Small-scale copy number variation and large-scale changes in gene expression

Yuriy Mileyko[a,1], Richard I. Joh[b], and Joshua S. Weitz[a,b,1]

[a]School of Biology and [b]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved September 11, 2008 (received for review June 30, 2008)

The expression dynamics of interacting genes depends, in part, on the structure of regulatory networks. Genetic regulatory networks include an overrepresentation of subgraphs commonly known as network motifs. In this article, we demonstrate that gene copy number is an omnipresent parameter that can dramatically modify the dynamical function of network motifs. We consider positive feedback, bistable feedback, and toggle switch motifs and show that variation in gene copy number, on the order of a single or few copies, can lead to multiple orders of magnitude change in gene expression and, in some cases, switches in deterministic control. Further, small changes in gene copy number for a 3-gene motif with successive inhibition (the ''repressilator'') can lead to a qualitative switch in system behavior among oscillatory and equilibrium dynamics. In all cases, the qualitative change in expression is due to the nonlinear nature of transcriptional feedback in which duplicated motifs interact via common pools of transcription factors. We are able to implicitly determine the critical values of copy number which lead to qualitative shifts in system behavior. In some cases, we are able to solve for the sufficient condition for the existence of a bifurcation in terms of kinetic rates of transcription, translation, binding, and degradation. We discuss the relevance of our findings to ongoing efforts to link copy number variation with cell fate determination by viruses, dynamics of synthetic gene circuits, and constraints on evolutionary adaptation.

gene duplication | gene regulation | network motifs | nonlinear dynamics

Copy number variation (CNV) is an important and widespread component of within and between population genetic variation. The copy number of genes and gene fragments varies significantly over physiological to evolutionary time scales with multiple effects on phenotype. For example, CNV can cause statistically significant changes in concentrations of RNA associated with growth rate changes in bacteria (1, 2) and enzyme concentrations associated with nutrient intake in humans (3, 4). The copy number of viral genomes undergoes dynamical changes during multiple infection of bacteria by phages, leading to qualitative changes in gene regulation that may lead to alternative modes of exploitation (5, 6). The duplication of a gene can facilitate subsequent diversification—a mechanism considered to be a dominant cause of phenotypic innovation (7–10). In extreme cases, whole-genome duplications have led to lineage diversification within yeast (11). In humans, large-scale deletions and duplications of chromosomes are known to cause severe genetic disorders (12, 13) and are imputed in the onset of other diseases including cancer (14, 15). Finally, multiple studies have demonstrated that CNV in humans is far more extensive than previously believed, although its impact on phenotype is yet to be fully resolved (16–20).

Despite its ubiquity, CNV has been nearly universally overlooked in quantitative models of gene regulation. In those cases where quantitative models of CNV have been developed, the primary focus has been on changes in the copy number itself, as in the case of plasmid maintenance (21) and dynamics of transposable elements (22). In some instances, gene copy number is integrated into dynamic models of regulation to account for cell-to-cell variability of regulatory elements found on plasmids (23). More commonly, recent studies have attempted to identify statistical relations between CNV and fitness (4), protein interactions (24), or combinations of both (25). To understand the progression from CNV to changes in phenotype to changes in fitness, it seems necessary to carefully examine gene regulation itself. The dynamics of a gene regulatory network depends on network topology, the quantitative nature of feedbacks and interactions between DNA, RNA and proteins, epigenetic modifications of regulatory elements, and the biochemical state of the intracellular and surrounding environment. Additionally, as we argue here, gene regulatory dynamics can also depend sensitively on the copy number of genes and promoters. For example, in synthetically designed networks, small changes in the copy number of gene regulatory modules have been shown to lead to qualitative changes in gene expression (26, 27). In naturally occurring networks, there may be selection pressures on kinetic parameters such that normally occurring levels of copy number are far from or close to the critical threshold that would lead to a dramatic change in gene expression.

In this manuscript, we take a quantitative approach to assess when small changes in copy number can have a dramatic, nonlinear effect on gene expression. We study the effect of changing copy number within a series of small, regulatory networks commonly referred to as "network motifs" (28). These motifs are network subgraphs shown to be building blocks of complex regulatory networks (29). Increasing the number of motifs means that multiple networks are coupled together via a common pool of transcription factors. Changes in the number of promoter sites is directly linked to changes in the rate of regulated recruitment, which in turn leads to changes in translation and other transcriptional feedbacks (30). We demonstrate that small changes in gene copy number within motifs exhibiting positive and/or negative feedback can switch the network from one alternative steady state to another and switch gene expression to and from an oscillatory state. Thus, changes in copy number may act as knobs within a nonlinear dynamical system in much the same way that changes in environmental conditions can drive expression from one steady state to another (29, 31).

