.e., $c = 1.0$). C_D was found to be -1.24 , representing the most ordered nontrivial network possible. (A network with all rule outputs being all zeroes or ones yields $C_D = -\infty$, but such networks perform no useful differentiation between input states). Similarly, we built networks with $K = 3, 4$, and 5 with no canalizing inputs, yielding $C_D = 1.7$. Thus, the range $-1.24 \leq C_D \leq 1.7$ allowed us to quantify the degree of order-chaos in the gene regulatory networks built from the bias we observed in the experimental data. The averaged Derrida curve of members of the ensembles of networks with $K = 3$, matching the calculated high numbers of canalizing inputs per gene derived from known experiments is shown in Figure 3A,B. The calculation of C_D also is derived from the data in Figure 3B. As one can see the $C_D = -0.09$ derived from the experimental data indicates that these $K = 3$ networks lie in the ordered regime. In Figure 3C,D are similar presentations of analysis of $K = 4$ networks. The $K = 4$ networks built with $c = 0.708$, derived from the data, gave a $C_D = -0.254$, clearly in the ordered regime. Figure 3E,F shows the data for $K = 5$ networks in which $c = 0.649$, derived from modeling of the experimental data. $C_D = -0.395$ was calculated for this set of $K = 5$ networks, built from the experimental data. In all cases, the Derrida curve and the C_D indicate that the generic behavior of networks in each ensemble lies in the ordered regime, not far from the transition to chaos, which is defined as a $C_D = 0$. There is a trend of decreasing C_D values with increasing K , suggesting that networks with higher K inputs may be more stable and less sensitive to perturbation.

Mixed networks (a distribution of K values, with appropriate subdistributions of c values based on the experimental data) give results that 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 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. Although 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 within the ordered regime. The resulting similarity of the Derrida curves and calculated C_D for eukaryotic

FIGURE 3



Comparison Derrida plots using random rules versus rules based on the fraction canalizing rules (f.c) that were derived from the data for $K = 3, 4,$ and 5 networks (A, C, E). The Derrida coefficient (D_c) is derived from the initial slope of the Derrida plots and was then compared to networks that have random rules, 100% canalizing rules (f.c = 1), or fraction of the rules that are canalizing with a given K input and derived from the data for $K = 3, 4$ and 5 networks (B, D, F). Derrida plots are created by initializing two copies of a Boolean network on two states with different node values. The current “distance” $D(t)$, between these two states is given by the fraction of sites that differ (normalized Hamming distance) and is indexed by the x -axis. Each initial state is propagated one time step forward along its trajectory. The successor distance, $D(t + 1)$, is the normalized Hamming distance between these two successor states, (y -axis). The “Data” curves result from nets of size $N = 1000$, incorporating biases taken from the actual genetic data and using 10 pairs of initial states at 100 different distances for each of 10 networks. The “Random” curves are analytically derived in the $N \rightarrow \infty$ limit, hence their smoothness. Curves below the diagonal are in the “ordered” regime because all initial differences decrease with time, and curves above the diagonal are in the “chaotic” regime because small initial differences increase with time.

genes with $K = 3, 4,$ and 5 known inputs suggests that natural selection may tune each K class (Figure 3).

Because 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 toward high p values. A bias toward 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 tests should be experimentally feasible by use of a cell population in which randomly chosen genes have exogenously controllable promoters introduced upstream. Then initial pertur-

bations of one or several promoters activities can be tried, the corresponding initial unperturbed and perturbed transcription states assessed by gene microarray gene expression profiling, 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]. Some experiments in which one gene is introduced and activated and then the downstream gene activation analyzed by gene expression profiling have begun to appear in the literature. These experiments suggest that eukaryotic gene networks are robust

against major changes in expression patterns, unless the gene is critical for growth or survival.

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 toward 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.

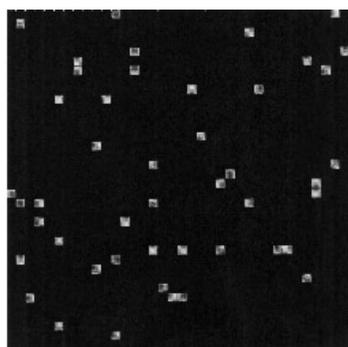
If a model genetic network is initiated at an arbitrary state far from an attractor state cycle, it flows along a "transient" trajectory to its corresponding attractor. For networks in the ordered regime with the observed canalizing bias, almost all of the nodes turn on and off in complex patterns initially. 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 con-

nected (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).

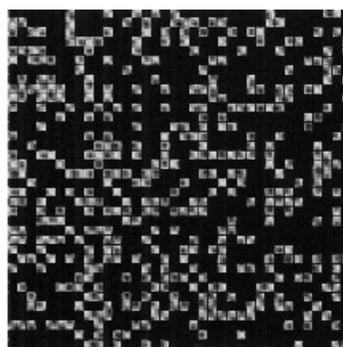
The functional isolation is due to the fact that changes of gene activities within one twinkling island cannot propagate changes of gene activities through the percolating frozen component to another twinkling island. Hence, once the frozen component forms, the islands are cut off from one another. By 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 and Derrida coefficient, because network parameters, such as the fraction of canalizing inputs increases, appear 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, because each such island typically has more than one attractor itself, such islands may represent the basic decision making circuitry of the genome. A cell type is then comprised of a kind of combinatorial epigenetic code [16–18], consisting of a specific choice of one of the possible attractors for each of the different isolated twinkling islands.

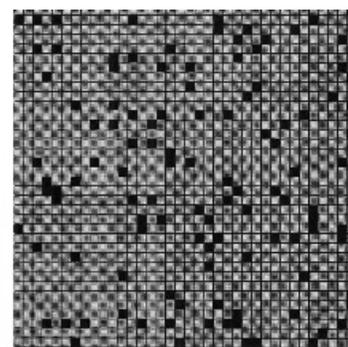
FIGURE 4



1. Chaos (f.c. = 0.22)



2. Near Order/Chaos boundary (f.c. = 0.5) (Slightly Chaotic)



3. Data (f.c. = 0.78) (Ordered Regime)

Tuning canalizing proportion across order-chaos boundary. The three graphics shown here depict in a 2-D format a Boolean Net of size $N = 1296$ (36×36). The connections are random ($K = 3$). Genes have a Poisson distribution of numbers of canalizing inputs, with the mean fraction indicated by f.c. After 1000 time steps sites that have not changed for 50 time steps are colored ("frozen"), with 1 = orange, 0 = green. Sites that continue to change ("twinkle") are black. The eukaryotic data correspond to a regime where the majority of sites are fixed and a small proportion (<10%) change.

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, whereas 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.

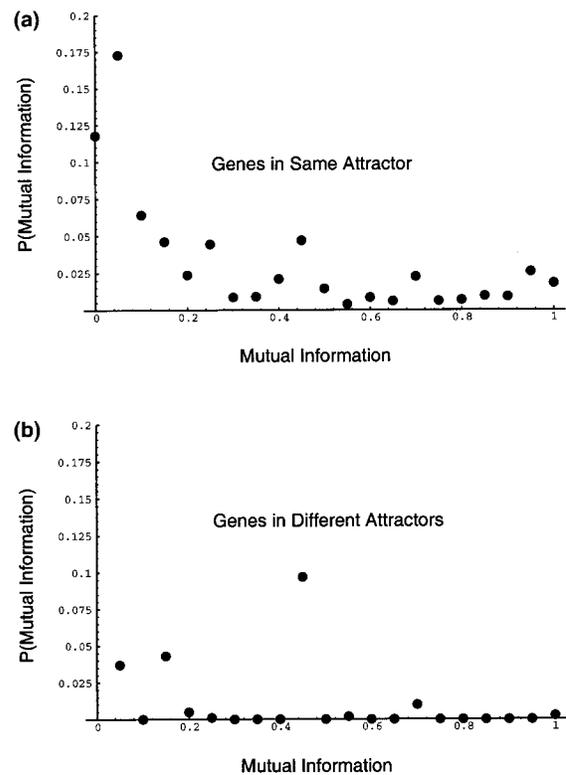
If either or both of genes A or B are frozen active or frozen inactive, the mutual information measure is 0. If genes A and B are twinkling on and off in an entirely uncorrelated way, then their mutual entropy equals the sum of their entropies, and the mutual information is again 0, with statistical 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 5, a and 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. Recent data on continuous models, using our derived fraction canalizing values, support our main conclusions [41]. In addition, experimental observation of fluctuating (unfrozen) gene activities may often be difficult.

