

Review



## Systems Biology as an Integrated Platform for Bioinformatics, Systems Synthetic Biology, and Systems Metabolic Engineering

Bor-Sen Chen \*, Chia-Chou Wu

Laboratory of Control and Systems Biology, Department of Electrical Engineering,  
National Tsing Hua University, HsinChu 30013, Taiwan

\* Author to whom correspondence should be addressed; E-Mail: bschen@ee.nthu.edu.tw;  
Tel.: +886-3-579-1155; Fax: +886-3-571-5971.

Received: 16 April 2013; in revised form: 12 September 2013 / Accepted: 19 September 2013 /

Published: 11 October 2013

---

**Abstract:** Systems biology aims at achieving a system-level understanding of living organisms and applying this knowledge to various fields such as synthetic biology, metabolic engineering, and medicine. System-level understanding of living organisms can be derived from insight into: (i) system structure and the mechanism of biological networks such as gene regulation, protein interactions, signaling, and metabolic pathways; (ii) system dynamics of biological networks, which provides an understanding of stability, robustness, and transduction ability through system identification, and through system analysis methods; (iii) system control methods at different levels of biological networks, which provide an understanding of systematic mechanisms to robustly control system states, minimize malfunctions, and provide potential therapeutic targets in disease treatment; (iv) systematic design methods for the modification and construction of biological networks with desired behaviors, which provide system design principles and system simulations for synthetic biology designs and systems metabolic engineering. This review describes current developments in systems biology, systems synthetic biology, and systems metabolic engineering for engineering and biology researchers. We also discuss challenges and future prospects for systems biology and the concept of systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering.

**Key words:** systems biology; bioinformatics; systems synthetic biology; systems metabolic engineering

---

## 1. Introduction

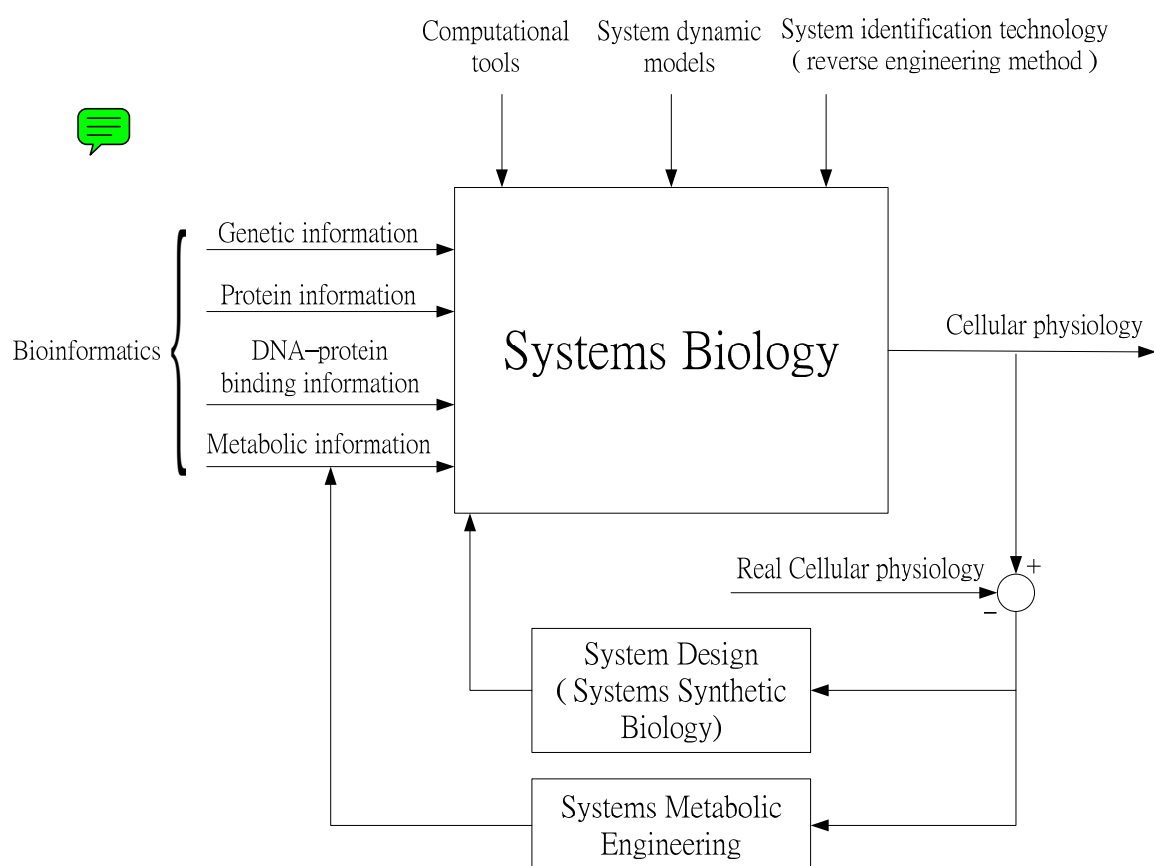
The Human Genome Project and high-throughput experimental methodologies such as microarray chromatin-immunoprecipitation DNA chips (ChIP-chip) have led to the development of biology as an increasingly information-rich science encompassing transcriptomes, proteomes, metabolomes, interactomes, and so forth [1,2]. Some have suggested that systems biology is nothing more than a new name for integrative physiology, which has been practiced for the past 50 years or more. Because of these novel technologies, biologists have been able to collect data at a rate that was unimaginable a decade ago. The context of biology has profoundly changed over the past 20 years. These changes provide a powerful new framework for systems biology that moves it far beyond classical integrative physiology. A systems biology approach implies that every system of any level of biological systems can be analyzed with respect to the system's structure, in particular, in terms of its dynamics, method of control, and method of system design. Systems biology involves genomic, transcriptomic, proteomic, and metabolic investigations from a systematic perspective. As a result, systems biology has become the frontier of modern biological research; large amounts of new omics data cannot be understood without a network or systems viewpoint and without highly sophisticated computational analyses [3–11].

The role of systems biology in modern biological research (Figure 1) requires powerful computational tools to mine large-scale data sets of information on genetics, proteins, DNA–protein binding, metabolism, and so forth. These tools are used to construct dynamic system models for the interpretation of specific mechanisms of some cellular phenotype (or behaviors) from a system (or network) perspective [12–17]. To construct a dynamic system model of biological networks from omics data, system identification technologies (*i.e.*, reverse-engineering schemes) are needed to estimate the parameter values of dynamic models and the order of biological networks [18–44]. Synthetic biology metabolic engineering has recently been developed for designing synthetic genetic networks for the production of specific cellular functions in host cells [45–51]. Based on system models and mechanisms in systems biology, synthetic genetic circuits and metabolic engineering pathways can be designed to investigate cellular behaviors [52–63]. These synthetic biological technologies can be employed to investigate the models and mechanisms of systems biology. Discrepancies between the real behavior of synthetic genetic networks and the desired behavior predicted by the models and mechanisms of systems biology can be fed back to modify the models through methodologies of systems synthetic biology and systems metabolic engineering. Based on the role of systems biology (Figure 1), this review describes current developments in bioinformatics, systems synthetic biology, and systems metabolic engineering. It discusses how systems biology can serve as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering in the future.

Bioinformatics is crucial in genome-wide analyses for understanding cell physiology at different cellular levels (*i.e.*, genome, transcriptome, proteome, and metabolome levels) [12,13]. The various disciplines of bioinformatics provide invaluable information on the global cellular status for systems biology, systems synthetic biology, and systems metabolic engineering, as well as a thorough analysis of the cell. Genomic information in bioinformatics represents the whole genetic makeup of the organism [12–17], and comparative genomic analysis may contribute to systems synthetic biology or systems metabolic biology for targeting and engineering genetic circuits to create desirable cellular phenotypes. Transcriptome profiling uses DNA microarrays to decipher the expression levels of thousands of genes

under various biological conditions [14,15]. The results can be used to select candidate genes for modification based on systematic analysis of regulatory genes in response to genetic variations and environmental changes, or to identify novel factors for the enhancement of heterologous product secretion in metabolic pathways [64–68]. Proteome profiling is also useful in obtaining transcriptome-profiling data at the protein level. The metabolome comprises the entirety of information on metabolites present within and/or outside the cell under specified conditions [61–63]. It is expected to contribute significantly to the understanding of the cell and of synthetic circuit engineering in its metabolic pathways. In this paper, recent advances in the application of bioinformatics to systems synthetic biology and systems metabolic engineering through systems biology are reviewed using specific examples.

**Figure 1.** The role of systems biology as an integrated platform in modern biological research Systems biology integrates information on genetics, proteins, DNA-protein binding, and metabolism with system dynamics modeling and system identification technology to develop models and mechanisms for the interpretation of phenotypes or behaviors of cellular physiology. Since large-scale data sets need to be mined, powerful computational tools are necessary. Based on system models and mechanisms in systems biology, synthetic genetic circuits are designed to investigate specific desired cellular behaviors of cellular physiology. Discrepancies between real and desired cellular behaviors are used as feedback to adjust system models and mechanisms. Systems biology is thus positioned to play the role of integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering.



Although bioinformatic information (e.g., data on genetics, protein binding, and metabolism) is available, several stages of systems biology are required to help us understand via system dynamics modeling the underlying molecular mechanisms of genetic regulatory (GR) networks [18–24], protein–protein interaction (PPI) networks [25–29], and metabolic networks [61–63] under various biological conditions. At the first stage, a putative GR or PPI network is created by large-scale integration of knowledge such as information from publications and databases, and high-throughput data (from data mining or deep curation). Based on this network and dynamic modeling, the actual GR or PPI network of cellular physiology can be identified with system identification methods (reverse-engineering scheme [34]) by using specific microarray gene or protein expression data [18–28]. For example, GR networks (GRNs) have been constructed by dynamic modeling via microarray data for cell cycles [18,23,24], environmental stressors [28,44], photosynthesis [69], aging [34], and cancer [39]. PPI networks have been constructed for cancer [3,9,33,39,70], inflammation [41], biofilm formation [43], and infection by *Candida albicans*. Comparison of PPI networks between healthy and cancer cells can provide network markers for the investigation of the systematic mechanism of cancer [70]. The integration of cellular networks of GRs and PPIs provides deeper insight into actual biological networks and is more predictive than an approach without integration [71]. A systematic and efficient method to integrate different kinds of omics data for the construction of integrated cellular networks via microarray data have been provided based on coupling dynamic models and statistical assessments [44]. This method has been shown to be powerful and flexible, and can be used to construct integrated networks at different cellular levels to investigate cellular machinery under different biological conditions and for different species. Coupling dynamic models of the whole integrated cellular network is very useful for theoretical analyses and for further experiments in the field of network biology and synthetic biology.

In short, synthetic biology is the engineering of biological systems to fulfill a particular purpose. It is possible to build living machines from off-the-shelf genetic devices by employing many of the same strategies that electrical engineers use to manufacture computer chips [47–49]. The main goal of this nascent field is the design and construction of biological systems with desired behaviors [51]. Synthetic biology envisions the redesign of natural biological systems as well as the construction of functional “genetic circuits” by using a set of powerful biotechniques for the automated synthesis of DNA molecules and their assembly into genes and microbial genomes [47]. Synthetic biology is predicted to have important applications in biotechnology, metabolic engineering, and medicine. It may revolutionize how we conceptualize and approach the engineering of biological systems [49]. As illustrated in Figure 1, synthetic genetic circuits can furthermore be employed to confirm network mechanisms derived using systems biology methods, and can be used as feedback for their improvement or revision. Synthetic biology is therefore expected to contribute significantly to a better understanding of the functioning of complex biological systems such as metabolic pathways.