Results

CNV and Network Motifs. We systematically analyze the dependency of 4 network motifs, positive feedback, bistable feedback, toggle switch, and the repressilator, on the copy number, $\mathcal{N}$. The method for analyzing each of these motifs is largely the same, and illustrated in Fig. 1. Although $\mathcal{N}$ is not explicitly present in the mathematical models presented in Fig. 1 *A–D*, it factors in

Author contributions: Y.M. and J.S.W. designed research; Y.M., R.I.J., and J.S.W. performed research; and Y.M. and J.S.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: yuriy.mileyko@biology.gatech.edu or jsweitz@gatech.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0806239105/DCSupplemental.

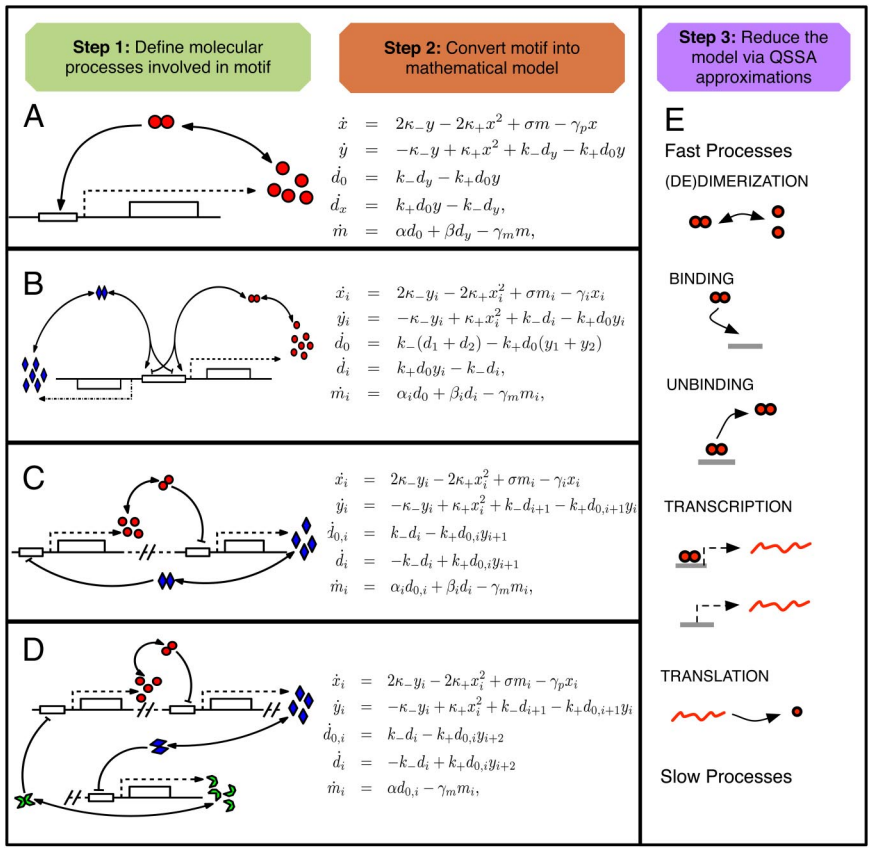

Fig. 1. Schematic of the quantitative approach to linking CNV with gene expression in the case of 4 motifs. (*A*) Positive feedback. (*B*) Bistable feedback. (*C*) Toggle switch. (*D*) Repressilator. (*E*) Catalogs the process we consider in order from fast to slow processes (degradation of mRNA and proteins are not shown). Variables refer to concentrations of protein monomers (x), protein dimers (y), unoccupied and occupied promoters (d_0 and d_1), and mRNA (m). Parameters $\kappa_\pm$ are the dimerization and dedimerization rates, $k_\pm$ are the binding and dissociation rates of the dimers to the promoter site, α is the basal transcription rate, β is the regulated transcription rate, σ is the translation rate, γ_m is the degradation rate of mRNA, γ_p and γ_i denote the degradation rate of proteins. The subscript i denotes the index of the promoter, monomer, or dimer in the motif. (*C*) The notation for the toggle switch is $i = 1,2$, and, to simplify the index notation we assume that $2 + 1 = 1$; that is, the index wraps back to 1 once it becomes >2. The variables and parameters have the same meaning as in the other cases except that $d_{0,i}$ and d_i denote free and occupied promoters of the ith gene. (*D*) For the repressilator system, we again employ our ''wrapping'' notation, this time using the rule $3 + 1 = 1$, so that when the index becomes >3 it wraps back to 1. Note that in the motif schematics in the first column, transcription and translation are denoted using a single dashed line.