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 is 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 mesen-

FIGURE 5



Mutual information measures between genes. Mutual information measures correlation between time sequences of discrete states. The nets used above have $N = 2560$ and $K = 3$, with the canalizing distribution given by the eukaryotic data. The difference between mutual information in same (a) and different (b) attractors is noticeable, though noisy. The “signal-to-noise ratio” improves with net size, however, and should be quite good for nets of size $\sim 80,000$.

chyme 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 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 that appears to scale as a function of $\sim N$

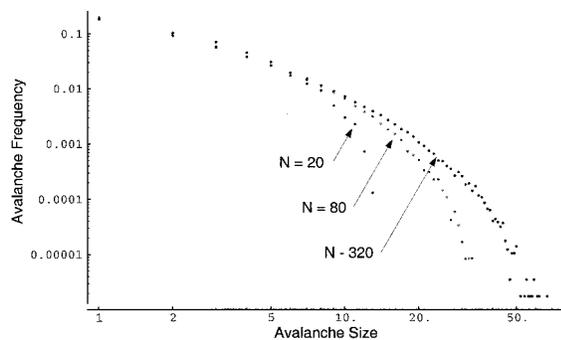
(Figure 6). Thus, for the human genome with an estimated 30 to 40,000 genes [1], maximum avalanches once the frozen component has formed should involve about 200 genes. The predicted size distribution of avalanches of changes of gene activities is directly testable and if carried out and confirmed, would 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 and Their Scaling Properties

A tentative interpretation of genetic network models says that a cell type may correspond to one or more attractors [14–18]. If so, then any scaling relation between numbers of attractors and network size, as a function of position in the

FIGURE 6



Damage avalanches. Damage avalanches are created by flipping one bit randomly in a Boolean Net and observing the spread of “damage.” A site is damaged after a perturbation if its state of activity is ever different than what it would have been. Once a site is damaged, it remains damaged whether or not it returns to normal. The summation stops after the number of affected bits does not change for 20 time steps. The nets in this model were constructed using canalizing bias from the data and a 50/50 mixture of $K=3$ and $K=4$ inputs. Finite size effects cut the curves off at about $2\sqrt{N}$. The above data represents about 300,000 avalanches total.

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 the square root of the size of the genome (data not shown). This scaling behavior for $K=3, 4,$ and 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 [30–33], though this transition shows conventional phase transition behavior and becomes sharper for larger networks.

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 and observed asymptotic convergence of the number of attractors.

Gene Expression Overlaps Between Attractor Clusters

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 cells share a common core of genes in the same fixed activities: fixed on or fixed off. This prediction appears to fit 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 has been re-examined with gene expression microarrays and SAGE techniques [2–6]) or quantitative PCR technique.

Second, in addition to a common core of expressed genes, the typical distribution of differences in gene expression patterns between cells in different states 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, as they move from one state to another [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 distribu-

tions are directly testable by microarray, SAGE, quantitative PCR, or other means to test the transcriptional state of thousands of genes in different defined cell types on various transients of differentiation.

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, or finding new constraints; hence the next improved ensemble [14–18,24–27,30–32,38]. 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, toward 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 equations. 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 [10–13,41] 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 en-

semble studies to nonlinear and stochastic network models as well as discrete multi-state 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 subnetwork arises in which genes are in fixed active or inactive states on all attractors—model cell types or states.
3. The percolating frozen subnetwork 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 overlaps in gene activity patterns in cell types and states along trajectories, and a combinatorial epigenetic code for the alternative cell states of a higher eukaryote are also open to test.

The above predictions demonstrate that an ensemble approach, although 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.

ACKNOWLEDGMENTS

We thank Andreas Wagner for critical discussion. Graphical simulations were generated using “*DDLab*” software, courtesy of A.W. [19–21]. This research was partially funded by National Institutes of Health grant GM49619-04 and The Coleman Foundation grant 96-05-3003. The general research program of the Santa Fe Institute is supported by core funding from the John D. and Catherine T. MacArthur Foundation, the National Science Foundation (PHY 9021437), and the U.S. Department of Energy (DE-FG03-94ER61951), and by gifts and grants from individuals and members of the Institute’s Business Network for Complex Systems research.