However, the development of synthetic gene networks is still difficult. Most newly created gene networks are nonfunctional because of intrinsic parameter fluctuations, environmental disturbances, and functional variations in the intra- and extracellular context. For this reason, the design method based on dynamic models for robust synthetic gene networks has become an important topic in synthetic biology [52–60,68]. These system-dynamics-based design methods for synthetic biology lead to systems synthetic biology.

Heterologous genes have previously been combined into pathways in metabolic engineering, generating a myriad of non-native biochemical products, including isoprenoids, hydroxyacids, biofuels, polypeptides, and biopolymers [64–67]. Synthetic biologists developed synthetic tools to engineer genetic devices capable of performing complex biological functions such as sensing cell states, counting cellular events, and implementing computational logic. These tools have been applied to the modification and control of metabolic pathways in several organisms. They consist of one or more parts that have been combined to perform a complex function, and provide metabolic engineers with novel ways of exerting cellular control over heterologous production pathways. Some synthetic biological devices with potential relevance in metabolic engineering include orthogonal inducible promoters, light-sensitive promoters, state sensors, spatiotemporal controllers, and logic gates, as well as promoter and ribosome binding site (RBS) libraries [36,37]. Since metabolic engineering seeks to control cellular metabolism and manipulate through heterologous pathways to maximize production of a desired molecule, metabolic engineers need elegant methods for gathering bio-information about cells, their environment, and modulating gene expression in responses [37,61–63]. Hence, devices of synthetic biology promise to be a useful addition to the metabolic engineering toolbox. Some synthetic devices have already been used to increase product titers. However, many remain largely untested in an industrial setting, and the complexity of biology makes their application a feat of engineering [36,37,72]. From the systems biology perspective, continuous work with these devices can help elucidate design rules or aid the development of system dynamics models that facilitate their integration into metabolic industrial processes [36,37,72], and thereby lead to the development of systems metabolic engineering.

Several mathematical techniques based on systems biology have been developed to analyze the systematic properties of complex biological networks. For example, system sensitivity of a biological network in response to various parameter variations has been analyzed to determine the systematic properties that affect the robustness and fragility of a biological network. System sensitivity analysis not only can reveal the robust stability of a biological network against various perturbations, but may also provide information about the controllability of a biological network [7–9]. The system response ability of a biological network is a measure of response to environmental signals or disturbances [28,34]. From the system theory perspective, robustness to intrinsic system variation and the ability to respond to external stimuli are two important and complementary system characteristics in the evaluation of system performance [33,34]. A more biological system that is robust toward a large amount of intrinsic parameter fluctuations is less responsive to environmental disturbances, and vice versa. A systems biology investigation of the aging-related gene network via microarray data found that network robustness increases and network response ability decreases during the aging process [34]. The sensitivity of a biological genetic system to environmental molecular noise is considered as an indication of the noise-filtering ability of the gene network [42]. Systems biology allows the measurement of its system characteristics, as well as the capabilities of the signal transduction pathway [8]. Similarly, flux amplification of the metabolic pathway can be estimated, by using both system dynamics models.

As nonlinear biological networks operate under different conditions of cellular homeostasis and homeodynamics, systems biology studies on complex biological models in the landscape of phenotypes are highly informative. These studies help discover possible equilibrium points (phenotypes) and dynamic behaviors, such as bifurcation, oscillation, robust stability, and phase drift to other equilibrium points (phenotype transition). Bifurcation analysis and phase-plane analysis of nonlinear dynamic

networks can be useful in predicting system behavior of biological networks under intrinsic parameter changes. Through systems biology approach and dynamic modeling [38–42], network robustness and noise filtering ability can be improved via feedback, redundancy, and modular schemes. This is why there are so many feedback loops, redundant genes, and modular structures at different scales of biological networks. A unifying mathematical framework based on nonlinear stochastic dynamic models [73–75] was recently proposed to describe different levels of stochastic biological networks under different parameter fluctuations, genetic variations, and environmental disturbances [29,30,59]. The phenotype robustness criteria of biological networks in systems, evolutionary, ecological, and synthetic biology were investigated from a systematic perspective on the basis of robust stabilization and filtering ability. Network robustness of biological networks can confer intrinsic robustness toward intrinsic parameter fluctuations, genetic robustness for buffering genetic variations, and environmental robustness for resisting environmental disturbances. It was found that if the sum of intrinsic robustness, genetic robustness, and environmental robustness is less than or equal to the network robustness, then the phenotype is robust in different levels of biological networks in systems, evolutionary, ecological, synthetic, and metabolic biology. These phenotype criteria at different levels of biological networks are useful for the design of synthetic and metabolic systems. A systems biology approach based on dynamic models can clearly provide not only a systematic insight into behaviors at different levels of biological networks, but also a design platform to improve system robustness, filtering ability, and transduction ability of synthetic and metabolic system networks, which are discussed in detail in the following sections.

## 2. Systems Biology Approach to GRNs and PPI Networks via Bioinformatics Methodology

### 2.1. Construction of GRNs via Microarray Data

The construction of a GRN using a systems biology approach can be divided into two steps. The first step is construction of a potential (or putative) GRN from two high-throughput data sets, namely, a ChIp network and a mutant network. The ChIp data are based on experimental data published by Harbison *et al.* [14], who also used a genomic tiling array to identify the genomic region bound by transcription factors (TFs). The mutant data are the gene expression data matrix published by Hughes *et al.* [76] with different gene deletion mutants. In general, the potential GRN is suitable for all possible biological conditions. Therefore, the GRN for a specific biological condition needs to be confirmed using microarray gene expression data of the specific biological condition; that is, the real GRN is derived by pruning the potential GRN with specific microarray data. Let  $x_i(t)$  denote the gene expression of the  $i$ -th target gene in the potential GRN. The dynamic equation for the regulation of the  $i$ -th gene is then modeled as [21,22]

$$x_i(t+1) = x_i(t) + \sum_{j=1}^N a_{ij}x_j - \lambda_i x_i(t) + k_i + n_i(t) \quad (2.1)$$

where  $x_i(t)$  represents the mRNA expression level of target gene  $i$  at time  $t$ ,  $x_j(t)$  represents the regulation function of the  $j$ -th TF binding to the target gene  $i$ ,  $a_{ij}$  denotes the regulatory ability from the  $j$ -th regulatory gene to the  $i$ -th target gene (the positive sign indicates activation and the negative sign indicates repression),  $\lambda_i$  indicates the degradation effect of the present time point on the next time point  $t+1$ ,  $k_i$  represents the basal level, and  $n_i(t)$  is the stochastic noise due to model uncertainty and fluctuation

of microarray data of the target gene. Expression of the  $i$ -th gene in (2.1) can be represented by the following regression equation [77]

$$x_i(t+1) = [x_1(t) \quad \cdots \quad x_N(t) \quad 1] \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_{i-1} \\ (1-\lambda_i) \\ a_{i+1} \\ \vdots \\ a_N \\ k_i \end{bmatrix} + n_i(t) \quad (2.2)$$

$$= \phi_i(t)\theta_i + n_i(t)$$

where  $\phi_i$  denotes the regression vector, which can be obtained from microarray data.  $\theta_i$  is the regulatory parameter vector of target gene  $i$ , which is to be estimated.

The regulatory parameter can be estimated from microarray data of the corresponding gene through the least-squares or maximum-likelihood parameter estimation [77] (known as the reverse-engineering method). After regulator parameters in  $\theta_i$  are estimated, the system order (*i.e.*, the number of regulatory genes) is determined by model comparison (e.g., Akaike's information criterion (AIC)), and the  $P$ -value statistical method is employed to determine the significant regulatory genes for target gene  $i$ . This is done by pruning false-positive regulations in the potential GRN. That is, some  $a_{ij}$  is pruned because of false positive deletion. Based on the dynamic model in (2.1), the true GRN can then be constructed one target gene at a time through microarray data. Using similar methods, GRNs for yeast cell cycles [18,23,24], cancer cell cycles [78], stress response [44], and inflammation [41] can be constructed.

## 2.2. Construction of PPI Networks

The construction of PPI network via a systems biology approach is also a two-step process. The first step is constructing a potential PPI network via data mining from literature and databases such as BioGRID, SGD, and GO [16,17]. As this is only a candidate network based on many biological conditions, the second step is pruning false positive PPIs by using a protein expression microarray of a specific biological condition. For a target protein  $i$  in the potential PPI network, the dynamic model of protein activity is [19,20]

$$y_i(t+1) = y_i(t) + \sum_{j=1}^M b_{ij}y_i(t)y_j(t) - \beta_i y_i(t) + h_i + v_i(t) \quad (2.3)$$

where  $y_i(t)$  represents the protein activity level of the target protein  $i$  at time  $t$ ,  $b_{ij}$  denotes the interaction ability of the  $j$ -th interactive protein with the  $i$ -th protein,  $y_j(t)$  represents the activity level of the  $j$ -th protein interacting with the target protein  $i$ ,  $\beta_i$  denotes the degradation effect of the protein,  $h_i$  represents the basal activity level, and  $v_i(t)$  is the stochastic noise. The rate of PPI is proportional to the product of the concentrations of both proteins [7], that is, it is proportional to the probability of molecular collisions between them. The physical interpretation of Equation (2.3) is that the activity of target protein  $i$  at time  $t+1$  is the combination of present protein activity, regulatory interactions with  $M$  interactive proteins,

levels of basal protein from other sources and  $M$  interactive proteins in the cell, and stochastic noise, less the protein degradation of the present state. The PPI dynamic equation of target protein  $i$  in the potential PPI network can be represented by the following regression equation [77]:

$$y_i(t+1) = [y_i y_1 \quad \cdots \quad y_i y_{i-1} \quad y_i \quad \cdots \quad y_i y_M \quad 1] \cdot \begin{bmatrix} b_{i1} \\ \vdots \\ b_{ii-1} \\ 1 - \beta_i \\ \vdots \\ b_{iM} \\ h_i \end{bmatrix} + v_i \quad (2.4)$$

$$= \Phi_i(t) \theta_i + v_i$$

The interaction parameter  $\theta_i$  can be estimated from protein profile microarray data by least-squares or maximum-likelihood parameter estimation [77] (if protein profile microarray data are unavailable, ten mRNA microarray data could be used to replace them, with some modification [19,20]). By using AIC to prune false positive interactions, the real PPI network can then be constructed one target protein at a time by following the above two-step procedure. Some dynamic metabolic pathways [20] and PPI networks of cancer [39] and inflammation [41] have recently been constructed by using the microarray data and AIC method. Comparison of PPI networks between healthy and cancer cells can provide network-based biomarkers for molecular investigation and diagnosis of cancer [70].

### 2.3. Construction of Integrated GRN and PPI Cellular Networks

Living organisms have evolved complex mechanisms to respond to changes in environmental conditions. This is the case even in unicellular organisms like the yeast *Saccharomyces cerevisiae*. Such environmental changes, commonly termed as “stress”, are harmful or even lethal to the survival of the cells, especially of microorganisms whose environment is very changeable. Cellular responses to stresses require complex cellular networks of sensing and signal transduction pathways, as well as metabolic pathways to adapt cell growth and proliferation for adjustments of gene expression programs, metabolic activities, and other features of the cell (see Figure 2).