implicitly. Note that $\mathcal{N}$ is proportional to the total concentration of promoter sites, d, i.e., $\mathcal{N} = d/C$, where $C \approx 10^{-9}$ M is the conversion factor denoting the molar concentration of a single molecule in the volume of an *Escherichia coli* cell (29). Hence, from the outset, it is evident that copy number can directly modify basic kinetic rates of transcription, binding, and unbinding and indirectly modify others. In the quasi-steady-state approximation (QSSA) version of all models (see *Methods* and supporting information (SI) *Appendix*), the rescaled translation rates are proportional to copy number. Likewise, changes in cell volume may also have global effects in changing gene expression. The estimates of other kinetic parameters are approximate. They are in range with experimental measurements and typical values for dimerization, binding, transcription, translation, and degradation in bacteria and viruses (29) (a list of all parameters used in numerical simulations can be found in *SI Appendix*). Note that we do not include degradation of dimers for the sake of analytical tractability, however, numerical tests including degradation of dimers do not qualitatively change any of the results. In presenting results, we emphasize how the steady-state gene expression changes as a function of $\mathcal{N}$. In so doing, we use the term bifurcation to mean a qualitative change in steady-state gene expression (31, 32). More information on the derivation of the QSSA and bifurcation conditions may be found in *SI Appendix*.

Positive Feedback. The positive feedback motif system consists of a single gene whose protein, when dimerized, enhances its own transcription and subsequent translation (see Fig. 1*A*). For this system, the monomer concentration is x, the basal mRNA transcription rate is α, but when dimers of the protein bind to the promoter site the transcription rate increases to β. The positive feedback motif has been analyzed in various ways (29, 33–35), but the question we are considering here is different. What is the effect of changing the copy number (which in this case is equal to the number of motifs) on the steady-state behavior of the system?

In the positive feedback motif considered here, dimerization precedes regulated recruitment. When $\mathcal{N}$ is below some threshold, there will be insufficient concentration of dimers to enhance transcription. Thus, transcription will occur predominantly at basal levels. However, for copy numbers above some threshold the coupling between motifs will lead to enhanced transcription at activated levels. Hence, the steady-state gene expression will jump nonlinearly as a function of $\mathcal{N}$. The effect of copy number on expression dynamics is depicted in Fig. 2. Analytical predictions of a nonlinear jump in the steady-state gene expression are confirmed by numerical simulation (see Fig. 3*A*, where the threshold is $\mathcal{N} = 3$ for the parameters considered). This discontinuity in steady-state expression can be formally explained as

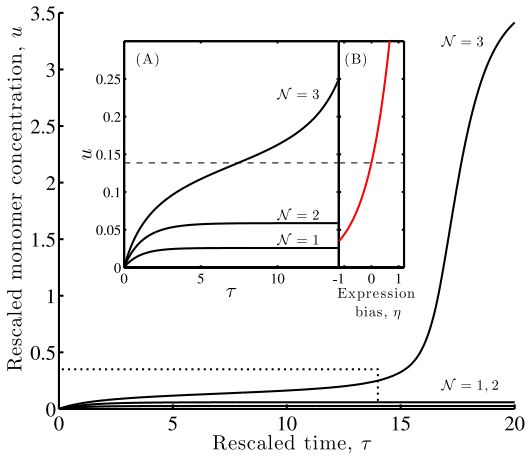

Fig. 2. Dynamics of a positive feedback loop (Eq. **S15** in *SI Appendix*) with rescaled translation rates $\hat{\alpha} = 0.025$, $\hat{\beta} = 1.3$ given copy numbers $\mathcal{N} = 1$, 2, and 3. (*Inset*) (*A*) Blow-up of the rescaled concentration dynamics, $u(\tau)$, for the rescaled time $0 \leq \tau \leq 14$; (*B*) Concentration u vs. expression bias $\eta = \log_{10}(\hat{\beta}u^2/\hat{\alpha})$. The dashed line denotes the value of u such that the rate of basal transcription is equal to the rate of activated transcription, i.e., $\hat{\beta}u^2 = \hat{\alpha}$. Note that expression bias is on a log-scale and denotes the relative contribution of activated vs. basal transcription, e.g., when $\eta = 1$, activated transcription occurs at 10-times the rate of basal transcription.

being due to the presence of a series of saddle-node bifurcations in which an alternative stable state (with high gene expression) appears in the system and eventually dominates the original (low gene expression) state.

We cannot expect that all positive feedback motifs will lead to such a dramatic sensitivity to copy number variation. The condition under which the system exhibits dramatic sensitivity to CNV is very robust, $\beta > 9\alpha$, and is derived in *SI Appendix*. Thus, so long as enhanced transcription is sufficiently greater than the basal transcription rate, changing the copy number may lead to a jump in the steady-state monomer concentration. The value of the copy number at which the dramatic change of expression occurs is a tunable feature, that will have important consequences for how the entire system responds to CNV. For example, the theoretical prediction of the critical threshold, $\mathcal{N}_c$ may be too low ($\mathcal{N}_c < 1$) or too high ($\mathcal{N}_c \gg 1$) to be of any biological relevance. In the former case, the system should always be dominated by activated transcription, and in the later case, the copy number may never reach a level where activated transcription is feasible.