REFERENCES

- Schuler, G.D.; Boguski, M.S.; Stewart, L.D.; et al. A gene map of the human genome. *Science* 1996, 274, 540.
- Eberwine, J.; Yeh, H.; Miyashiro, K.; Cao, Y.; Nair, S.; Finnell, R.; Zettel, M.; Coleman, P. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci USA* 1992, 89, 3010.
- Lockhart, D.; Dong, H.; Byrne, M.; Follettie, M.; Gallo, M.; Chee, M.; Mittmann, M.; Wang, C.; Kobayashi, M.; Horton, H.; Brown, E. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol* 1996, 14, 1675.
- Schena, D.; Shalon, R.; Davis, R.W.; Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995, 270, 467.
- Velculescu, V.E.; Zhang, L.; Vogelstein, B.; Kinzler, K.W. Serial analysis of gene expression. *Science* 1995, 270, 484.
- Berry, N.J. Quantification of viral DNA by a nested PCR radiometric incorporation assay. In: *PCR: Essential Techniques*. J. Burke (Ed.), John Wiley and Sons, New York, 1996, 22–28.
- Somogyi, R.; Wen, X.; Ma, W.; Barker, J.L. Developmental kinetics of gad family mRNAs parallel neurogenesis in the rat spinal cord. *J Neurosci* 1995, 15, 2575.
- Alberts, A.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J.D. *Molecular Biology of the Cell*. Garland, New York, 1983.
- Orphanides, G.; Lagrange, T.; Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev* 1996, 10, 2657.
- Glass, L.; Kauffman, S.A. The logical analysis of continuous, nonlinear biochemical control networks. *J Theor Biol* 1973, 39, 103.
- Glass, L.; Young, R.E. Structure and dynamics of neural network oscillators. *Brain Res* 1979, 179, 207.
- Lewis, J.E.; Glass, L. Steady states, limit cycles, and chaos in models of complex biological networks. *Int J Bifurcation Chaos* 1991, 2, 477.
- Mestl, T.; Lemay, C.; Glass, L. Chaos in high-dimensional neural and gene networks. *Physica D* 1996, 1385, 1.
- Kauffman, S.A. Metabolic stability and epigenesis in randomly connected nets. *J Theor Biol* 1969, 22, 437.
- Kauffman, S.A. Gene regulation networks: A theory for their global structure and behavior. *Curr Topics Dev Biol* 1971, 6, 145.
- Kauffman, S.A. *Origins of Order*. Oxford University Press, Oxford, 1993.
- Kauffman, S.A. The large-scale structure and dynamics of gene control circuits: An ensemble approach. *J Theor Biol* 1974, 44, 167.
- Kauffman, S.A. Emergent properties in random complex automata. *Physica D* 1984, 10, 145.
- Wuensche, A.; Lesser, M. *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, MA, 1992.
- Wuensche, A. 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, MA, 1994.
- Wuensche, A. Discrete Dynamics Lab (DDLab) www.ddlab.com.
- Tsien, J.; Feng Chen, D.; Gerber, D.; Tom, C.; Mercer, E.; Anderson, D.; Mayford, M.; Kandel, E.; Tonegawa, S. Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* 1996, 87, 1317.
- Kadonga, J.; Carner, K.; Masiarz, S.; Tijian, R. Isolation of cDNA encoding transcription factor sp1 and functional analysis of the dna-binding domain. *Cell* 1987, 51, 1079.
- Derrida, B.; Pomeau, Y. Random networks of automata: A simple annealed approximation. *Europhys Lett* 1986, 1, 45.
- Derrida, B.; Stauffer, D. Phase transitions in two-dimensional kauffman cell automata. *Europhys Lett* 1986, 2, 739.
- Derrida, B.; Weisbuch, G. Evolution of overlaps between configurations in random boolean networks. *J Physique* 1986, 47, 1297.
- Weisbuch, G.; Stauffer, D. Phase transition in cellular random boolean nets. *J Physique* 1987, 49, 11.
- Cover, T.M.; Thomas, J.A. *Elements of Information Theory*. John Wiley and Sons, 1991.
- Shannon, C.; Weaver, E. *The Mathematical Theory of Communication*. University of Illinois Press, Chicago and Urbana, 1949.
- Bastolla, U.; Parisi, G. Closing probabilities in the kauffman model: An annealed computation. *Physica D* 1996, 98, 1.
- Bastolla, U.; Parisi, G. The critical line of kauffman networks: A numerical study. Submitted to *J Theor Biol* 1996.
- Bhattacharjya, A.; Liang, S. Power-law distributions in some random boolean networks. *Phys Rev Lett* 1996, 77, 1644.
- Bhattacharjya, A.; Liang, S. Median attractor and transients in random boolean nets. *Physica D* 1996, 95, 29.
- Monahan, J.J.; Harris, S.E.; O'Malley, B.W. Analysis of cellular messenger RNA using complementary DNA probes. In: *Receptors and Hormone Action*. Academic Press, New York, Vol. I, 1977, 297–329.
- Lander, E.S. The new genomics: Global view of biology. *Science* 1996, 274, 536.
- McAdams, H.H.; Arkin, A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA* 1997, 94, 814.
- Somogyi, R.; Fuhrman, S.; Askenazi, M.; Wuensche, A. The Gene Expression Matrix: Towards the Extraction of Genetic Network Architectures. In: *Proc of the Second World Congress of Nonlinear Analysis*. Elsevier Science, New York, 1996.
- Bienenstock, E.; Fogelman-Soulie, F.; Weisbuch, G. (Eds.) *Disordered Systems and Biological Organization*. Series F: Computer and Systems Sciences, Vol. 20. Springer, New York, 1986.
- Fell, D.; Wagner, A. The Small World of Metabolism, *Nat Biotechnol* 2000, 18, 1121–1122.
- Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabasi, A.L. The large scale organization of metabolic networks. *Nature* 2000, 407, 651–654.
- Hill, C.; Sawhill, B.; Kauffman, S.; Glass, L. Transition to Chaos in Models of Genetic Networks. Proceedings of the XIV Sitges Conference. In: *Lecture Notes in Physics*. J.M. Rubi and J.M. Vilar (Eds.), Springer-Verlag, New York, 1999.

APPENDIX A

K = 3 References

- Qian, S.; Capovilla, M.; Pirrotta, V. The bx region enhancer, a distant cis-control element of the *Drosophila* Ubx gene and its regulation by hunchback and other segmentation genes. *EMBO J* 1991, 10, 1415–1425.
- Shimell, M.J.; Simon, J.; Bender, W.; O'Connor, M.B. Enhancer point mutation results in a homeotic transformation of *Drosophila*. *Science* 1994, 264, 968–971.
- Kwok, R.P.S.; Lundblad, J.R.; Chrivia, J.C.; Richards, J.P.; Bachlinger, H.P.; Brennan, R.G.; Roberts, S.G.E.; Green, M.R.; Goodman, R.H. Nuclear protein CBO is a coactivator for the transcription factor CREB. *Nature* 1994, 370, 223–226.
- Liu, M.; Freedman, L.P. Transcriptional synergism between vitamin D3 receptor and other nonreceptor transcription factors. *Mol Endo* 1994, 8, 1593–1604.
- Vachon, G.; Cohen, B.; Pfeifle, C.; McGuffin, M.E.; Botas, J.; Cohen, S.M. Homeotic genes of the bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-Less*. *Cell* 1992, 71, 437–450.
- Hoch, M.; Gerwin, N.; Taubert, H.; Jackle, H. Competition for overlapping sites in the regulatory region of the *Drosophila* gene *Kruppel*. *Science* 1992, 256, 94–97.