These regulatory systems and signal and metabolic pathways can therefore be suitably described by an integrated cellular network of transcription regulation and PPIs. The flowchart for constructing the integrated cellular network is shown in Figure S1. The same two steps already described apply (construction of a potential cellular network and pruning of false positives via specific microarrays of gene expression and protein expression data). Several kinds of omics data and database information were integrated, including data on microarray gene expression and protein expression, regulatory associations between TFs and genes, and PPI. From these data, the potential integrated GRN and PPI networks are retrieved. In the second step, based on the dynamic model of integrated gene/protein interactions in the potential cellular network, we find that [71]



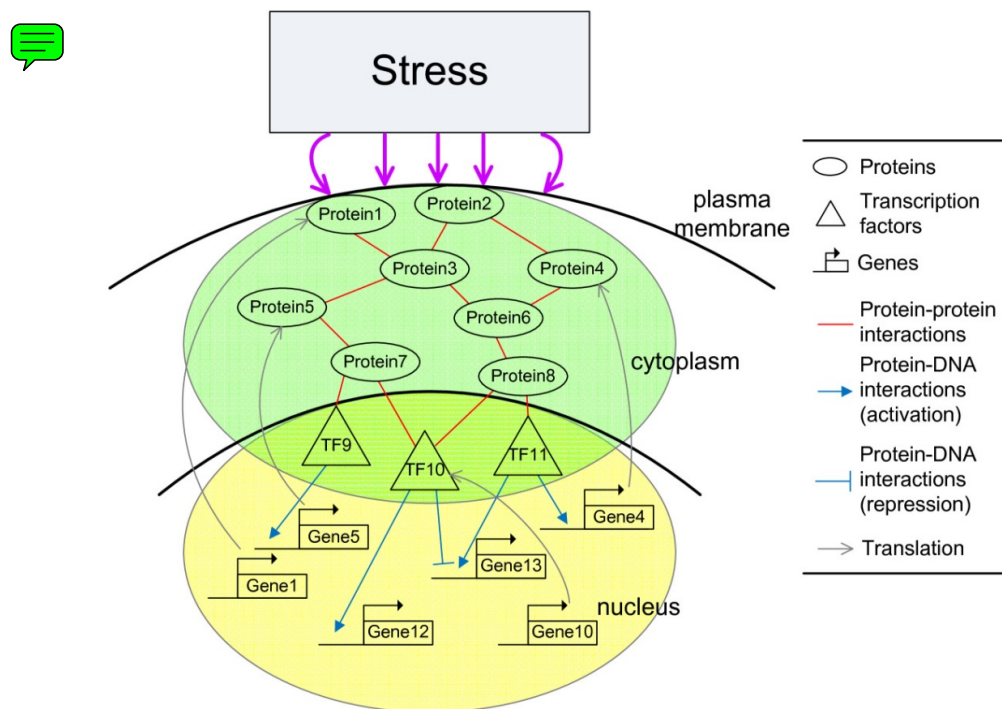
$$\begin{aligned}
 x_i(t+1) &= x_i(t) + \sum_{j=1}^N a_{ij} z_j(t) - \lambda_i x_i(t) + k_i + n_i(t) \\
 y_n(t+1) &= y_n(t) + \sum_{m=1}^M b_{nm} y_n(t) y_m(t) + \alpha_n x_n(t) - \beta_n y_n(t) + h_n + v_n(t)
 \end{aligned}
 \tag{2.5}$$

where the regulation function  $z_j(t)$  is modeled as the sigmoid regulation function of  $y_j(t)$ , that is, for the protein activity profiles of transcription factor  $j$  [71],

$$z_j(t) = f_j(y_j(t)) = \frac{1}{1 + \exp\left\{-\left(y_j(t) - u_j\right)/\sigma_j\right\}}
 \tag{2.6}$$

where  $f_j(y_j(t))$  denotes the sigmoid function,  $u_j$  and  $\sigma_j$  represent the mean and deviation of protein activity level of TF  $j$ , and  $\alpha_n$  denotes the translation effect from mRNA  $x_n(t)$  to protein  $y_n(t)$ .

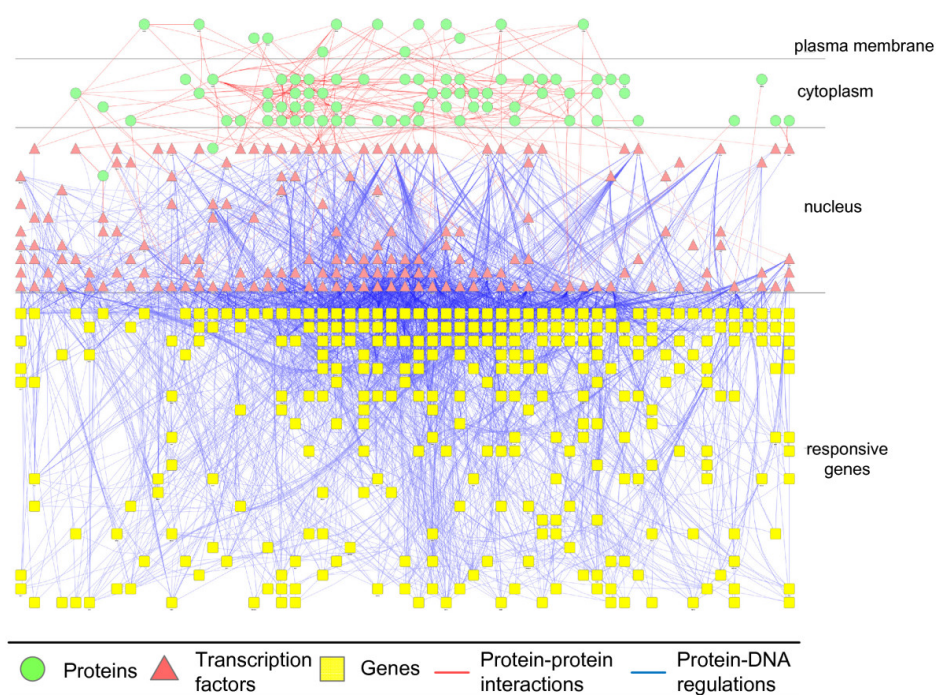
**Figure 2.** Schematic diagram of the integrated cellular network. The integrated cellular network consists of two subnetworks. The signaling regulatory pathway (green) contains protein–protein interaction (PPIs). The gene regulatory network (yellow) contains transcription regulations. The transcription factors serve as the interface between the two subnetworks.



Interaction and coupling among genes and proteins in the integrated cellular network in (2.5) are described as follows. If some TFs  $y_j(t)$  at the end of the signal regulatory pathway regulate their target genes through the regulation function  $z_j(t) = f_j(y_j(t))$  in (2.6), then the regulatory genes influence their corresponding proteins in signal and metabolic pathways through translation effect  $\alpha_n x_n(t)$ . The interplay between genes and proteins can be seen from their coupling dynamic equations in (2.5) and **Figure 2**. Here, the TFs serve as the interface between the signaling regulatory pathway and gene regulatory network. In other words, the interplay between transcription regulation and PPIs constitutes the framework of the integrated cellular network.

Based on the dynamic coupling equation in (2.5), the potential integrated cellular network can be linked through the regulatory parameter  $a_{ij}$  between genes and their possible regulatory TFs and through the translation parameter  $\alpha_n$  for gene expression to protein expression. The potential signaling or metabolic pathways can be linked through the interaction parameter  $b_{nm}$  between possible interaction proteins. Since omics data on the potential gene regulatory network and potential signaling or metabolic pathway only indicate possible TF-gene regulation and protein interactions, they should be confirmed using microarray data of gene and protein expressions. In particular, values of  $a_{ij}$  and  $b_{nm}$  in (2.5) should be identified and validated by least-squares estimation via microarray data in a specific biological condition or phenotype [71]. Significant regulations and interactions between genes and proteins were detected using model selection methods such as AIC and statistical assessments like such as Student's t-test [77]. Based on the interface between gene regulatory and signal/metabolic networks (*i.e.*, transcription factors) in a specific biological condition, the two networks are coupled together to form the integrated cellular network. The integrated cellular network for *S. cerevisiae* under hyperosmotic stress is shown in Figure 3. After the construction of GRN and PPI network from microarray data and bioinformatic method, various system characteristics of the biological network are estimated or measured using systems biology methods in the following sections so that these systematic methodologies can be applied the systematic design of systems synthetic biology and systems metabolic engineering.

**Figure 3.** The *S. cerevisiae* integrated cellular network under hyperosmotic stress. By following the schematic diagram of an integrated genetic regulatory network (GRN) and PPI cellular network in Figure S1, the integrated cellular network of signaling regulatory pathway and GRN for hyperosmotic stress in *S. cerevisiae* is identified by dynamic modeling and data mining. Receptor proteins in the plasma membrane, signal regulatory pathways in the cytoplasm, and transcription factors and GRNs in the nucleus are used to construct an integrated cellular network for *S. cerevisiae* under hyperosmotic stress.



#### 2.4. Network Robustness and Sensitivity Estimation via Microarray Data Using a Systems Biology Approach

After the GRNs or PPI networks or their integrated cellular networks are constructed by dynamic modeling using microarray data, some characteristics of these biological networks need to be estimated to gain insight into their systematic mechanisms. “Robustness”, the network ability to maintain systematic performance under intrinsic perturbations, and “response ability”, the network sensitivity to respond to external stimuli or to transduce them to downstream regulators, are two important complementary or antagonistic system characteristics that must be considered when discussing biological network performance. However, these systematic features cannot be measured directly for all network components in the experimental procedure. Even biological processes such as development, differentiation, tumorigenesis, and aging are increasingly being described in terms of temporal ordering of highly orchestrated transcriptional programs [33,34]. The term robustness is encountered widely in diverse scientific fields, from engineering and control theory [73–75] to dynamic systems [79] and biology [80]. It is important to note that robustness describes a relative property, not an absolute one, as no system can maintain stability in all functions when it is perturbed. In other words, robustness is not immutable.

Let the GRN of interest consist of  $N$  genes. After parameter identification via microarray data, the GRN in (2.1) can be represented by the following linear discrete-time dynamic system [33,34]

$$x(t+1) = \hat{A}x(t) + \hat{K} + n(t) \quad (2.7)$$

$$\hat{A} = \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} & \cdots & \hat{a}_{1N} \\ \hat{a}_{21} & \hat{a}_{22} & & \\ \vdots & & \ddots & \\ \hat{a}_{N1} & & & \hat{a}_{NN} \end{bmatrix}, \quad \hat{K} = \begin{bmatrix} \hat{k}_1 \\ \hat{k}_2 \\ \vdots \\ \hat{k}_N \end{bmatrix}$$

where the state vector  $x(t) = [x_1(t) \cdots x_N(t)]^T$  stands for the discrete-time mRNA expression levels of total  $N$  genes at times  $t = 1, 2, \dots, K$ . The system matrix  $\hat{A}$  denotes gene interactions in the gene network estimated by microarray data, that is,  $\hat{a}_{ij}$  denotes the estimated interaction of gene  $j$  with gene  $i$ . If  $i \neq j$ , then  $\hat{k}_i$  denotes the estimated basal level of the  $i$ -th gene.  $n(t)$  denotes the model residual and measurement noise. The steady state  $x_s$  of  $x(t)$  is obtained as  $t \rightarrow \infty$ :

$$x_s = \hat{A}x_s + \hat{K} \quad \text{or} \quad x_s = (1 - \hat{A})^{-1} \hat{K} \quad (2.8)$$

To simplify analysis of “robustness” of the steady state (phenotype), the origin of the dynamic system is shifted to the steady state  $x_s$ , that is,  $x(t) = \tilde{x}(t) + x_s$ . This shift allows the following shifted dynamic system to be achieved by subtracting Equation (2.7) from Equation (2.8) [34]:

$$\tilde{x}(t+1) = \hat{A}\tilde{x}(t) \quad (2.9)$$

Therefore, the robustness of the phenotype (*i.e.*, the steady state) of the gene network becomes the robustness of the shifted gene network in Equation (2.9) at the origin  $\tilde{x}(t) \equiv 0$ . In the following, its network robustness (tolerance to intrinsic network perturbation) is tested. Suppose the linear GRN suffers from intrinsic molecular perturbations mainly due to process noise, thermal fluctuation, or genetic mutations. The interactive matrix  $\hat{A}$  is then perturbed as  $\hat{A}(1 + \eta)$ , where  $\eta$  denotes the ratio of

intrinsic perturbation. That is, the corresponding additional system perturbation is  $\Delta\hat{A} = \eta\hat{A}$  [80]. A GRN with intrinsic perturbation can then be represented by

$$\tilde{x}(t+1) = \hat{A}(1+\eta)\tilde{x}(t) = (\hat{A} + \Delta\hat{A})\tilde{x}(t) \quad (2.10)$$