Bistable Feedback. The bistable feedback motif consists of 2 genes whose protein monomer concentrations we denote by x_1 and x_2, respectively (see Fig. 1*B*). These 2 genes have overlapping promoters such that each bound dimer halts the transcription of the other gene and regulates its own. The basal transcription rates are $\alpha_{1,2}$ and the regulated transcription rates are $\beta_{1,2}$. Bistable feedback is characteristic of networks that involve switching between alternative gene expression states (29, 36). Here, we are particularly interested in the case where one gene has higher basal transcription $\alpha_2 > \alpha_1$ but the other gene has higher regulated transcription $\beta_1 > \beta_2$. These conditions are motivated by studies of the genetic switch that regulates the decision between lysis and lysogeny within bacteriophage λ (2, 6, 30, 37, 38). Based on previous analysis of the case where $\beta_2 = 0$ (6), we expect that changes in $\mathcal{N}$ can lead to drastic changes in protein monomer concentrations. However, it is not immediately clear what kind of changes will occur. Will both concentrations jump up or only one of them? Will any of the concentrations drop? We answer these questions by determining the relation

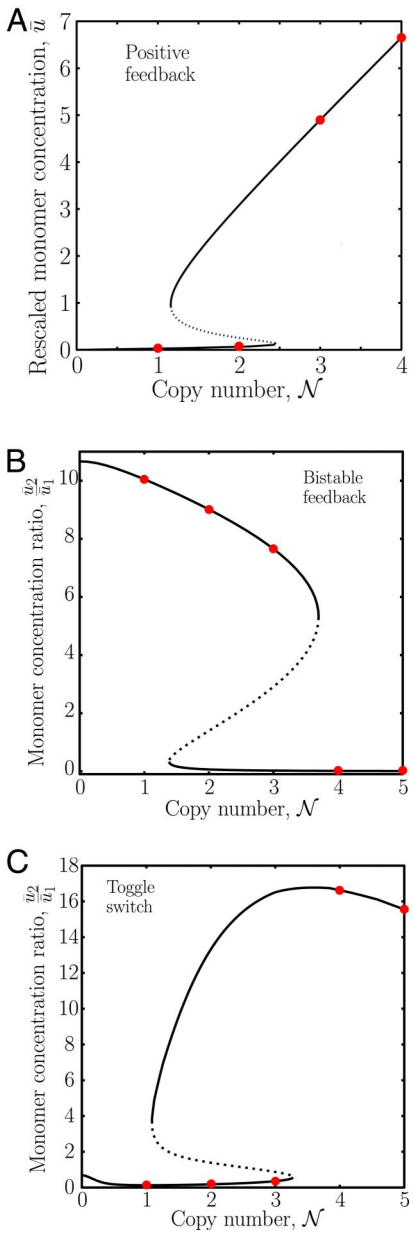

Fig. 3. Steady state gene expression (or ratios) as a function of copy number, $\mathcal{N}$ for 3 motifs. In each case, solid lines are stable equilibria from theory, dashes lines are unstable equilibria from theory, and circles denote numerical simulation of the appropriate QSSA model. (*A*) Positive feedback: Numerically computed steady-state monomer concentration as a function of copy number with $\hat{\alpha} = 0.025$ and $\hat{\beta} = 1.7$. (*B*) Bistable feedback: relation between the copy number, $\mathcal{N}$, and the ratio of the steady-state monomer concentrations. $\hat{\alpha}_1 = 0.063$, $\hat{\beta}_1 = 1.5$, $\hat{\alpha}_2 = 0.67$ and $\hat{\beta}_2 = 0.0083$. (*C*) Toggle switch: relation between the copy number, $\mathcal{N}$, and the ratio of steady-state monomer concentrations. $\hat{\alpha}_1 = 6.3$, $\hat{\beta}_1 = 0.14$, $\hat{\alpha}_2 = 4.3$, and $\hat{\beta}_2 = 0.43$.

between $\mathcal{N}$ and the steady-state monomer concentrations (see *SI Appendix*). For the bistable motif, it is possible that, for small values of $\mathcal{N}$, there is only one steady gene expression state, which is a stable node. As $\mathcal{N}$ increases, the system will enter a bistable region in which alternative stable states corresponding to dominance by either of the 2 genes is possible. Finally, as $\mathcal{N}$ increases even further, the system will have a different steady state, again a stable node. This implies that one gene will dominate at low $\mathcal{N}$, and the other gene will dominate at high $\mathcal{N}$. A graphical representation of the above observations can be obtained by