7. Kruse, F.; Rose, S.D.; Swift, G.H.; Hammer, R.E.; MacDonald, R.J. Cooperation between elements of an organ-specific transcriptional enhancer in animals. *Mol Cell Biol* 1995, 15, 4385–4394.
8. Melamed, D.; Tiefenbrun, N.; Yarden, A.; Kimchi, A. Interferons and interleukin-6 suppress the DNA-binding activity of E2F in growth-sensitive hematopoietic cells. *Mol Cell Biol* 1993, 13, 5255–5265.
9. Halachmi, S.; Marden, E.; Martin, G.; MacKay, H.; Abbondanza, C.; Brown, M. Estrogen receptor-associated proteins: Possible mediators of hormone-induced transcription. *Science* 1994, 264, 1455–1458.
10. Yang, Y.; Minucci, S.; Zand, D.J.; Ozato, K.; Ashwell, J.D. T cell activation and increases in protein kinase C activity enhance retinoic acid-induced gene transcription. *Mol Endo* 1994, 8, 1370–1376.
11. Ayer, D.E.; Eisenman, R.N. A switch from Myc:Max to Mad:Max heterocomplexes accompanies monocyte/macrophage differentiation. *Gene Dev* 1993, 7, 2110–2219.
12. Scott, M.L.; Fujita, T.; Liou, H.-C.; Nolan, G.P.; Baltimore, D. The p65 subunit of NF- κ B regulates I κ B by two distinct mechanisms. *Gene Dev* 1993, 7, 1266–1276.
13. Zhou, J.; Olson, E.N. Dimerization through the helix-loop-helix motif enhances phosphorylation of the transcription activation domains of myogenin. *Mol Cell Biol* 1994, 14, 6232–6243.
14. Ganss, R.; Weih, F.; Schutz, G. The cyclic adenosine 3', 5'-monophosphate- and the glucocorticoid-dependent enhancers are targets for insulin repression of tyrosine aminotransferase gene transcription. *Mol Endo* 1994, 8, 895–903.
15. Kaushal S.; Schneider, J.W.; Nadal-Ginard, B.; Mahdavi, V. Activation of the myogenic lineage by MEF2A, a factor that induces and cooperates with myoD. *Science* 1994, 266, 1236–1240.
16. Kopan, R.; Nye, J.S.; Weintraub, H. The intracellular domain of mouse notch: A constitutively activated repressor of myogenesis directed at the basic helix-loop-helix region of myoD. *Development* 1994, 120, 2385–2396.
17. Lee, I.J.; Driggers, P.H.; Medin, J.A.; Nikodem, V.M.; Ozato, K. Recombinant thyroid hormone receptor and retinoid X receptor stimulate ligand-dependent transcription *in vitro*. *Proc Natl Acad Sci USA* 1994, 91, 1647–1651.
18. Schneider, J.W.; Gu, W.; Zhu, L.; Mahdavi, V.; Nadal-Ginard, B. Reversal of terminal differentiation mediated by p107 in Rb+ muscle cells. *Science* 1994, 264, 1467–1471.
19. Voronova, A.F.; Lee, F. The E2A and tal-1 helix proteins associate *in vivo* and are modulated by Id proteins during interleukin-6-induced myeloid differentiation. *Proc Natl Acad Sci USA* 1994, 91, 5952–5956.
20. MacDonald, P.N.; Dowd, D.R.; Nakajima, S.; Galligan, M.A.; Reeder, M.C.; Haussler, C.A.; Ozato, K.; Haussler, M.R. Retinoid X receptors stimulate and 9-cis retinoic acid inhibits 1,25-dihydroxyvitamin D3-activated expression of the rat osteocalcin gene. *Mol Cell Biol* 1993, 13, 5907–5917.
21. Cheng, T.-C.; Wallace, M.C.; Merlie, J.P.; Olson, E.N. Separable regulatory elements governing myogenin transcription in mouse embryogenesis. *Science* 1993, 261, 215–218.
22. Martin, J.F.; Li, I.; Olson, E.N. Repression of myogenin function by TGF- β 1 is targeted at the basic helix-loop-helix motif and is independent of E2A products. *J Biol Chem* 1992, 267, 10956–10960.
23. Towler, D.A.; Bennett, C.D.; Rodan, G.A. Activity of the rat osteocalcin basal promoter in osteoblastic cells is dependent upon homeodomain and CP1 binding motifs. *Mol Endo* 1994, 8, 614–624.
24. Smith, C.L.; Conneely, O.M.; O'Malley, B.W. Modulation of the ligand-independent activation of the human estrogen receptor by hormone and antihormone. *Proc Natl Acad Sci USA* 1993, 90, 6120–6124.
25. Ignar-Trowbridge, D.M.; Teng, C.T.; Ross, K.A.; Parker, M.G.; Korach, K.S.; McLachlan, J.A. Peptide growth factors elicit estrogen receptor-dependent transcriptional activation of an estrogen-responsive element. *Mol Endo* 1993, 7, 992–998.
26. Shrivastava, A.; Saleque, S.; Kalpana, G.V.; Artandi, S.; Goff, S.P.; Calame, K. Inhibition of transcriptional regulator yin-yang-1 by association with c-myc. *Science* 1993, 262, 1889–1892.
27. Neuhold, L.A.; Wold, B. HLH forced dimers: Tethering myoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* 1993, 74, 1033–1042.
28. Chan, S.-K.; Jaffe, L.; Capovilla, M.; Botas, J.; Mann, R.S. The NDS binding specificity of ultrabithorax is modulated by cooperative interactions with Extradenticle, another homeoprotein. *Cell* 1994, 78, 603–615.
29. Widom, R.L.; Rhee, M.; Karathanasis, S.K. Repression by ARP-1 sensitizes apolipoprotein AI gene responsiveness to RXRa and retinoic acid. *Mol Cell Biol* 1992, 12, 3389–2289.
30. Ikonen, T.; Palvimo, J.J.; Kallio, P.J.; Reinikainen, P.; Janne, O.A. Stimulation of androgen-regulated transactivation by modulators of protein phosphorylation. *Endocrinology* 1994, 135, 1359–1366.
31. van Dijk, M.A.; Mure, C. Extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell* 1994, 78, 617–624.
31. Studer, M.; Popperl, H.; Marshall, H.; Kuroiwa, A.; Krumlauf, R. Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. *Science* 1994, 265, 1728–1732.
32. Ohtani-Fujita, N.; Fujita, T.; Takahashi, R.; Robbins, P.; Dryja, T.P.; Sakai, T. A silencer element in the retinoblastoma tumor-suppressor gene. *Oncogene* 1994, 9, 1703–1711.
33. Yee, S.-P.; Rigby, P.W.J. The regulation of myogenin gene expression during the embryonic development of the mouse. *Gene Develop* 1993, 7, 1277–1289.
34. Darnell, J.E., Jr.; Kerr, I.M.; Startk, G.R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 1994, 264, 1415–1421.
35. Cserjesi, P.; Olson, E.N. Myogenin induces the myocyte-specific enhancer binding factor MEF-2 independently of other muscle-specific gene products. *Mol Endo* 1996, 10, 4854–4862.
36. Patrone, C.; Ma, Z.Q.; Pollio, G.; Agrati, P.; Parker, M.G.; Maggi, A. Cross-coupling between insulin and estrogen receptor in human neuroblastoma cells. *Mol Endo* 1996, 10, 499–507.
37. Chen, G.; Baltimore, D. TANK, a co-inducer with TRAF2 of TNF- and CD40L-mediated NF κ B activation. *Gene Dev* 1996, 10, 963–973.
38. Foulkes, N.S.; Duval, G.; Sassone-Corsi, P. Adaptive inducibility of CREM as transcriptional memory of circadian rhythms. *Nature* 1996, 381, 83–85.
39. Kamei, Y.; Xu, L.; Heinzel, T.; Torchia, J.; Kurokawa, R.; Glass, B.; Lin, S.C.; Heyman, R.A.; Rose, D.W.; Glass, C.K.; Rosenfeld, M.G. A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* 1996, 85, 403–414.
40. Cooper, J.P.; Roth, S.Y.; Simpson, R.T. The global transcription regulators, SSN6 and TUP1, play distinct roles in the establishment of a repressive chromatin structure. *Gene Dev* 1994, 8, 1400–1410.
41. Gray, S.; Szymanski, P.; Levine, M. Short-range repression permits multiple enhancers to function autonomously within a complex promoter. *Gene Dev* 1994, 7, 491–503.
42. Arnold, S.F.; Klotz, D.M.; Collins, B.M.; Vonier, P.M.; Guillet, L.J., Jr.; McLachlan, J.A. Synergistic activation of estrogen receptor with combinations of environmental chemicals. *Science* 1996, 272, 1489–1492.
43. Norris, J.L.; Manley, J.L. Functional interactions between the Pelle kinase, toll receptor, and tube suggest a mechanism for activation of dorsal. *Gene Dev* 1996, 10, 862–872.
44. Forman, B.M.; Umesono, K.; Chen, J.; Evans, R.M. Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell* 1995, 81, 541–550.
45. Beardsley, T. Smart genes. *Sci Am* 1991, August, 87–95.