Because quadratic stability with the Lyapunov energy-like function  $V(\tilde{x}) = \tilde{x}^T(t)P\tilde{x}(t) > 0$ , with  $V(0) = 0$  for a positive symmetric matrix,  $P = P^T > 0$ . The perturbative GRN in Equation (2.10) is robustly stable if  $\Delta V(\tilde{x}) = V(\tilde{x}(t+1)) - V(\tilde{x}(t)) \leq 0$ , *i.e.*, the energy of the gene network is not increased by intrinsic perturbations [79]. Based on this idea of robust stability, if the following inequality has a positive definite solution  $P = P^T > 0$  [34],

$$\left[ \hat{A}(1+\eta) \right]^T P \left[ \hat{A}(1+\eta) \right] \leq P \quad (2.11)$$

then the perturbative gene network is robustly stable or the phenotype of the gene network is maintained under parametric perturbation ratio  $\eta$ . The network robustness  $\eta^\circ$  of the perturbative gene network is equal to the tolerance of the largest perturbation that does not violate network stability or still maintains the phenotype of the perturbative gene network:

$$\begin{aligned} \eta^\circ &= \max \eta \\ \text{subject to } P &> 0, \end{aligned} \quad (2.12)$$

That is, the network robustness is the maximum perturbation ratio  $\eta^\circ$  tolerated by the GRN such that network stability (or phenotype) is still maintained. The constrained optimization problem in (2.12) can be solved by increasing  $\eta$  until no positive solution  $P$  exists in Equation (2.11) up to the highest value possible without violating the robust stability in Equation (2.11). A positive definite solution  $P > 0$  in Equation (2.11) can be easily obtained by using the linear matrix inequality (LMI) Toolbox of Matlab. This network robustness method has been used to measure the relative network robustness of multiple loops of a gene regulatory network associated with aging-related pathophysiological phenotypes by using previously reported microarray data. It profiles the effects of aging on gene expression in the thymus, spinal cord, and eye tissues in mice [34]. This aging-related GRN includes 16 genes (Figure 4). The relative robustness  $\eta^\circ$  toward normalized perturbation for GRNs of young, aged, and calorie-restriction (CR) groups is shown in Table 1.

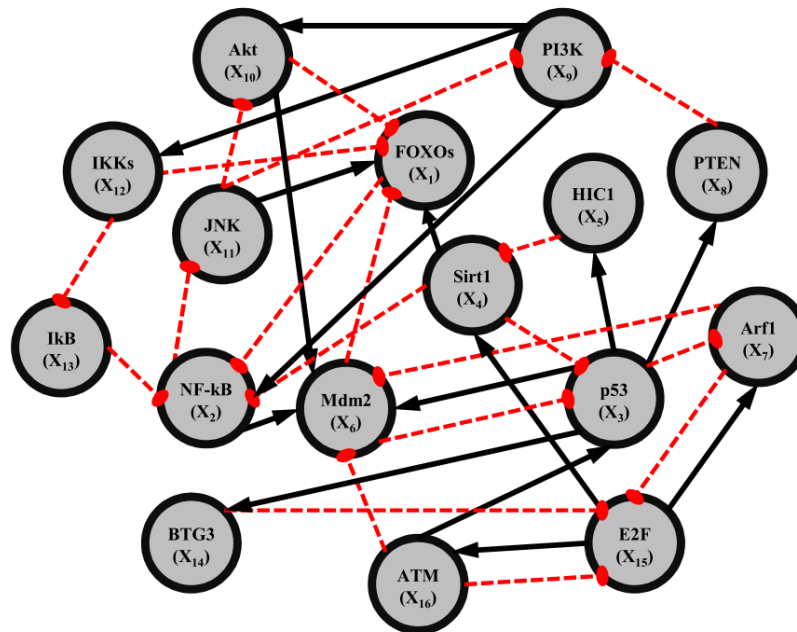
After the network robustness of aging-related GRNs is measured, the response abilities of genes to external stimuli are examined. Assuming that the GRN responds to external stimuli  $U(t)$ , including upstream regulatory signals and external signals outside the network (e.g., hormones, carcinogens, oxidative stress, or ambient biomedical molecules) that could produce an effect on the gene network, the dynamic Equation (2.9) could be modified as follows:

$$\begin{aligned} \tilde{x}(t+1) &= \hat{A}\tilde{x}(t) + U(t) \\ Y(t) &= C\tilde{x}(t) \end{aligned} \quad (2.13)$$

where  $U(t) = [u_1(t) \cdots u_N(t)]^T$  represents external stimuli and  $Y(t)$  denotes the output signal response of specific genes of interest. For example, if the output signal response of the gene  $i$  to external stimuli  $U(t)$  is to be analyzed, then  $C = \text{diag}(0, 0, \dots, 0, 1, 0, 0, \dots, 0)$ , *i.e.*, all elements of  $C$  are zero except for a single element at the  $i$ -th diagonal component. For example,  $C = \text{diag}(0, 0, 1, 0, \dots, 0)$  describes the gene

response of the P53 gene. If the network response ability of the entire GRN to external stimuli is analyzed, then  $C = I$  (identity matrix).

**Figure 4.** Multiple loops of a gene regulatory network associated with aging-related pathophysiological phenotypes. This aging-related gene regulatory network includes 16 genes: *FOXO*, *NF-kB*, *P53*, *SIRT1*, *HIC1*, *Mdm2*, *Arf1*, *PTEN*, *PI3K*, *Akt*, *JNK*, *IKK*, *IκB*, *BTG3*, *E3F1*, and *ATM*. Dashed red lines and black arrows indicate negative and positive parameters of regulated interaction, respectively.



The effect of input signals  $U(t)$  on output signal  $Y(t)$  is less than or equal to a positive value  $\delta$ , if the following inequality holds [79]

$$\frac{\sum_{t=0}^{\infty} Y^T(t)Y(t)}{\sum_{t=0}^{\infty} U^T(t)U(t)} \leq \delta^2 \quad (2.14)$$

$\delta$  denotes the upper bound of the effect of  $U(t)$  on  $Y(t)$  for all bounded input signals  $U(t)$ .  $\delta^\circ$ , the smallest upper bound of  $\delta$  in (2.14), is called the “gene response ability (or sensitivity)” of the GRN.  $\delta^\circ$  permits us to obtain a more systematic insight into the ability of gene response to external stimuli for individual genes or the entire GRN. From the system gain result in [79], the network response of the dynamic GRN has an upper bound  $\delta$  in (2.14), if there exists a positive definite  $P = P^T > 0$  solution to the following LMI [34]:

$$\begin{bmatrix} \hat{A}^T P \hat{A} + C^T C - P & \hat{A}^T P \\ P \hat{A}^T & P - \delta^2 I \end{bmatrix} \leq 0 \quad (2.15)$$

That is, if the above LMI holds for some  $P > 0$ , then the effect of  $U(t)$  on  $Y(t)$  must be less than or equal to  $\delta$ , and Equation (2.14) holds. The response ability (minimal  $\delta$ ) of the GRN to external stimuli can be obtained by solving the following constrained optimization problem:

$$\delta^0 = \min_{P>0} \delta \quad (2.16)$$

subject to (2.15)

This can be solved by decreasing the upper bound  $\delta$  in Equation (2.14) until no  $P > 0$  exists in Equation (2.15) by using Matlab LMI Toolbox. By following Equation (2.16), the network response abilities of aging-related GRNs in Figure 4 can be measured and compared at different life stages. Results are shown in Table 1.

Table 1 shows that the “aged” group has higher network robustness but lower network response ability. In order to tolerate the intrinsic parameter fluctuations accumulated due to genetic mutation, the elderly GRN becomes more robust than the young GRN. Robust GRNs are less responsive to external stimuli; consequently, the network protective ability against external stimuli decreases in the older gene networks. However, a more robust GRN may harbor more accumulated genetic mutations, which through random drift might provide more evolutionary paths to other phenotypes of gene network and thus lead to some aging-related chronic diseases like cancers, metabolic disorders, and arthritis. Table 1 also shows that the “young” network groups are less robust, with greater response to external stimuli. These GRNs are therefore also less robust to intrinsic perturbations and elicit a strong response toward external stimuli. This might imply that some gene expressions such networks could be easily reprogrammed to mediate downstream genes or regulators for further reactions. It also allows for modulation of gene expression in response to external stimuli, such as exposure to oxidative stress, carcinogens, and pro-inflammation molecules. These observations suggest that the gene regulatory network at the earlier stage of life or under conditions of CR may opportunely have protective adaptations to maintain intact regulatory structures and homeostasis of cellular functions. The purpose of the above analysis is to gain a greater understanding of the systematic protective and/or defensive mechanisms inherent to aging-related GRNs. The proposed robustness measurement methods may be used for future studies of GRNs involved in various biological processes and may have potential applications in gene therapy and drug target selection. It was also found that the network robustness of cancer cells is higher than that of normal cells and that the reverse is true for network response ability [33,34].

**Table 1.** The network robustness ( $\eta^\circ$ ) and network response ability ( $\delta^\circ$ ) of a gene regulatory network with 16 aging-related genes (Figure 4) across different tissues at young, aged, and calorie-restrictive (CR) stages.

Tissue	Young	Aged	CR
Thymas			
$\eta^\circ$	0.2310	0.3750	0.2050
$\delta^\circ$	1.1653	0.9463	1.2279
Spinal Cord			
$\eta^\circ$	0.2410	0.6270	0.1400
$\delta^\circ$	1.1630	0.9367	1.2645
Eye			
$\eta^\circ$	0.1600	0.2910	
$\delta^\circ$	1.2376	0.8798	



## 2.5. Network-Based Biomarker Determination via Sample Microarray Data Using a Systems Biology Approach

Biomarkers are used either in diagnostic evaluation to determine the health of a patient with or without a disease, or as a prognostic indicator to determine a patient's prognosis. Present biomarker identification methods, which strictly use gene expression profiles, cannot show how the different genes within the biomarker gene set are related to each other; that is, biomarkers are not identified from a systems biology perspective. Furthermore, the gene lists obtained for similarly diagnosed patients by different research groups differ widely and share few common genes. Here, a systems biology approach is introduced for the integration of microarray data and PPI information to develop a network-based biomarker for systematic investigation into the network mechanism of lung carcinogenesis and the diagnosis of lung cancer [70]. The network-based biomarker consists of two protein association networks constructed from cancer and noncancer samples. The proposed method can be widely applied to determining network-based biomarkers for other diseases. The overall flowchart of the proposed network-based biomarker approach is shown in Figure S2 [71]. First, PPI information and microarray sample data for smokers with and without cancer are used to construct potential PPI networks for cancer and noncancer samples. Since data for both samples are limited, the number of proteins used in potential PPI network construction is also restricted: to avoid overfitting in network construction, the maximum degree of proteins in the potential network are less than the number of samples, imposing a limit on the size of the potential network.

Using a simple regression model, the potential PPI network is then further validated by the sample microarray data to highlight independent protein associations of both samples relative to their respective data sets. For a target protein  $i$  in the potential PPI network, the protein is described using the following protein association model [71]:

$$y_i(n) = \sum_{k=1}^{N_i} \alpha_{ik} y_k(n) + \varepsilon_i(n) \quad (2.17)$$

where  $y_i(n)$  represents the gene expression level of the target protein  $i$  for the sample  $n$ , and  $\alpha_{ik}$  denotes the association ability between the target protein  $y_i(n)$  and protein  $y_k(n)$  for sample  $n$ .  $N_i$  represents the number of proteins interacting with the target protein  $i$ ; it can be obtained from the rough PPI network.  $\varepsilon_i(n)$  denotes stochastic noise associated with other factors or model uncertainty. Equation (2.17) states that, biologically, the expression level of the target protein  $i$  is associated with the expression levels of interacting proteins.