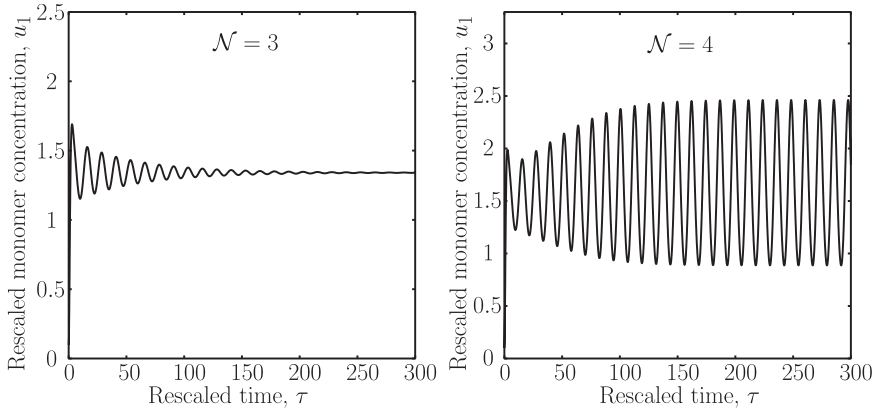

Fig. 4. An example of the onset of oscillations in the repressilator as the copy number changes. Here, $\hat{\alpha} = 1.25$, rescaled degradation of proteins is $\hat{\gamma} = 0.80$, and the critical value for the onset of oscillations as predicted in *SI Appendix* is $\mathcal{N} = 3.43$.

plotting $\mathcal{N}$ along the x axis and the ratio of the 2 concentrations, $\bar{u}_2/\bar{u}_1$, along the y axis. In this case, it is apparent that the ratio is large at the first stable node but drops down significantly at the second one (see Fig. 3*B*). In biological terms, this means that the bistable feature of the network depends on copy number. For sufficiently low or high values of $\mathcal{N}$, the coupled set of motifs will have deterministic outcomes. At low $\mathcal{N}$, gene 2, with the higher basal transcription rate, will dominate, whereas at high $\mathcal{N}$, gene 1, with the higher regulated transcription rate, will dominate.

Analysis of the steady-state behavior demonstrates that the above bifurcations occur when $\beta_1 > 9\alpha_1$ along with a second algebraic condition described in *SI Appendix*. The first condition implies that enhanced transcription must be at least 9 times as great as basal transcription in one of the genes. The second condition is more complicated and involves transcription rates and protein degradation rates—the condition is satisfied for a wide range of parameters (see *SI Appendix*). Thus, we obtain robust conditions for a copy number controlled genetic switch. The switch in abundance of regulatory proteins can lead to radically different phenotypic effects inside a cell or organism. For example, the fate of bacterial cells infected by multiple phages exhibit an acute sensitivity to changes in the multiplicity of infection (5, 6, 39). In addition, the values at which bifurcations occur, the sharpness between alternative gene expression states, and other features are tunable by this copy number dependent effect.

Toggle Switch. The toggle switch motif consists of 2 genes with different promoters such that the product of one gene inhibits transcription of the other (40). Here, we consider the general case in which each gene product dimerizes before binding and then partially inhibits transcription of the other, not completely. The basal transcription rates are $\alpha_{1,2}$ and the repressed transcription rates are $\beta_{1,2}$, where $\beta_{1,2} < \alpha_{1,2}$. The schematics of this system is shown in Fig. 1*C*. The motifs we have analyzed thus far share a common feature: positive feedback loops. Moreover, we saw that the conditions that guarantee existence of essential bifurcations in the 2 motifs would not be satisfied without positive feedback. So, how does the copy number affect behavior of genetic networks with only negative regulation?

For the toggle switch motif, there is a single steady state for low values of the copy number. As $\mathcal{N}$ increases, 2 consecutive saddle-node bifurcations occur, first creating a new stable node and a saddle and then colliding the saddle with the old stable node. We also observe that the steady-state concentration ratio, $\bar{u}_2/\bar{u}_1$, is small before the first bifurcation and large after the second one (see Fig. 3*C*). Biologically speaking this means that the dominant gene in a toggle switch can depend on the copy number of the motif.

Analysis of the steady-state expression shows that existence of the 2 bifurcations and the resulting jump in the ratio of the steady-state concentrations are quite robust with respect to the parameter values (see *SI Appendix*). Thus, even without positive feedback a genetic network may switch to a drastically different state as the copy number changes. Note that the toggle switch of Gardner and Collins (40) corresponds to the case of complete mutual halting of transcription ($\beta_i = 0$). The Gardner and Collins switch exhibits behavior significantly different from the case $\beta_i > 0$ (analysis not shown). Although at small values of $\mathcal{N}$ there is still a single stable node, increasing the copy number will lead to only one saddle-node bifurcation. At larger copy number, 2 gene expression states are possible, corresponding to dominance by either gene respectively. Hence, increasing copy number leads to bistable behavior, in which steady-state outcomes depend on initial conditions and the strength of stochastic effects.