46. Sauer, F.; Hansen, S.K.; Tijan, R. DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* Bicoid. *Sci* 1995, 270, 1825–1828.
47. Tremml, G.; Bienz, M. Induction of labial expression in the *Drosophila* endoderm: Response elements for *dpp* signalling and for autoregulation. *Development* 1992, 116, 447–456.
48. Johg, M.T.C.; Raaka, B.M.; Samuels, H.H. A sequence in the rat Pit-1 gene promoter confers synergistic activation by glucocorticoids and protein kinase-C. *Mol Endo* 1994, 8, 1320–1327.
49. Scholz, A.; Truss, M.; Beato, M. Hormone-induced recruitment of Sp1 mediates estrogen activation of the rabbit uteroglobin gene in endometrial epithelium. *J Biol Chem* 1998, 273, 4360–4366.
50. Wissink, E.; van Heerde, E.C.; van der Burg, B.; van der Saag, P.T. A dual mechanism mediates repression of NF- κ B activity by glucocorticoids. *Mol Endocrinol* 1998, 12, 355–363.
51. Huang, J.D.; Dubnicoff, T.; Liaw, G.J.; Bai, Y.; Valentine, S.A.; Shirokawa, J.M.; Lengyel, J.A.; Courey, A.J. Binding sites for transcription factor NTF-1/Elf-1 contribute to the ventral repression of decapentaplegic. *Genes Dev* 1995, 9, 3177–3189.
52. Gould, A.; Morrison, A.; Sproat, G.; White, R.A.H.; Krumlauf, R. Positive cross-regulation and enhancer sharing: Two mechanisms for specifying overlapping HOX expression patterns. *Genes Dev* 1997, 11, 900–913.
53. Chen, L.; Glover, J.N.M.; Hogan, P.G.; Rao, A.; Harrison, S.C. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* 1998, 392, 42–48.
54. Naya, F.J.; Stellrecht, C.M.M.; Tsai, M.J. Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. *Genes Dev* 1995, 9, 1009–1019.
55. Rachez, C.; Lemon, B.D.; Suldan, Z.; Bromleigh, V.; Gamble, M.; Naar, A.M.; Erdjument-Bromage, H.; Tempst, P.; Freedman, L.P. Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex. *Nature* 1999, 398, 824–828.
56. Dai, P.; Akimaru, H.; Tanaka, Y.; Maekawa, T.; Nakafuku, M.; Ishii, S. Sonic hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. *J Biol Chem* 1999, 274, 8143–8152.
57. Atkins, G.B.; Hu, X.; Guenther, M.G.; Rachez, C.; Freedman, L.P.; Lazar, M.A. Coactivators for the orphan nuclear receptor ROR α . *Mol Endocrinol* 1999, 13, 1550–1557.
58. Schreiber, M.; Kolbus, A.; Piu, F.; Szabowski, A.; Mohle-Steinlein, U.; Tian, J.; Karin, M.; Angel, P.; Wagner, E.F. Control of cell cycle progression by c-Jun is p53 dependent. *Genes Dev* 1999, 13, 607–619.
59. Nakashima, K.; Yanagisawa, M.; Arakawa, H.; Kimura, N.; Hisatsune, T.; Kawabata, M.; Miyazono, K.; Taga, T. Synergistic signaling in fetal brain by STAT3-smad1 complex bridged by p300. *Science* 1999, 284, 479–482.
60. Shikama, N.; Lee, C.W.; France, S.; Delavaine, L.; Lyon, J.; Krstic-Demonacos, M.; La Thangue, N.B. A novel cofactor for p300 that regulates the p53 response. *Mol Cell* 1999, 4, 365–376.
61. Kurokawa, R.; Kalafus, D.; Ogliaastro, M.H.; Kioussi, C.; Xu, L.; Torchia, J.; Rosenfeld, M.K.G.; Glasfs, C.K. Differential use of CREB bending protein-coactivator complexes. *Science* 1998, 279, 700.
62. Hua, X.; Liu, X.; Ansari, D.O.; Lodish, H.F. Synergistic cooperation of TFE3 and Smad proteins in TGF β -induced transcription of the plasminogen activator inhibitor-1 gene. *Genes Dev* 1998, 12, 3084–3095.
63. Clarkson, R.W.E.; Shang, C.A.; Levitt, L.K.; Howard, T.; Waters, M.J. Ternary complex factors Elk-1 and Sap-1a mediate growth hormone-induced transcription of Egr-1 (early growth response factor-1) in 3T3-F442A preadipocytes. *Mol Endocrinol* 1999, 13, 619–631.
2. Sauer, F.; Fondel, J.D.; Ohkuma, Y.; Roeder, R.G.; Jaeckle, H. Control of transcription by krippel through interactions with TFIIIB and THIEP. *Nature* 1995, 375, 162–164.
3. Giese, K.; Kingsley, C.; Kirshner, J.R.; Grosschedl, R. Assembly and function of a TCRA enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Gene Dev* 1995, 9, 995–1008.
4. Love, J.J.; Li, X.; Case, D.A.; Giese, K.; Grosschedl, R.; Wright, P.E. Structural basis for DNA bending for the architectural transcription factor LEF-1. *Nature* 1995, 376, 791–795.
5. Martin, K.; Trouche, D.; Hangemeier, C.; Sorensen, T.S.; LaThangue, N.B.; Kouzarides, T. Stimulation of E2F1/DP1 transcriptional activity by MDM2 oncoprotein. *Nature* 1995, 375, 691–697.
6. Beardsley, T. *Smart genes*. *Sci Am* 1991, August, 86–95.
7. Jackson, S.M.; Gutierrez-Hartmann, A.; Hoefler, J.P. Upstream stimulatory factor, a basic-helix-loop-helix-zipper protein, regulates the activity of the u-glycoprotein hormone subunit gene in pituitary cells. *Mol Endo* 1995, 9, 278–291.
8. Rajnarayan, S.; Chiono, M.; Alexander, L.M.; Gutierrez-Hartmann, A. Reconstitution of protein kinase A regulation of the rat prolactin promoter in HeLa nonpituitary cells: Identification of both GHF-1/Pit-1-dependent and -independent mechanisms. *Mol Endo* 1995, 9, 401–412.
9. Caldenhoven, E.; Liden, J.; Wissink, S.; Van de Stolpe, A.; Raaijmakers, J.; Koenderman, L.; Okert, S.; Gustafsson, J.-A.; Van der Saag, P.T. Negative cross-talk between RelA and the glucocorticoid receptor: A possible mechanism for the antiinflammatory action of glucocorticoids. *Mol Endo* 1995, 9, 401–412.
10. Eraly, S.A.; Mellon, P.L. Regulation of gonadotropin-releasing hormone transcription by protein kinase C is mediated by evolutionarily conserved promoter-proximal elements. *Mol Endo* 1995, 9, 848–859.
11. Keller, H.; Givel, F.; Perroud, M.; Wahil, W. Signaling cross-talk between peroxisome proliferator-activated receptor/retinoid X receptor and estrogen receptor through estrogen response elements. *Mol Endo* 1995, 9, 794–804.
12. Piedrafita, F.J.; Bendik, I.; Ortiz, M.A.; Pfahl, M. Thyroid hormone receptor homodimers can function as ligand-sensitive repressors. *Mol Endo* 1995, 9, 563–578.
13. Popperl, H.; Bienz, M.; Studer, M.; Chan, S.-K.; Aparicio, S.; Brenner, S.; Mann, R.S.; Krumlauf, R. Segmental expression of Hoxb-1 is controlled by highly conserved autoregulatory loop dependent upon exd/pbx. *Cell* 1995, 81, 1031–1042.
14. Whyte, D.B.; Lawson, M.A.; Belsham, D.D.; Early, S.A.; Bond, C.T.; Adelman, J.P.; Mellon, P.L. A neuron-specific enhancer targets expression of the gonadotropin-releasing hormone gene to hypothalamic neurosecretory neurons. *Mol Endo* 1995, 9, 467–477.
15. Manak, J.R.; Mathies, L.D.; Scott, M.P. Regulation of a decapentaplegic midgut enhancer by homeotic protein. *Development* 1995, 120, 3065–3619.
16. Ogura, T.; Evans, R.M. A retinoic acid-triggered cascade of Hoxbl gene activation. *Proc Natl Acad Sci USA* 1995, 92, 387–391.
17. Chiang, C.-M.; Roeder, R.G. Cloning of an intrinsic human TFIIID subunit that interacts with multiple transcriptional activators. *Science* 1995, 276, 531–535.
18. Singh, P.; Coe, J.; Hong, W. A role for retinoblastoma protein in potentiating transcriptional activation by the glucocorticoid receptor. *Nature* 1995, 374, 562–565.
19. Chawla, A.; Lazar, M.A. Peroxisome proliferator and retinoid signaling pathways co-regulate preadipocyte phenotype and survival. *Proc Natl Acad Sci USA* 1994, 91, 1786–1790.
20. Ayer, D.E.; Lawrence, Q.A.; Eisenman, R.N. Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* 1995, 80, 767–776.