The associated interaction parameter  $\alpha_{ik}$  in Equation (2.17) is identified through maximum likelihood estimation [77] on microarray data. AIC and Student's t-test were employed for model order selection and for tests on the statistical significance of protein associations. Based on  $\alpha_{ik}$ , two matrices are established to represent the cancer protein association network (CPAN) and the non-cancer protein association network (NPAN) as follows [77]

$$C = \begin{bmatrix} \alpha_{11,C} & \alpha_{12,C} & \cdots & \alpha_{1K,C} \\ \alpha_{21,C} & \alpha_{22,C} & \cdots & \alpha_{2K,C} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K1,C} & \alpha_{K2,C} & \cdots & \alpha_{KK,C} \end{bmatrix}, \quad N = \begin{bmatrix} \alpha_{11,N} & \alpha_{12,N} & \cdots & \alpha_{1K,N} \\ \alpha_{21,N} & \alpha_{22,N} & \cdots & \alpha_{2K,N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K1,N} & \alpha_{K2,N} & \cdots & \alpha_{KK,N} \end{bmatrix} \quad (2.18)$$

where  $\alpha_{ij,C}$  and  $\alpha_{ij,N}$  indicate the quantitative protein association ability between protein  $i$  and protein  $j$  for CPAN and NPAN, respectively, and  $K$  is the number of proteins in the protein association network. The resulting CPAN and NPAN constitute the network-based biomarker used for identifying significant proteins in lung carcinogenesis through the diagnostic evaluation

$$Y_C = C \cdot Y_N + E_C, \quad Y_N = N \cdot Y_N + E_N \quad (2.19)$$

where  $Y_C = [y_{1,C(n)} \cdots y_{K,C(n)}]^T$ ,  $Y_N = [y_{1,N(n)} \cdots y_{K,N(n)}]^T$  denotes the vectors of expression levels, and  $E_C$  and  $E_N$  indicate the noise vectors in cancer and non-cancer cases, respectively. A matrix indicating the difference between two protein association networks is defined as  $C - N$ , i.e.,

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1K} \\ d_{21} & d_{22} & \cdots & d_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ d_{K1} & d_{K2} & \cdots & d_{KK} \end{bmatrix} = \begin{bmatrix} \alpha_{11,C} - \alpha_{11,N} & \alpha_{12,C} - \alpha_{12,N} & \cdots & \alpha_{1K,C} - \alpha_{1K,N} \\ \alpha_{21,C} - \alpha_{21,N} & \alpha_{22,C} - \alpha_{22,N} & \cdots & \alpha_{2K,C} - \alpha_{2K,N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K1,C} - \alpha_{K1,N} & \alpha_{K2,C} - \alpha_{K2,N} & \cdots & \alpha_{KK,C} - \alpha_{KK,N} \end{bmatrix} \quad (2.20)$$

where  $d_{ij}$  denotes the difference in protein association ability between CPAN and NPAN among proteins  $i$  and  $j$ . Using matrix  $D$  to represent the difference in network structure between CPAN and NPAN, a carcinogenesis relevance value (CRV) was derived to quantify the correlation of each protein significant to lung carcinogenesis. To identify significant proteins, two important issues are taken into consideration. First, the magnitude of the association ability  $\alpha_{ij}$  denotes the significance of association of one protein to another. A higher absolute value of  $\alpha_{ij}$  implies that the two proteins are more tightly associated. Second, if a protein plays a more crucial role in lung carcinogenesis, then the difference in association numbers linked to the protein for CPAN and NPAN would be larger. For example, if one protein shares a strong association with many proteins in CPAN, but has weaker associations (no protein) in NPAN, then the protein in question is more likely to be involved in lung carcinogenesis. As a result, CRV is determined based on the difference in protein association abilities using the following equation [70]:

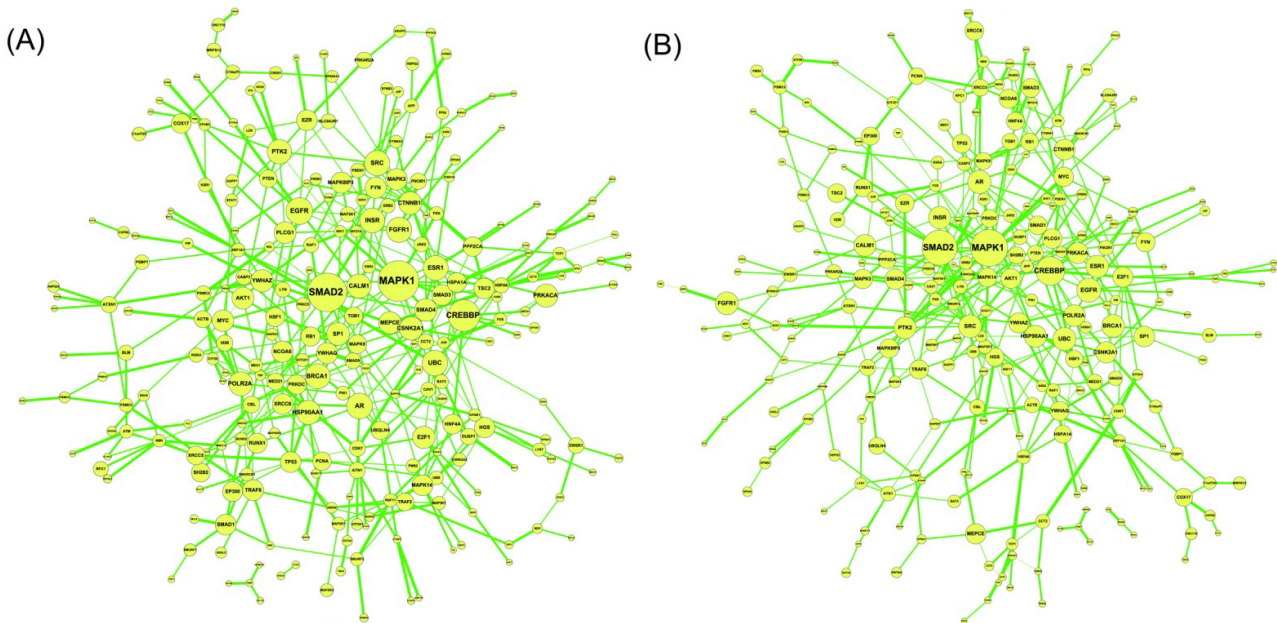
$$CRV_i = \sum_{j=1}^K |d_{ij}| \quad (2.21)$$

For the  $i$ -th protein in the network-based biomarker, the implication of Equation (2.21) is that the CRV quantifies the extent of protein associations that differentiate CPAN from NPAN.

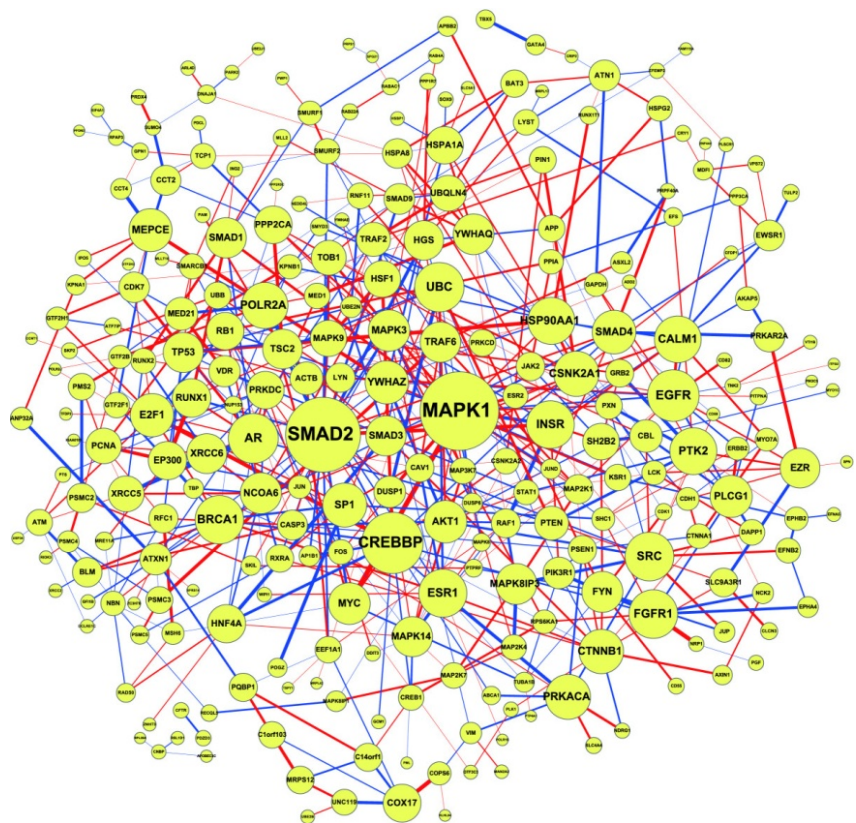
The above network-based biomarker approach is applied to the molecular investigation and diagnosis of lung cancer. The primary data set of GSE4115 (79 smokers with lung cancer and 73 smokers without lung cancer; obtained from the GEO database, <http://www.ncbi.nlm.nih.gov/geo/>) was used for construction of the network-based biomarker. The CPAN and NPAN in Equation (2.18), which consist of 399 and 393 protein associations respectively, constitute the network-based biomarker of lung cancer (Figure 5). The difference between CPAN and NPAN is shown in Figure 6. The CPAN indicates that 40 identified proteins play significant roles in lung carcinogenesis (Table S1).



**Figure 5.** The constructed network-based biomarker. (A) Cancer protein association network (CPAN) obtained from  $C$  in (2.18) by maximum likelihood estimation, Akaike's information criterion (AIC) selection, and Student's t-test. (B) Non-cancer protein association network (NPAN) obtained from  $N$  in Equation (2.18) using the same criteria.



**Figure 6.** The difference between CPAN and NPAN obtained from Equation (2.20) for network-based biomarkers for lung cancer. The significance of proteins (indicated by circle size) to the network-based marker is dependent on their CRVs in Equation (2.21), which are listed in Table S1.



## 2.6. On the Network Robustness and Filtering Ability versus Molecular Noise in GRNs Using a Stochastic System Approach

Molecular noise has been shown to play many roles in the cellular functions of GRNs, including noise-driven divergence of cell fates and population heterogeneity, noise-induced amplification of signals, generation of errors in DNA replication leading to mutation and evolution, and maintenance of the quantitative individuality of cells [50]. Other cellular processes influenced by noise include ion-channel gating, neural firing, developmental modules, cytoskeleton dynamics, and motors [50]. Phase variation in pathogenic bacteria, in which cells alternate randomly between expressing certain genes and silencing others, is thought to be a form of cultivated noise [25]. These molecular-level noisy phenomena are deeply rooted in the statistical mechanical behavior of so-called nanoscale chemical systems, where concentrations of reactive molecule species are extremely low and, consequently, fluctuations (noises) in the reaction rates are large [50]. Even though the molecular fluctuations leading to phase variation seem random in the individual, regulatory factors tune the variation to ensure mean levels of heterogeneity for the population, *i.e.*, the random molecular noises can be shown to be filtered or attenuated by the GRNs [25].

Since cellular molecular events are subject to significant thermal fluctuations and noisy processes with transcriptional control, alternative splicing, translation, diffusion, and chemical modification reactions, gene expression is best viewed as a stochastic process. Many observations suggest that molecular events underlying cellular physiology are subject to random fluctuations; these observations have led to the proposal of a stochastic model for gene expression and cellular functions [25]. Noise filtering can be considered from a signal processing perspective [74]. From this perspective, a pathway is viewed as an analog filter in terms of its frequency response. In terms of signal processing, these cellular pathways function as biological filters that transduce molecular signals and filter molecular noise [25].