Repressilator. The repressilator motif consists of 3 genes with a circular network structure such that gene 1 represses gene 2, gene 2 represses gene 3, and gene 3 represses gene 1 (see Fig. 1*D*) (41). Unlike the previous motifs, it is known that the repressilator can exhibit sustained oscillations. The corresponding stable limit cycle emerges from a stable node via a supercritical Hopf bifurcation (32, 33, 41). What we are interested in is whether changes in copy number can switch the system between a single steady state and sustained oscillations.

In the repressilator motif, as before, we consider the situation where dimerization precedes binding to promoters. As we show in *SI Appendix*, the above system has a single symmetric steady state where each of the 3 protein concentrations are at identical levels. Analysis of this steady state reveals that a Hopf bifurcation occurs when the copy number passes a critical threshold. A numerical simulation, shown in Fig. 4, confirms the above finding. Biologically speaking, when copy number is low, the circular series of transcriptional feedbacks is insufficient to allow dominance by a single gene in time. Increases in copy number allow a single gene to dominate for a short period, followed by the rise of its inhibitor and so on. Thus, at sufficiently large copy number, the repressilator can exhibit oscillatory behavior. Copy number itself acts as a proxy for the degree of coupling in this system. We should keep in mind, however, that the estimate of the bifurcation point where the switch from steady to oscillatory behavior occurs is, in fact, a rather crude one deriving from the use of the QSSA. Nevertheless, it does demonstrate that changing the copy number can both drive a genetic network to a different state and make it oscillate and that the threshold of oscillations is a tunable quantity.

Discussion

We have demonstrated that copy number is a key control parameter in the expression dynamics of simple network motifs. Changing the copy number can make a network switch to an entirely different equilibrium gene expression state and move it to and from an oscillating regime. Our results stand in contrast to previous assertions that target gene expression is proportional to gene copy number (42, 43). Gene expression can be nonlinearly related to gene copy number because of feedbacks found in even the simplest of network motifs. Such nonlinearities are found even when the balance among gene components is maintained. Although not every small-scale CNV will lead to large-scale changes in gene expression, we have found a set of principles to understand when such a link may occur. In the cases of positive feedback, bistable feedback, and toggle switch motifs, we are able to find general conditions for the presence of qualitative sensitivity to copy number. In more technical terms, we have solved for the sufficient conditions for the existence of saddle-node bifurcations within a set of nonlinear dynamical systems. This has dramatic consequences for systematic analysis of the emergence and maintenance of CNV.

Importantly, our findings hold despite significant variation in parameter values associated with the molecular details of regulated recruitment (see *SI Appendix*). Thus, sensitivity of motifs to CNV may apply to a broad range of cellular contexts. The robustness of genetic regulatory networks to noise (44, 45) and gene duplication (46, 47) have been highlighted. Our findings suggest that there are limits to robustness, particularly with respect to gene duplication. The bifurcation conditions we derived for each motif provide guidance as to the range of kinetic parameters in which network fragility may be expected.

There are many challenges remaining in the study of the link between CNV and phenotypic effects. The networks we have considered are small components of complex gene regulatory networks. It remains an open question whether and to what extent these results scale up to larger, more complex networks (23, 29, 48). For example, how have actual networks evolved with respect to the critical values of copy number which can lead to qualitative shifts in system behavior? Although we have studied the effect of varying the copy number of motifs, it is worthwhile to examine the effects of copy number imbalances in complex motifs. Note that, in this article, we have assumed fully coupled network motifs, whereas the dynamics of intracellular transport of regulatory elements is certainly more complex (38, 49). There are a number of areas where we believe further examination is likely to yield successes in applying the theory presented here: host-phage dynamics, synthetic biology, and evolution via gene duplication. We discuss each of these areas below.

First, in the case of host-phage dynamics, there may be selection pressure favoring sensitivity to copy number, as in the case of temperate viruses whose exploitation strategy depends on the multiplicity of infection (5, 39). In ref. 6, we demonstrated that the number of phage DNA copies inside a bacterial cell has a dynamical effect on the decision making circuit of bacteriophage λ. Hence, coinfecting phages can in principle make collective decisions about a cell's fate. A small number of viruses can direct regulatory machinery toward lysis, whereas the coinfection of a single host by many viruses leads to a latent infection. Different phages differ in their response to coinfection, and so the response to coupling decision modules is likely to be an evolvable feature of phages' life histories. An alternative hypothesis for the link between cell fate and multiple infection is that each injected phage genome experiences a distinct microenvironment (38). Even in such a case, coordination of phage response depends on synchronization of decision modules, although perhaps on different time scales.