K = 4 References

1. Roberts, S.G.E.; Green, M.R. Dichotomous regulators. *Nature* 1995, 375, 105–106.

21. Forman, B.M.; Umesono, K.; Chen, J.; Evans, R.M. Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell* 1995, 81, 541–55D.
22. Zeng, C.; Pinsonneault, J.; Gellon, G.; McGinnis, N.; McGinnis, W. Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in *Drosophila* embryos. *EMBO J* 1994, 13, 2362–2377.
23. Onate, S.A.; Tsai, S.Y.; Tsai, M.-J.; O'Malley, B.W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* 1995, 270, 1354–1355.
24. Puschel, A.W.; Balling, R.; Gruss, P. Separate elements cause lineage restriction and specify boundaries of Hox-1.1 expression. *Development* 1991, 112, 279–287.
25. Starr, D.B.; Thomas, J.R.; Yamamoto, K.R. A single amino acid residue governs a switch between transcription activation and repression by the glucocorticoid receptor. Abstract 445. Keystone symposium/steroid-thyroid-retinoid gene family, March 17, 1996.
26. Robertson, L.M.; Kerppola, T.K.; Vendrell, M.; Luk, D.; Smayne, R.J.; Bocchiaro, C.; Morgan, J.L.; Curran, T. Regulation of c-fos expression in transgenic mice require multiple interdependent transcription control elements. *Neuron* 1995, 14, 241–252.
27. Genetta, T.; Ruezinsky, D.; Kadesch, T. Displacement of a E-box binding repressor by basic helix loop helix proteins: Implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol Cell Biol* 1994, 14, 6153–6163.
28. Laney, J.D.; Biggin, M.D. Redundant control of Ultrabithorax by Zeste involves functional levels of Zeste protein binding at the Ultrabithorax promoter. *Development* 1996, 122, 2303–2311.
29. Kim, J.B.; Spiegelman, B.M. ADD/SREBP1 promotes adipocyte differentiation and gene expression linked to fatty acid metabolism. *Genes Dev* 1996, 10, 1096–1107.
30. Wang, J.-C.; Stromstead, P.-E.; O'Brien, R.; Granner, D.K. Hepatic nuclear factor (HNF 3) serves as a accessory factor for the stimulation of phosphoenolpyruvate carboxy kinase gene by glucocorticoids. Abstract 459. Keystone symposium/steroid-thyroid-retinoid gene family, March 17, 1996.
31. Seol, W.; Choi, H.-S.; Moore, D.D. An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors. *Science* 1996, 272, 1336–1339.
32. Mermelstein, F.; Yeung, K.; Cao, J.; Inostronza, J.A.; Erdjument-Bromage, H.; Egelson, K.; Landsman, D.; Levitt, P.; Tempst, P.; Reinberg, D. Requirement of a corepressor for Drl-mediated repression of transcription. *Genes Dev* 1996, 10, 1033–1048.
33. Spicer, D.B.; Rhee, J.; Cheung, W.L.; Lassar, A.B. Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein twist. *Science* 1996, 272, 1476–1480.
34. Kephart, D.D.; Walfish, P.G.; DeLuca, H.; Butt, T.R. Retinoid X receptor isotype identity directs human vitamin D receptor heterodimer transactivation from the 24-hydroxylase vitamin D response elements in yeast. *Mol Endo* 1996, 10, 408–419.
35. Grueneberg, D.A.; Henry, R.W.; Brauer, A.; Novina, C.D.; Cheriya, V.; Roy, A.L.; Gilman, M. A multifunctional DNA-binding protein that promotes the formation of serum complexes: Identity to TFII-I. *Genes Dev* 1997, 11, 2482–2493.
36. Hunter, T. Prolyl isomerases and nuclear function. *Cell* 1998, 92, 141–143.
37. Zwijsen, R.M.L.; Buckle, R.S.; Hijmans, M.; Loomans, C.J.M.; Bernards, R. Ligand-independent recruitment of steroid receptor coactivators to estrogen receptor by cyclin D1. *Genes Dev* 1998, 12, 3488–3498.
38. Mattias, M.; Nibu, Y.; Zhang, H.; Levine, M. Transcriptional coregulators in development. *Science* 1999, 284, 606–609.
39. Riese, J.; Yu, X.; Munnerlyn, A.; Eresh, S.; Hsu, S.C.; Grosschedl, R.; Bienz, M. LEF-1, a nuclear factor coordinating signaling inputs from wingless and decapentaplegic. *Cell* 1997, 88, 777–787.
40. Gervois, P.; Torra, I.P.; Chinetti, G.; Grotzinger, T.; Dubois, G.; Fruchart, J.C.; Fruchart-Najib, J.; Leitersdorf, E.; Staels, B. A truncated human peroxisome proliferator-activated receptor α splice variant with dominant negative activity. *Mol Endocrinol* 1999, 13, 1535–1549.
41. Partridge, N.; et al. The collagenase 3 promoter is synergistically regulated by PTH through phosphorylation of Cbfa1 by PKA dependent pathway. *J Biol Chem* (in press).
42. Nishihara, A.; Hanai, J.; Imamura, T.; Miyazono, K.; Kawabata, M. E1A inhibits transforming growth factor β signaling through binding to Smad proteins. *J Biol Chem* 1999, 274, 28716–28723.
43. Chen, G.; Fernandez, J.; Mische, S.; Courey, A.J. A functional interaction between the histone deacetylase Rpd3 and the corepressor Groucho in *Drosophila* development. *Genes Dev* 1999, 13, 2218–2230.
44. Wang, E.H.; Zou, S.; Tjian, R. TAF11250-dependent transcription of cyclin A is directed by TF activator proteins. *Genes Dev* 1997, 11, 2658–2669.
45. Zhou, S.; Zewel, L.; Lengauer, C.; Kinzler, K.W.; Vogelstein, B. Characterization of human FAST-1, a TGF β and activin signal transducer. *Mol Cell* 1998, 2, 121–127.
46. Forman, B.M.; Tzameli, I.; Cho, H.S.; Chen, J.; Simha, D.; Seol, W.; Evans, R.M.; Moore, D.D. Androstane metabolites bind to and deactivate the nuclear receptor CAR- β . *Nature* 1998, 395, 612–615.
47. Ray, A.; Pregontaine, D.E. Physical association and functional antagonism between the p65 subunit of transcription factor NF- κ B and the glucocorticoid receptor. *Proc Natl Acad Sci USA* 1994, 91, 752–756.