For the convenience of illustration, the following linear biochemical dynamics of  $n$  genes GRN in Figure 7 [25] is simply considered initially:

$$\frac{dx(t)}{dt} = Nx(t) \quad (2.22)$$

where the concentration vector  $x(t)$  and stoichiometric matrix  $N$  are given by

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad N = \begin{bmatrix} N_{11} & \cdots & N_{1n} \\ \vdots & N_{ij} & \vdots \\ N_{n1} & \cdots & N_{nn} \end{bmatrix}$$

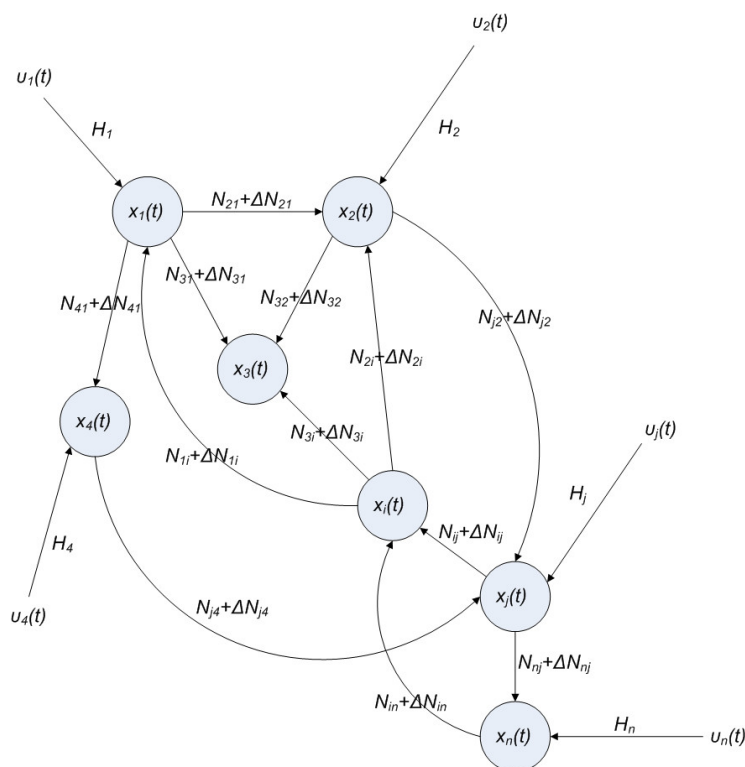
where  $x_i(t)$  denotes the concentration of the  $i$ -th gene, and  $N_{ij}$  denotes the interaction between genes  $j$  and  $i$ .

Suppose the linear GRN suffers intrinsic molecular fluctuations so that the stoichiometric matrix  $N$  is perturbed as  $N + \Delta N$ ,

$$\Delta N = \begin{bmatrix} \Delta N_{11}(t) & \cdots & \Delta N_{1n}(t) \\ \vdots & \Delta N_{ij}(t) & \vdots \\ \Delta N_{n1}(t) & \cdots & \Delta N_{nn}(t) \end{bmatrix} = \begin{bmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & M_{ij} & \vdots \\ M_{n1} & \cdots & M_{nn} \end{bmatrix} n(t) = Mn(t)$$

where  $\Delta N_{ij}$  denotes the random parametric fluctuation of  $N_{ij}$ ;  $M_{ij}$  denotes the deterministic part (amplitude) of fluctuation; and  $n(t)$  is white Gaussian noise with zero mean and unit variance, and denotes the stochastic part of fluctuation, *i.e.*, the stochastic part of fluctuation is absorbed to  $n(t)$ .

**Figure 7.** The linear  $n$  genes GRN with interaction  $N_{ij}$ , intrinsic fluctuation  $\Delta N_{ij}$ , gene expression  $x_i(t)$ , and extrinsic fluctuation  $v_i(t)$ .



The GRN with random parametric fluctuation can then be represented by

$$\begin{aligned} \frac{dx(t)}{dt} &= (N + \Delta N)x(t) \\ &= Nx(t) + Mx(t)n(t) \end{aligned} \quad (2.23)$$

For the convenience of analysis by the stochastic process, the above stochastic GRN can be represented as follows [74,75]:

$$dx(t) = Nx(t)dt + Mx(t)d\omega(t) \quad (2.24)$$

where  $d\omega(t) = n(t)dt$ .  $\omega(t)$  is called the stochastic Wiener process or Brownian motion [73].

The stochastic dynamic equations of GRNs are always nonlinear. In order to meet the nonlinear stochastic regulatory networks, Equation (2.23) should be generalized as the following Langevin equation [25]

$$dx(t) = N(x)dt + M(x)d\omega(t) \quad (2.25)$$

where  $N(x)$  denotes the nonlinear interaction equation of the nonlinear GRN, and  $M(x)d\omega(t)$  is due to nonlinear intrinsic random fluctuation.

Based on stochastic Lyapunov theory, let  $V(x) > 0$  with  $V(0) = 0$  denote the Lyapunov power-like function, then the stochastic GRN in Equation (2.23) or (2.24) is stochastically stable if  $E[dV(x(t))/dt] \leq 0$  [73]. With the choice of  $V(x) = x^T(t)Px(t)$  for some positive definite matrix  $P$ , the following result is derived.

**Proposition 1** [25]:

The linear GRN with stochastic perturbation in Equation (2.23) or (2.24) is stochastically stable if the following LMI

$$PN + N^T P + M^T P M \leq 0 \quad (2.26)$$

has a symmetric positive definite solution  $P > 0$ , *i.e.*, the phenotype of the linear stochastic GRN is maintained under intrinsic stochastic fluctuation.

**Remark 1:** (i) In the intrinsic noise-free case, the stable condition in (2.26) is reduced to  $PN + N^T P \leq 0$ , *i.e.*, the eigenvalues of system matrix  $N$  should be on the left-hand side of the complex domain. Obviously, if the LMI in Equation (2.26) has a positive solution  $P > 0$ , then the eigenvalues of  $N$  should be located on the far left-hand side of the complex domain with more negative real values in order to overcome the additional term  $M^T P M$  due to intrinsic random noise. (ii) If some eigenvalues of system interaction matrix  $N$  are near the  $j\omega$  axis, then intrinsic random molecular fluctuations across the  $j\omega$  axis perturb these modes more easily such that the linear GRN becomes unstable. The LMI in Equation (2.26) can be rearranged to

$$\underbrace{M^T P M}_{\text{intrinsic robustness}} \leq \underbrace{-(PN + N^T P)}_{\text{network robustness}} \quad (2.27)$$

that is,  $-(PN + N^T P)$  in Equation (2.27) can be taken as a measure of network robustness, and  $M^T P M$  due to the random parametric fluctuation can be taken as a measure of intrinsic robustness. The physical interpretation of Equation (2.27) is that if the network robustness can confer enough intrinsic robustness for tolerating intrinsic random parameter fluctuation, then the phenotype of the GRN is maintained.

For the nonlinear stochastic GRN in Equation (2.25), the following result is also derived.

**Proposition 2** [25]:

The nonlinear stochastic GRN in Equation (2.25) is still stochastically stable if the following Hamilton–Jacobi inequality (HJI) has a positive Lyapunov solution  $V(x) > 0$  with  $V(0) = 0$

$$\left( \frac{\partial V(x)}{\partial x} \right)^T N(x) + \frac{1}{2} M^T(x) \frac{\partial^2 V(x)}{\partial x^2} M(x) \leq 0 \quad (2.28)$$

That is, the phenotype of the nonlinear stochastic GRN in Equation (2.25) is maintained under intrinsic stochastic fluctuation. Similarly, the HJI in Equation (2.28) can be rearranged as the following phenotype robustness criterion:

$$\underbrace{\frac{1}{2} M^T(x) \frac{\partial^2 V(x)}{\partial x^2} M(x)}_{\text{intrinsic robustness}} \leq \underbrace{-\left( \frac{\partial V(x)}{\partial x} \right)^T N(x)}_{\text{network robustness}} \quad (2.29)$$

That is, if the network robustness of the nonlinear stochastic GRN in Equation (2.25) can confer

enough intrinsic robustness to tolerate intrinsic stochastic fluctuation, then the phenotype of the GRN is maintained.

After the robust stability of GRN is guaranteed under intrinsic biochemical stochastic fluctuation, the effect of environmental random molecular noises on the GRN may be discussed. If the linear GRN in Equation (2.24) also suffers from environmental molecular noises  $v(t) = [v_1(t) \cdots v_n(t)]^T$  outside the network (see Figure 7), then

$$\begin{aligned} dx(t) &= (Nx(t) + Hv(t))dt + Mx(t)d\omega(t) \\ Z(t) &= Cx(t) \end{aligned} \quad (2.30)$$

where  $H$  is a coupling matrix denoting the influence of environmental molecular signals  $v(t)$  on the GRN.  $Z(t)$  denotes the concentration of specific genes of interest. For example, if we want to examine the effect of molecular noises of  $\omega(t)$  and  $v(t)$  on gene  $i$  (i.e.,  $x_i(t)$ ), then we let  $C = \text{diag}(0 \dots 010 \dots 0)$ . That is, every element of  $C$  is zero except for the  $i$ -th element. To investigate the effect of molecular noises on the whole GRN, then  $C = I$ , the identity matrix. The positive value  $\rho$  in the following inequality is then called the effect of environmental noises (or signals) on  $Z(t)$  in the stochastic GRN in Equation (2.30) with  $x(0) = 0$

$$\frac{E \int_0^\infty Z^T(t)Z(t)dt}{E \int_0^\infty v^T(t)v(t)dt} \leq \rho^2 \quad (2.31)$$

That is,  $\rho$  is the upper bound of the effect of all environmental molecular signals or noises  $v(t)$  on the GRN. It is called the response level of the GRN.

From this is derived the following response level result of the GRN in Equation (2.30).

**Proposition 3** [25]:

The response level  $\rho$  of the linear stochastic GRN in Equation (2.30) is guaranteed if the following matrix inequality has a positive solution  $P > 0$

$$PN + N^T P + M^T P M + C^T C + \frac{1}{\rho^2} P H H^T P \leq 0 \quad (2.32)$$

By Shur complement [79], the above inequality is equivalent to the following LMI:

$$\begin{bmatrix} PN + N^T P + C^T C + M^T P M & P H \\ (P H)^T & -\rho^2 \end{bmatrix} \leq 0 \quad (2.33)$$

That is, if the above LMI has a positive solution  $P > 0$ , then the linear stochastic GRN in Equation (2.30) is robustly stable under random intrinsic molecular fluctuation and has a response level  $\rho$  to environmental molecular noises.