Next, of relevance to synthetic biology, CNV may alter dynamics of gene regulatory networks that have been engineered *de novo* or modified to acquire new functions (50). Here, we briefly discuss two experimental studies in which qualitative changes in gene expression were observed in synthetic networks as a consequence of small scale changes in the copy number of gene regulatory components. In one case, an *E. coli* gene regulatory circuit was designed to exhibit both sustained oscillations and toggle switch behavior (26). The copy number of a key activator module in the circuit (controlled by the *glnAp2* promoter) was increased by inserting it closer to the origin of replication. Comparison of gene expression showed a 20% decrease in the degree of damping of oscillations when the activator was located near the origin as opposed to near the terminus. In another case, a reengineered budding yeast pheromone response pathway was designed to exhibit bistable response to pheromone induction (27). Bistability depended sensitively on the number of positive feedback modules inserted into the yeast cells. A minimum of 3 tandem copies of the P_{FUS1J1}−STE*11*ΔN construct was necessary for a sustained positive feedback response, whereas 1 or 2 copies did not lead to a sustained response. Although these are only two examples, they both suggest that experimental studies of the sensitivity of small genetic circuits to CNV may be necessary if regulatory motifs are to be used as reliable building blocks of more complex networks (23).

Finally, gene duplication is considered to be a major factor in the evolution of novel phenotypes. According to the theory of neofunctionalization, duplicated genes are initially redundant, and, on occasion, one of a duplicate pair may diverge to perform some new function (8). In fact, the number of retained gene duplicate pairs is unexpectedly high, with extensive experimental evidence that duplicate genes retain functional compensation over long periods of time (4, 51, 52). Duplicated genes or motifs may not be strictly redundant, even initially. The evolution of network motifs subsequent to duplication may depend on global network context (24). In the current theoretical framework, it is apparent that an extra copy of a gene or motif caused by a duplication event can lead to a shift in expression past some functional threshold. Thus, a new feature could emerge immediately, augmenting or modifying previous function. The possibility that duplicated genes are not redundant is supported by a number of evolutionary studies (25, 53). This is not to say that large-scale gene expression given a gene duplication event must be the norm. To the contrary, if the effect of an extra copy was somehow buffered, then the present dynamical framework of gene regulation would be consistent with a model of evolution via neofunctionalization.

These three biological examples reflect a small fraction of ongoing research by scientists from many disciplines to understand how CNV impacts a broad range of biological phenomena. Although our treatment of gene regulation is closest to the mechanisms of regulated recruitment within bacteria and viruses, we envision that a copy number effect may be present from viruses to higher eukaryotes. This effect may have as its hallmark, a dramatic change in gene expression given a small change in copy number. Even if such a dramatic change represents the exception in gene regulatory networks, when such a change does occur it may have exceptional implications in modifying biological function. Whether in the case of genomic structural variation in humans or bacteriophage infections, variation in copy number is ubiquitous. At minimum, we hope to have provided some first steps toward constructing quantitative models of regulated recruitment that take into account CNV.

Methods

Modeling Framework. There are four steps involved in our quantitative approach to linking copy number variation with gene expression dynamics: (step 1) Define the molecular processes involved in the gene regulatory network motif; (step 2) convert motif into a mathematical language of gene regulatory

dynamics; (step 3) simplify the mathematical model, using a series of quasi-steady-state approximations (QSSAs) (33); (step 4) solve for the steady-state gene expression as a function of copy number. A schematic of steps 1–3 can be found in Fig. 1. For each network motif, we model the following molecular processes: transcription of mRNA, translation of mRNA into proteins, dimerization of monomers into dimers, dedimerization, binding of dimers to promoters upstream of genes, unbinding of dimers from promoters, degradation of mRNA, and degradation of proteins. Each of these processes is assumed to obey simple mass-action kinetics with corresponding kinetic rates such that any particular network motif can easily be transformed, in step 2, into a nonlinear dynamical system (29). In step 3, concentrations within the QSSA model are described in terms of the slowly varying monomer concentration. Importantly, the QSSA model retains the equilibrium values of the original model and is analytically tractable. In step 4, we are able to find the copy number dependence of gene expression via analysis of the QSSA model confirmed by computer simulation (see *SI Appendix*).

ACKNOWLEDGMENTS. We thank Russell Monds for helpful conversations and Anjali Iyer-Pascuzzi, Soojin Yi, and two anonymous referees for providing feedback on the manuscript. This work was supported by Defense Advanced Research Projects Agency Grant HR0011-05-1-0057 (to Princeton University). J.S.W. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

1. Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA oepron copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 66:1328–1333.
2. Stevenson BS, Schmidt TM (2004) Life history implications of rRNA gene copy number in Escherichia coli. *Appl Environ Microbiol* 70:6670–6677.
3. Perry, G. H., *et. al.* (2007) Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39:1256–1260.
4. Deluna A, *et al.* (2008) Exposing the fitness contribution of duplicated genes. *Nat Genet* 40:676–681.
5. Kobiler O, *et al.* (2005) Quantitative kinetics analysis of the bacteriophage λ genetic network. *Proc Natl Acad Sci USA* 102:4470–4475.
6. Weitz JS, Mileyko Y, Joh RI, Voit EO (2008) Collective decision making in bacterial viruses. *Biophys J* 95:2673–2680.
7. Ohno S (1970) *Evolution by Gene Duplication* (Allen and Unwin, London).
8. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
9. Li WH, Yang J, Gu X (2005) Expression divergence between duplicate genes. *Trends Genet* 21:602–607.
10. Korbel JO, *et al.* (2008) The current excitement about copy-number variation: How it relates to gene duplications and protein families. *Curr Opin Struct Biol* 18:366–374.
11. Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* 428:617–624.
12. Conrad B, Antonarakis SE (2007) Gene duplication: A drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet* 8:17–35.
13. Roper RJ, Reeves RH (2006) Understanding the basis for Down Syndrome phenotypes. *PLoS Genet* 2:e50.
14. Pollack JR, *et al.* (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 23:41–46.
15. Pollack JR, *et al.* (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 99:12963–12968.
16. Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: Closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 8:639–646.
17. Redon R, *et al.* (2006) Global variation in copy number variation in the human genome. *Nature* 444:444–454.
18. Stranger BE, *et al.* (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–853.
19. Kidd JM, *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64.
20. Cooper GM, Nickerson DA, Eichler EE (2007) Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 39:S22–S29.
21. Paulsson J, Ehrenberg M (2001) Noise in a minimal regulatory network: Plasmid copy number control. *Q Rev Biophys* 34:1–59.
22. Brookfield JFY, Badge RM (1997) Population genetics models of transposable elements. *Genetica* 100:281–294.
23. Guido NJ, *et al.* (2006) A bottom-up approach to gene regulation. *Nature* 439:856–860.
24. Presser A, Elowitz MB, Kellis M, Kishony R (2008) The evolutionary dynamics of the Saccharomyces cerevisiae protein interaction network after duplication. *Proc Natl Acad Sci USA* 105:950–954.
25. Guan Y, Dunham MJ, Troyanskaya OG (2007) Functional analysis of gene duplications in Saccharomyces cerevisiae. *Genetics* 175:933–943.
26. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell* 113:597–607.
27. Ingolia NT, Murray AW (2007) Positive-feedback loops as a flexible biological module. *Curr Biol* 17:668–677.
28. Milo R, *et al.* (2002) Network motifs: Simple building blocks of complex networks. *Science* 298:824–827.
29. Alon U (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall, London).
30. Ptashne M, Gann A (2002) *Genes and Signals.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
31. Hasty J, McMillen D, Isaacs F, Collins JJ (2001) Computational studies of gene regulatory networks: In numero molecular biology. *Nat Rev Genet* 2:268–279.
32. Strogatz S (1994) *Nonlinear Dynamics and Chaos* (Addison Wesley, Reading, MA).
33. Bennett MR, Volfson D, Tsimring L, Hasty J (2007) Transient dynamics of genetic regulatory networks. *Biophys J* 92:3501–3512.
34. Becskei A, Seraphin B, Serrano L (2001) Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J* 20:2528–2535.
35. Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci USA* 100:7714–7719.
36. Cherry JL, Adler FR (2000) How to make a biological switch. *J Theor Biol* 203:117–133.
37. Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by λ phage repressor. *Proc Natl Acad Sci USA* 79:1129–1133.
38. Rotem E, *et al.* (2008) Bacteriophage infection is targeted to cellular poles. *Mol Microbiol* 5:1107–1116.
39. Kourilsky P (1973) Lysogenization by bacteriophage-lambda 1. Multiple infection and lysogenic response. *Mol Gen Genet* 122:183–195.
40. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* 403:339–342.
41. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transciptional regulators. *Nature* 403:335–338.
42. Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: Biological implications. *Trends Genet* 4:219–226.
43. Griffiths AJF, Wessler SR, Lewontin RC, Carroll SB (2008) *Introduction to Genetic Analysis* (W. H. Freeman, New York).
44. Kollmann M, Lovdok L, Bartholome K, Timmer J, Sourjik V (2005) Design principles of a bacterial signalling network. *Nature* 438:504–507.
45. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307:1965–1969.
46. Wagner A (2000) Robustness against mutations in genetic networks of yeast. *Nat Genet* 24:355–361.
47. Ciliberti S, Martin OC, Wagner A (2007) Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput Biol* 3:e15.
48. Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. *Science* 311:796–800.
49. Slutsky M, Mirny LA (2004) Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys J* 87:4021–4035.
50. Endy D (2005) Foundations for engineering biology. *Nature* 438:449–453.
51. Dean EJ, Davis JC, Davis RW, Petrov DA (2008) Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet* 4:e1000113.
52. Musso G, *et al.* (2008) The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res* 18:1092–1099.
53. Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677–681.