K = 5 References

1. Liaw, G.-J.; Rudolph, K.M.; Huang, J.-D.; Dubnicoff, T.; Coffey, A.J.; Lengyel, J.A. The torso response element binds GAGA and NTF-1/Elf-1, and regulates tailless by relief of repression. *Gene Dev* 1995, 9, 3163–3178.
2. Jian, J.; Cai, H.; Zhou, Q.; Levine, M. Conversion of a dorsal-dependent silencer into an enhancer evidence for dorsal corepressors. *EMBO J* 1993, 12, 3201–3209.
3. Sauer, F.; Hansen, S.K.; Tjian, R. Multiple TAFs directing synergistic activation of transcription. *Science* 1995, 276, 1783–1788.
4. Huang, J.-D.; Dubnicoff, T.; Liaw, G.-J.; Bai, Y.; Valentine, S.A.; Shirokawa, J.M.; Lengyel, J.A.; Courey, A.J. Binding sites for transcription factor NTF-1/ELE-1 contribute to the ventral repression of Decapentaplegic. *Gene Dev* 1995, 9, 3177–3189.
5. Jiang, J.; Levine, M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* 1993, 72, 741–752.
6. Kurokawa, R.; Soderstrom, M.; Horlein, A.; Halachmi, S.; Brown, M.; Rosenfeld, M.G.; Glass, C.K. Polarity-specific activities of retinoic acid receptors determined by a co-repressor. *Nature* 1994, 377, 451–454.
7. Gu, W.; Schneider, J.W.; Condorelli, G.; Kaushal, S.; Mahdavi, V.; Nadal-Ginard, B. Interaction of myogenic factors and the retinoblastoma protein mediates muscle cell commitment and differentiation. *Cell* 1993, 72, 309–327.
8. Falvo, J.V.; Thanos, D.; Maniatis, T. Reversal of intrinsic DNA bends in the IFN β gene enhancer by transcription factors and the architectural protein HMGI(Y). *Cell* 1995, 83, 1101–1111.
9. Paroush, Z.; Finley, R.L., Jr.; Kidd, T.; Wainwright, S.M.; Ingham, P.W.; Brent, R.; Ish-Horowicz, D. Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with Hairy-related bHLH proteins. *Cell* 1994, 79, 805–815.

10. Lehming, N.; Thanos, D.; Brickman, M.J.; Ma, J.; Maniatis, T.; Ptashne, M. An HMG-like protein that can switch a transcriptional activator to a repressor. *Nature* 1994, 371, 175–179.
11. Kurokawa, R.; Kalafus, D.; Ogliaastro, M.H.; Kioussi, C.; Xu, L.; Torchia, J.; Rosenfeld, M.G.; Glass, C.K. Differential use of CREB bending protein-coactivator complexes. *Science* 1998, 279, 700–703.
12. Korzus, E.; Torchia, J.; Rose, D.W.; Xu, L.; Kurokawa, R.; McInerney, E.M.; Mullen, T.M.; Glass, C.K.; Rosenfeld, M.G. Transcription factor-specific requirements for coactivators and their acetyltransferase functions. *Science* 1998, 279, 703.
13. Lin, R.J.; Nagy, L.; Inoue, S.; Shaos, W.; Miller, W.H., Jr.; Evans, R.M. Role of the histone deacetylase complex in acute promyelocytic leukaemia. *Nature* 1998, 391, 811–814.
14. Labbe, E.; Silvestri, C.; Hoodless, P.A.; Wrana, J.L.; Attisano, L. Smad2 and Smad3 positively and negatively regulate TGF β -dependent transcription through the forkhead DNA-binding protein FAST2. *Mol Cell* 1998, 2, 109–120.
15. Seagroves, T.N.; Krnacik, S.; Raught, B.; Gay, J.; Burgess-Beusse, B.; Darlington, G.J.; Rosen, J.M. C/EBP β , but not C/EBP α , is essential for ductal morphogenesis, lobuloalveolar proliferation, and functional differentiation in the mouse mammary gland. *Genes Dev* 1998, 12, 1917–1928.
16. Bruhn, L.; Munnerlyn, A.; Grosschedl, R. ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCR α enhancer function. *Genes Dev* 1997, 11, 640–653.
17. Duan, R.; Porter, W.; Samudio, I.; Vyhlidal, C.; Klade, M.; Safe, S. Transcriptional activation of c-fos protooncogene by 17 β -estradiol: mechanism of aryl hydrocarbon receptor-mediated inhibition. *Mol Endocrinol* 1999, 13, 1511–1521.
18. Luo, K.; Stroschein, S.L.; Wang, W.; Chen, D.; Martens, E.; Zhou, S.; Zhou, Q. The Ski oncoprotein interacts with the Smad proteins to repress TGF β signaling. *Genes Dev* 1999, 13, 2196–2206.
19. Waltzer, L.; Bienz, M. A function of CBP as a transcriptional coactivator during Dpp signaling. *EMBO J* 1999, 18, 1630–1641.
20. Moustakas, A.; Kardassis, D. Regulation of the human p21/WAF1/Cep1 promoter in hepatic cells by functional interactions between

Sp1 and Smad family members. *Proc Natl Acad Sci USA* 1998, 95, 6733–6738.

21. Stroschein, S.L.; Wang, W.; Zhou, S.; Zhou, Q.; Luo, K. Negative feedback regulation of TGF β signaling by the SnoN oncoprotein. *Science* 1999, 286, 771–774.
22. Arango, N.A.; Lovell-Badge, R.; Behringer, R.R. Targeted mutagenesis of the endogenous mouse *Mis* gene promoter: In vivo definition of genetic pathways of vertebrate sexual development. *Cell* 1999, 99, 409–419.
23. Puigserver, P.; Adelmant, G.; Wu, Z.; Fan, M.; Xu, J.; O'Malley, B.; Spiegelman, B.M. Activation of PPAR γ coactivator-1 through transcription factor docking. *Science* 1999, 286, 1368–1372.
24. Sauer, F.; Wassarman, D.A.; Rubin, G.M.; Tigan, R. TAFIIS mediate activation of transcription in the *Drosophila* embryo. *Cell* 1996, 87, 1271–1284.

K = 7, 8, 9 References

1. Korzus, E.; Torchia, J.; Rose, D.W.; Xu, L.; Kurokawa, R.; McInerney, E.M.; Mullen, T.M.; Glass, C.K.; Rosenfeld, M.G. Transcription factor-specific requirements for coactivators and their acetyltransferase functions. *Science* 1998, 279, 703–707.
2. Drenzo, J.; Soderstrom, M.; Kurokawa, R.; Ogliaastro, M.H.; Ricte, M.; Ingrey, S.; Horlein, A.; Rosenfeld, M.G.; Glass, C.K. Peroxisome proliferator-activated receptors and retinoic acid receptors differentially control the interactions of retinoid X receptor heterodimers with ligands, coactivators, and corepressors. *Mol Cell Biol* 1997, 17, 2166–2176.
3. Yuh, C.; Bolouri, H.; Davidson, E.H. Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* 1998, 279, 1896–1902.
4. Kirchhamer, C.V.; Davidson, E.H. Spatial and temporal information processing in the sea urchin embryo: Modular and intramodular organization of the *cylla* gene cis-regulatory system. *Development* 1996, 122, 333–348.
5. Yuh, C.H.; Bolour, H.; Davidson, E.H. Cis-regulatory logic in the endo 16 gene: Switching from a specification to a differentiation mode of control. *Development* 2001, 128, 617–629.