The optimal response level  $\rho_0$  of the linear stochastic GRN (Equation (2.30)) can be obtained by solving the following constrained optimization

$$\begin{aligned} \rho_0 &= \min \rho \\ \text{subject to } &\rho > 0, P > 0 \text{ and LMI in (2.33)} \end{aligned} \quad (2.34)$$

**Remark 2:** (i) If  $\rho_0 < 1$ , then environmental molecular noises  $v(t)$  are attenuated by the GRN and  $\rho_0$  is called the filtering ability of the GRN, i.e., the GRN is less sensitive to environmental noises. If  $\rho_0 > 1$ ,

then environmental molecular noises are amplified by the GRN, *i.e.*, the GRN is more sensitive to environmental noises. (ii) Substituting  $\rho_0$  for  $\rho$  in Equation (2.32) and rearranging, the following phenotype robustness criterion of the stochastic GRN is derived

$$\underbrace{M^T P M}_{\text{intrinsic robustness}} + \underbrace{C^T C + \frac{1}{\rho_0^2} P H H^T P}_{\text{environmental robustness}} \leq \underbrace{-(P N + N^T P)}_{\text{network robustness}} \quad (2.35)$$

The phenotype robustness criterion in (2.35) for the stochastic GRN in (2.30) can be denoted as “intrinsic robustness + environmental robustness  $\leq$  network robustness”. In other words, if the sum of intrinsic robustness and environmental robustness is less than network robustness, then the phenotype of the stochastic GRN remains robust under the influence of stochastic intrinsic fluctuation and environmental noises. In order to maintain the phenotype robustness criterion, GRNs need negative feedback loops to improve the network robustness on the right-hand side of Equation (2.35). Parallel loops and modular and redundant structures are also required to reduce the effect of intrinsic parameter variations on the GRN and to resist environmental disturbance, *i.e.*, to provide more intrinsic robustness and environmental robustness on the left hand side of Equation (2.35). This is why feedback loops, parallel loops, and modular and redundant structures are abundant in GRNs as they contribute to phenotype robustness and favor natural selection. (iii) If the network robustness in Equation (2.35) is not large enough to confer intrinsic robustness and environmental robustness simultaneously to maintain the phenotype of the GRN, some negative feedback gene loops are implemented as follows:

$$\begin{aligned} dx(t) &= ((N + K)x(t) + H v(t)) dt + M x(t) d\omega(t) \\ Z(t) &= C x(t) \end{aligned} \quad (2.36)$$

where  $K$  denotes specific negative feedback loops that improve the network robustness of the stochastic GRN. In this situation, the phenotype criterion in Equation (2.35) is modified as

$$\underbrace{M^T P M}_{\text{intrinsic robustness}} + \underbrace{C^T C + \frac{1}{\rho_0^2} P H H^T P}_{\text{environmental robustness}} \leq \underbrace{-[P(N + K) + (N + K)^T P]}_{\text{network robustness}} \quad (2.37)$$

It can be seen from the phenotype robustness criterion in Equation (2.37) that an adequate  $K$  can improve the network robustness of the stochastic GRN such that it better tolerates intrinsic stochastic fluctuation and is less responsive to environmental noise (*i.e.* smaller  $\rho_0$ ).

Similarly, the nonlinear stochastic GRN under environmental molecular noises (Equation (2.25)) should be modified as follows:

$$\begin{aligned} dx(t) &= (N(x) + H(x)v(t)) dt + M(x) d\omega(t) \\ Z(t) &= C x(t) \end{aligned} \quad (2.38)$$

The phenotype robustness criterion in Equation (2.37) is then modified as follows [25]:

$$\underbrace{x^T C^T C x + \frac{1}{4\rho_0^2} \left( \frac{\partial V(x)}{\partial x} \right)^T H H^T \frac{\partial V(x)}{\partial x}}_{\text{environmental robustness}} + \underbrace{\frac{1}{2} M^T(x) \frac{\partial^2 V(x)}{\partial x^2} M}_{\text{intrinsic robustness}} \leq \underbrace{- \left( \frac{\partial V(x)}{\partial x} \right)^T N(x)}_{\text{network robustness}} \quad (2.39)$$



In general, it is very difficult to solve the HJI in Equations (2.29) and (2.39). Global linearization techniques [79,81] or T-S fuzzy methods [82,83] can be employed to interpolate several local linearized systems to approximate the nonlinear stochastic system in Equation (2.38) as follows:

$$\begin{aligned} dx(t) &= \sum_{i=1}^L (h_i(z)N_i x + H_i v(t)) dt + M_i x d\omega(t) \\ Z(t) &= Cx(t) \end{aligned} \quad (2.40)$$

where  $h_i(z)$ ,  $i = 1, \dots, L$  denotes the interpolation functions of global linearization or fuzzy bases with  $h_i(z) \geq 0$ ,  $\sum_{i=1}^L h_i(z) = 1$ , and  $dx = (N_i x + H_i v)dt + M_i x d\omega$  is the  $i$ -th local linearized GRN.

In this situation, the network robustness criterion in Equation (2.39) can be modified accordingly [25]

$$\underbrace{C^T C + \frac{1}{\rho_0^2} P H_i H_i^T P}_{\text{local environmental robustness}} + \underbrace{M_i^T P M_i}_{\text{local intrinsic robustness}} \leq - \underbrace{(P N_i + N_i^T P)}_{\text{local network robustness}}, \quad i = 1, 2, \dots, L \quad (2.41)$$

From the phenotype robustness criterion in Equation (2.41), it is seen that if local network robustness is large enough for every local linearized GRN to simultaneously provide enough local intrinsic robustness for tolerating local random parameter fluctuations and enough local environmental robustness for resisting local environmental disturbance, then the phenotype of nonlinear stochastic GRN is maintained in spite of intrinsic stochastic parameter fluctuations and environmental molecular noises. If the phenotype robustness criterion in Equation (2.41) cannot be maintained, then negative feedback loops can be engineered to improve network robustness. These feedback loops may have potential applications for some types of therapy and drug target selection.

Systematic methodologies for the analysis of system characteristics of GRNs and PPINs detailed above are useful for the systematic designs of systems synthetic biology discussed in the following section.

### 3. Systems Synthetic Biology

Synthetic biology can be expected to have important applications in biotechnology and medicine, and to contribute significantly to a better understanding of the functioning of complex biological systems. Synthetic biology is concerned with the engineering of biological systems that fulfill a particular purpose. This is achieved by transformative innovation that makes it possible to build living machines from off-the-shelf chemical ingredients, employing many of the same strategies that electrical engineers use to make computer chips. The main goal of this nascent discipline is the design and construction of biological systems with desired behaviors [45–49]. Synthetic biology envisions the redesign of natural biological systems based on a set of powerful biomolecular techniques for the automated synthesis of DNA molecules and their assembly into genes and microbial genomes, for greater efficiency, as well as for the construction of functional “genetic circuits” and metabolic pathways for practical purposes [49]. The construction of GRNs has recently demonstrated the feasibility of synthetic biology [84–87]. The design of robust gene networks still presents a difficult challenge, and most newly designed gene networks cannot function properly. Such design failures are mainly due to both intrinsic perturbations such as gene expression noises, splicing, mutation, evolutionary changes, and environmental disturbances such as changing extracellular environments [52,53]. Designing robust synthetic gene

networks that can tolerate intrinsic parameter perturbations, attenuate extrinsic disturbances, and function properly in a host cell is therefore an important topic in synthetic biology [55].

### 3.1. Design of Specifications-Based Systems Synthetic Biology

Analysis of the dynamic properties of gene networks has been previously implemented using sensitivity analysis, either by qualitative simulation of coarse-grained models or by extensive numerical simulations of nonlinear differential equation models or stochastic dynamic models [56]. For applications in systems synthetic biology, however, these approaches are not satisfactory since local sensitivity analysis can provide only a partial description of all possible behaviors of a nonlinear gene network. In particular, it cannot guarantee that a synthetic network behaves as expected for all uncertain initial conditions, external disturbances, and parameter variations in a given range. Recently, Kuepfer *et al.* developed an approach based on semidefinite programming for partitioning parameter spaces of polynomial differential equation models into “feasible” and “infeasible” regions [68]. In this approach, “feasible” simply refers to the existence of a desired steady state of the synthetic network. Another approach using robustness analysis and tuning of synthetic networks was proposed by Batt *et al.* to provide a means to assess the robustness of a synthetic gene network with respect to parameter variations [51]. This approach allows searching for parameter sets for which a given property is satisfied, using a publicly available tool called RoVerGeNe. Several gene circuit design methods have recently been introduced insert or delete specific circuits in an existing gene network to modify its structure toward improved robust stability and filtering ability [38]. Robust synthetic gene network design, however, is a topic in itself, necessitating the design of a completely synthetic network with enough robust stability toward parameter fluctuations and with enough noise filtering ability to resist external disturbances, which allows it to function properly in a host cell.

A robust synthetic biological design incorporating molecular noises has been developed based on stochastic game theory [52]. However, the intrinsic parameter fluctuations of synthetic gene networks have not been considered in the design procedure. In a previous study [53] some system design specifications (engineering designs) had been provided by users so that the designer must engineer an artificial gene network to meet these design specifications. For the convenience of problem description, a simple design example is provided to give an overview of design problems for robust synthetic gene networks. A more general problem scenario is treated in preparation for this overview. First, consider the cross-inhibition network shown in Figure 8. This network is synthesized with two genes (a, b) that code for two repressor proteins (A, B). More specifically, protein A represses the expression of gene b, and, at higher concentration, the expression of its own gene. Protein degradation is not regulated. This synthetic system can be modeled by the following synthetic equations:

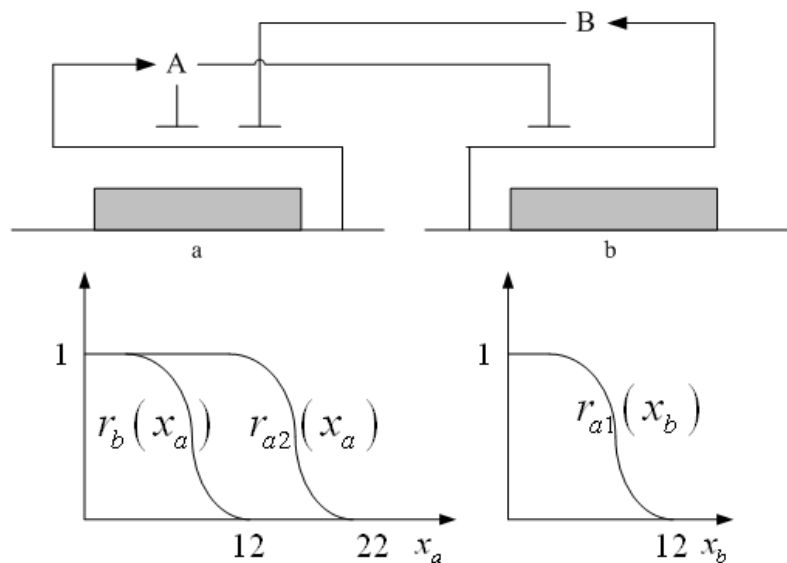
$$\dot{x}_a = k_a r_a(x_b) r_a(x_a) - \gamma_a x_a \quad (3.1)$$

$$\dot{x}_b = k_b r_b(x_a) - \gamma_b x_b \quad (3.2)$$

The state variables  $x_a$  and  $x_b$  denote the concentrations of proteins A and B, respectively.  $k$  and  $\gamma$  are the kinetic parameter and decay rate, respectively.  $r$  is the regulation function capturing the regulator effect of a transcriptional protein on gene expression, and has a smooth sigmoidal form (e.g., Hill function) [88].



**Figure 8.** A simple two-gene cross-inhibition network. The network's regulation functions are given in Equations (3.1) and (3.2).



The simple cross-inhibition network in (3.1) and (3.2) can be represented by the following stoichiometric matrix equation [7]

$$\begin{bmatrix} \dot{x}_a \\ \dot{x}_b \end{bmatrix} = \begin{bmatrix} k_a & 0 & -\gamma_a & 0 \\ 0 & k_b & 0 & -\gamma_b \end{bmatrix} \begin{bmatrix} r_{a1}(x_b)r_{a2}(x_a) \\ r_b(x_a) \\ x_a \\ x_b \end{bmatrix} \quad (3.3)$$

However, the stoichiometric matrix *in vivo* suffers from intrinsic parameter perturbations because of gene expression noises, splicing, mutation, evolutionary change, *etc.*, as in [52]

$$\begin{aligned} k_a &\rightarrow k_a + \Delta k_a n_a & \gamma_a &\rightarrow \gamma_a + \Delta \gamma_a n_a \\ k_b &\rightarrow k_b + \Delta k_b n_b & \gamma_b &\rightarrow \gamma_b + \Delta \gamma_b n_b \end{aligned} \quad (3.4)$$

where  $\Delta k_i$  and  $\Delta \gamma_i$  denote the amplitudes of fluctuations of the stochastic parameters and decay rates, respectively; and  $n_i$  is a random white noise with zero mean and unit variance.