APPENDIX B

TABLE B1

a: $K = 3$, 66 Cases

C	P					Statistical Significance
	4/8	5/8	6/8	7/8	8/8	
0						
Random	0.47 25.00%	0.47 25.0%	0.06 3.125%	0	0	NA
Data	0	0	0	0	0	
1						
Random	0.077 2.34%	0.615 18.75%	0.308 9.375%	0	0	$P < 0.004$ (left shift)
Data	0.055 (1/18)	0.945 (17/18)	0	0	0	
2						
Random	0	0	1.0 9.375%	0	0	NA
Data	0	0	1.0 (4/4)	0	0	
3						
Random	0	0	0	0.944 6.641%	0.056 0.39%	$P < 0.177$ (left shift)
Data	0	0	0	1.0 (44/44)	0	
Margin totals	1	17	4	44		

b: $K = 3$, 66 Cases

P	$C = 0$	$C = 1$	$C = 2$	$C = 3$	Statistical Significance
4/8					
Random	0.914 25.000%	0.086 2.344%	0	0	$P < 0.004$ (right shift)
Data	0	1.0 (1/1)			
5/8					
Random	0.571 25.000%	0.429 18.75%	0	0	$P < 0.001$ (right shift)
Data	0	1.0 (17/17)			
6/8					
Random	0.142 3.125%	0.429 9.375%	0.429 9.375%	0	$P < 0.123$ (right shift)
Data	0	0	1.0 (4/4)		
7/8					
Random	0	0	0	1.0 6.641%	NA
Data	0			1.0 (44/44)	
Margin totals	0	18	4	44	

TABLE B2

a: $K = 4$, 49 Cases

P	$C = 0$	$C = 1$	$C = 2$	$C = 3$	$C = 4$	Statistical Significance
9/16						
Random	0.994 34.7%	0.006 0.195%	0	0	0	$P < 0.006$
Data	0	1.0 (1/1)	0	0	0	(right shift)
11/16						
Random	0.897 11.96%	0.193 1.367%	0	0	0	$P < 0.193$
Data	0	1.0 (1/1)	0	0	0	(right shift)
12/16						
Random	0.665 3.918%	0.332 1.562%	0.012 0.073%	0	0	$P < 0.001$
Data	0	0.4 (2/5)	0.6 (3/5)	0	0	(right shift)
13/16						
Random	0.371 0.635%	0.457 0.781%	0.171 0.293%	0	0	$P < 0.001$
Data	0	0	1.0 (12/12)	0	0	(right shift)
14/16						
Random	0.067 0.024%	0.267 0.098%	0.40 0.146%	0.266 0.098%	0	$P < 0.001$
Data	0	0	0.154 (2/13)	0.846 (11/13)	0	(right shift)
15/16						
Random	0	0	0	0	1.0 0.05%	NA
Data	0	0	0	0	1.0 (17/17)	
Margin						
totals	0	4	17	11	17	

b: $K = 4$, 49 Cases

C	P									Statistical Significance
	8/16	9/16	10/16	11/16	12/16	13/16	14/16	15/16		
1										
Random	0.03 0.012%	0.042 0.195%	0.145 0.684%	0.291 1.367%	0.332 1.562%	0.166 0.781%	0.021 0.098%	0	$P < 0.379$	
Data	0	0.25 (1/4)	0	0.25 (1/4)	0.50 (2/4)	0	0	0	(No shift)	
2										
Random	0	0	0	0	0.134 0.073%	0.571 0.293%	0.286 0.146%	0	$P < 0.310$	
Data	0	0	0	0	0.176 (3/17)	0.706 (12/17)	0.118 (2/17)	0	(left shift)	
3										
Random	0	0	0	0	0	0	1.0 0.098%	0	NA	
Data	0	0	0	0	0	0	1.0 (11/11)	0		
4										
Random	0	0	0	0	0	0	0	1.0 0.05%	NA	
Data	0	0	0	0	0	0	0	1.0 (17/17)		
Margin										
totals		1	0	1	5	12	13	17		

TABLE B3

a: K = 5, 25 Cases

P	C = 0	C = 1	C = 2	C = 3	C = 4	C = 5	Statistical Significance
16/32							
Random	14.01%	0.000000%	0	0	0	0	P < .001
Data	0	1.0 (1/1)	0	0	0	0	(right shift)
17/32							
Random	1.0 26.33%	0.0000125%	0.000000%	0	0	0	P < .001
Data	0	0	1.0 (1/1)	0	0	0	(right shift)
24/32							
Random	0.987 0.485%	0.013 0.006%	<0.001 0.0000025%	0	0	0	P < .001
Data	0.50 (1/1)	0	0.50 (1/1)	0	0	0	(right shift)
25/32							
Random	0.968 0.152%	0.032 0.005%	<0.001 0.000025%	0	0	0	P < .001
Data	0	0	1.0 (1/1)	0	0	0	(right shift)
26/32							
Random	0.928 0.039%	0.071 0.003%	0.001 0.00005%	0	0	0	P < .001
Data	0	0.50 (1/1)	0.50 (1/1)	0	0	0	(right shift)
27/32							
Random	0.787 0.007%	0.200 0.001785%	0.013 0.000115%	0	0	0	P < .001
Data	0	0	1.0 (2/2)	0	0	0	(right shift)
28/32							
Random	0.593 0.001%	0.317 0.000535%	0.087 0.0001475%	0.003 0.000005%	0	0	P < .09
Data	0	0	1.0 (1/1)	0	0	0	(right shift)
29/32							
Random	0.086 0.00007%	0.136 0.00011%	0.741 0.0006%	0.037 0.00003%	0	0	P < .001
Data	0	0	0	1.0 (3/3)	0	0	(right shift)
30/32							
Random	0.143 0.0000025%	0.143 0.0000025%	0.143 0.0000025%	0.428 0.0000075%	0.143 0.0000025%	0	P < .012
Data	0	0	0	0	1.0 (3/3)	0	(right shift)
31/32							
Random	0	0	0	0	0	0.00000149%	NA
Data	0	0	0	0	0	1.0 (9/9)	
Margin totals	1	2	7	3	3	9	

b: K = 5, 23 Cases

C	Random/ Data	P																Statistical Significance	
		16/32	17/32	18/32	19/32	20/32	21/32	22/32	23/32	24/32	25/32	26/32	27/32	28/32	29/32	30/32	31/32		
0	Random	0.161 13.93%	0.305 26.34%	0.254 21.98%	0.187 16.19%	0.118 10%	0.069 6.014%	0.053 2.998%	0.015 1.294%	0.006 0.475%	0.002 0.146%	0.001 0.037%	0	0	0	0	0	0	P < 0.00 (right shift)
	Data	0	0	0	0	0	0	0	0	1.0 (1/1)	0	0	0	0	0	0	0	0	
1	Random	0	0	0	0.00027%	0.0033 0.00084%	0.067 0.002%	0.133 0.004%	0.166 0.005%	0.006%	0.005%	0.003%	0.001785%	0.00053%	0.00011%	<0.0001%	<0.0001%	<0.0001%	P < 0.633 (right shift)
	Data	0	0	0	0	0	0	0	0	0	0	1.0 (1/1)	0	0	0	0	0	0	
2	Random	0	0	0	0	0	0	0	0.006 0.000025%	0.062 0.000025%	0.124 0.00005%	0.286 0.000115%	0.367 0.0001475%	0.149 0.00006%	0.006 0.0000025%	0	0	0	P < 0.036 (left shift)
	Data	0	0	0	0	0	0	0	0	0.143 (1/6)	0.143 (1/6)	0.143 (1/6)	0.28 (2/6)	0.143 (1/6)	0	0	0	0	
3	Random	0	0	0	0	0	0	0	0	0	0	0	0.000005%	0.00003%	0.000075%	0	0	0	P < 1.000
	Data	0	0	0	0	0	0	0	0	0	0	0	0	1.0 (3/3)	0	0	0	0	
4	Random	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000025%	0	0	0	NA
	Data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0 (3/3)	0	0	
5	Random	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000000%	0	NA
	Data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0 (9/9)	0	
Margin Totals		0	0	0	0	0	0	0	2	1	2	2	1	3	3	3	9		