Suppose the synthetic gene network also suffers from environmental disturbances because of changing extracellular environments and interactions with the cellular context in its host cell. The stochastic gene network can be then represented as

$$\begin{aligned} \begin{bmatrix} \dot{x}_a \\ \dot{x}_b \end{bmatrix} &= \begin{bmatrix} k_a + \Delta k_a n_a & 0 & -\gamma_a - \Delta \gamma_a n_a & 0 \\ 0 & k_b + \Delta k_b n_b & 0 & -\gamma_b - \Delta \gamma_b n_b \end{bmatrix} \begin{bmatrix} r_{a1}(x_b)r_{a2}(x_a) \\ r_b(x_a) \\ x_a \\ x_b \end{bmatrix} + \begin{bmatrix} v_a \\ v_b \end{bmatrix} \\ &= \begin{bmatrix} k_a & 0 & -\gamma_a & 0 \\ 0 & k_b & 0 & -\gamma_b \end{bmatrix} \begin{bmatrix} r_{a1}(x_b)r_{a2}(x_a) \\ r_b(x_a) \\ x_a \\ x_b \end{bmatrix} + \begin{bmatrix} \Delta k_a & -\Delta \gamma_a \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{a1}(x_b)r_{a2}(x_a) \\ x_a \end{bmatrix} n_a \\ &\quad + \begin{bmatrix} 0 & 0 \\ \Delta k_b & -\Delta \gamma_b \end{bmatrix} \begin{bmatrix} r_b(x_a) \\ x_b \end{bmatrix} n_b + \begin{bmatrix} v_a \\ v_b \end{bmatrix} \end{aligned} \quad (3.5)$$

where  $x = [x_a \ x_b]^T$  and  $v = [v_a \ v_b]^T$  denote the state vector and environmental disturbance of the synthetic gene network in the host cell, respectively. These intrinsic parameter fluctuations and environmental disturbances may cause the engineered synthetic gene network to be dysfunctional in the host cell.

If a synthetic gene network consists of  $n$  genes, then the stochastic gene network of Equation (3.5) in the host cell can be extended according to the following  $n$ -gene network dynamics [52]

$$\dot{x} = Nf(x) + \sum_{i=1}^m M_i g_i(x) n_i + v \quad (3.6)$$

where the state vector  $x = [x_1 \cdots x_n]^T$  denotes the concentrations of proteins in the synthetic gene network,  $N$  denotes the corresponding stoichiometric matrix of the  $n$ -gene network,  $M_i$ ;  $i = 1, \dots, m$  denotes fluctuation matrices associated with independent random noise sources  $n_i$ ,  $i = 1, \dots, m$ ; and the elements of  $M_i$  denote the standard deviations of the corresponding parameter fluctuations.  $v = [v_1 \cdots v_n]^T$  denotes the vector of external disturbances. This stochastic system is used to mimic the realistic dynamic behavior of a synthetic gene network of  $n$  genes in the host cell. As the network is subject to intrinsic parameter fluctuations and environmental disturbances in the context of the host cell, a robust synthetic gene network should be designed with the ability to not only tolerate parameter fluctuations and attenuate external disturbances, but also to achieve desired steady-state behaviors.

For convenience of analysis and design, the stochastic dynamic in Equation (3.6) of a synthetic gene network *in vivo* can be represented by the following Ito stochastic differential equation [73,74]

$$dx = (Nf(x) + v) + \sum_{i=1}^m M_i g_i(x) dW_i \quad (3.7)$$

where  $W_i(t)$  is a standard Wiener process with  $dW_i(t) = n_i dt$ .

To ensure correct and efficient operation of the gene network, several systematic design specifications should be imposed on it from the systematic engineering point of view.

(i) The kinetic parameters and the decay rates in stoichiometric matrix should be chosen from a biologically feasible range

$$N \in [N_1, N_2]$$

(ii) The network should tolerate the stochastic kinetic parameter and decay rate fluctuations with prescribed standard deviations in  $M_i$  in the state-dependent noise term

$$\sum_{i=1}^m M_i g_i(x) dW_i$$

(iii) The following desired steady state should be achieved

$$x \rightarrow x_d \text{ as } t \rightarrow \infty$$

(iv) The following prescribed attenuation level  $\rho$  of environmental disturbance should be achieved (*i.e.*, the  $H_\infty$  filtering ability)

$$\frac{E \int_0^\infty (x - x_d)^T Q (x - x_d) dt}{E \int_0^\infty v^T v dt} \leq \rho^2 \quad (3.8)$$

for all bounded environmental disturbances  $v(t)$ , where  $Q \geq 0$  is a symmetric weighting matrix and  $\rho$  is a prescribed attenuation level less than 1. That is, the effect of environmental disturbance  $v$  on the regulation error  $x - x_d$  should be less than the attenuation level  $\rho$  from the average energy perspective.

In order to achieve the desired steady state  $x_d$  and for convenience of design, the origin of the nonlinear stochastic system in Equation (3.7) should be shifted to  $x_d$ . Stabilizing the shifted nonlinear stochastic system at the origin would then also achieve  $x_d$ , simplifying the design procedure. Let  $\tilde{x} = x - x_d$ . The following shifted stochastic synthetic genetic system is then derived [79]:

$$d\tilde{x} = (Nf(\tilde{x} + x_d) + v)dt + \sum_{i=1}^m M_i g_i(\tilde{x} + x_d) dW_i \quad (3.9)$$

That is, the origin  $\tilde{x} = 0$  of the stochastic system in Equation (3.9) is at the desired steady state  $x_d$  of the original stochastic system in Equation (3.7).  $N \in [N_1, N_2]$  is then specified to tolerate the stochastic parameter fluctuation  $\sum_{i=1}^m M_i g_i(\tilde{x} + x_d) dW_i$  and efficiently attenuate the environmental disturbance  $v$  to the prescribed level

$$\frac{E \int_0^\infty \tilde{x}^T Q \tilde{x} dt}{E \int_0^\infty v^T v dt} \leq \rho^2 \quad (3.10)$$

From the stochastic synthetic gene network Equation (3.9) *in vivo*, we obtain the following result.

**Proposition 4** [53]:

If design kinetic parameters and decay rates in  $N \in [N_1, N_2]$  are chosen such that the following HJI has a positive solution  $V(\tilde{x}) > 0$

$$\begin{aligned} & \left( \frac{\partial V(\tilde{x})}{\partial \tilde{x}} \right)^T Nf(\tilde{x} + x_d) + \tilde{x}^T Q \tilde{x} + \frac{1}{4\rho^2} \left( \frac{\partial V(\tilde{x})}{\partial \tilde{x}} \right)^T \left( \frac{\partial V(\tilde{x})}{\partial \tilde{x}} \right) \\ & + \frac{1}{2} \sum_{i=1}^m g_i^T(\tilde{x} + x_d) M_i^T \frac{\partial^2 V(\tilde{x})}{\partial \tilde{x}^2} M_i g_i(\tilde{x} + x_d) < 0 \end{aligned} \quad (3.11)$$

then (a) the stochastic gene network in Equation (3.9) can achieve both the robust stabilization necessary to tolerate the intrinsic stochastic parameter perturbations and the prescribed attenuation level  $\rho$  of environmental disturbances, *i.e.*, design specifications (i), (ii), and (iv) are satisfied. (b) If the stochastic gene network is free of environmental disturbances ( $v(t) = 0$ ), then  $\tilde{x} \rightarrow 0$  or  $x \rightarrow x_d$  in probability, *i.e.*, design specification (iii) is achieved.

It is generally very difficult to specify  $N \in [N_1, N_2]$  to solve HJI in Equation (3.11) for  $V(\tilde{x}) > 0$  using a systematic method. If all global linearizations are bound by a polytope consisting of  $M$  vertices

$$\begin{bmatrix} \frac{\partial f(\tilde{x} + x_d)}{\partial \tilde{x}} \\ \frac{\partial g_1(\tilde{x} + x_d)}{\partial \tilde{x}} \\ \vdots \\ \frac{\partial g_m(\tilde{x} + x_d)}{\partial \tilde{x}} \end{bmatrix} \in \text{Co} \left[ \begin{bmatrix} F_1 \\ G_{11} \\ \vdots \\ Gm1 \end{bmatrix}, \dots, \begin{bmatrix} F_M \\ G_{1M} \\ \vdots \\ GmM \end{bmatrix} \right], \quad \forall \tilde{x} \quad (3.12)$$

where Co denotes the convex hull of a polytope with  $M$  vertices defined in Equation (3.12), then through the global linearization method [79,81], the state trajectories  $\tilde{x}(t)$  of the shifted gene network in Equation (3.9) can be represented by the convex combination of  $M$  linearized gene networks as

$$d\tilde{x} = \sum_{j=1}^M \alpha_j(\tilde{x}) \left( NF_j \tilde{x} dt + \sum_{i=1}^m M_i G_{ij} \tilde{x} dW_i \right) + v dt \quad (3.13)$$

where the interpolation function  $\alpha_j(\tilde{x})$  satisfies  $0 \leq \alpha_j(\tilde{x}) \leq 1$  and  $\sum_{j=1}^M \alpha_j(\tilde{x}) = 1$ . The trajectory of the nonlinear stochastic synthetic gene network in Equation (3.9) could thus be represented by the interpolated synthetic gene network in Equation (3.13). This yields the following result.

**Proposition 5** [53]:

Assuming that design kinetic parameters and decay rates in  $N \in [N_1, N_2]$  are chosen such that the following  $M$  LMIs have a common symmetric positive definite solution  $P > 0$

$$\begin{bmatrix} PNF_j + F_j^T N^T P + \sum_{i=1}^m G_{ij}^T M_i^T P M_i G_{ij} + Q & P \\ P & -\rho^2 I \end{bmatrix} < 0, \quad j = 1, \dots, M \quad (3.14)$$

then there are two results: (a) The stochastic synthetic gene network in Equation (3.7) is robustly stable toward intrinsic stochastic parameter perturbations and achieves a prescribed attenuation level  $\rho$  of environmental disturbances. That is, design specifications (i), (ii) and (iv) are satisfied. (b) If the synthetic gene network is free of environmental disturbance ( $v(t) = 0$ ), then the synthetic gene network in Equation (3.7) may asymptotically converge to the desired steady state. (An *in silico* design example is shown in **supplementary example 1**.)

**Remark 3:** (i) Gene circuit design can now be implemented using recombination technology to insert or delete TF binding sites in the promoter region of a regulated gene with the aim of increasing or decreasing the value of the kinetic parameter  $\kappa_i$  (i.e., different levels of affinity) of the regulated gene [89]. By inserting strong or weak binding sites, large or small values can be obtained. For example, the binding site of  $\kappa_i = 1$  is 10 times larger than that of  $\kappa_i = 0.1$  at the promoter region of target gene  $i$ . Changes to the decay rate  $\gamma_i$  can be implemented by shortening the 3' polyadenylate tail (referred to as deadenylation), which primarily triggers decapping, resulting in 5' to 3' exonucleolysis. Alternatively, removal of the 3' polyadenylate tail can increase  $\gamma_i$  [53]. Therefore, by shortening or elongating the gene's 3' polyadenylate tail, we can increase or decrease the decay rate  $\gamma_i$  of gene  $i$ . Directed evolution methods are also useful in changing the elasticity (kinetic property of  $\kappa_i$ ) and in designing biochemical circuits [53]. From a systems biology perspective, these advances in implementation techniques of  $\kappa_i$  and  $\gamma_i$  enable engineering of synthetic gene networks in the near future.

(ii) Because of the uncertain values of  $v$  and initial state  $\tilde{x}(0)$  of the synthetic gene network (3.9) in the host cell, the following minimax design has also been considered to reset the uncertainty of  $v(t)$  and  $\tilde{x}(0)$  *in vivo* [52].  $v(t)$  and  $\tilde{x}(0)$  are considered players maximizing their effects on regulation error  $\tilde{x}(t)$  when the kinetic parameters are specified as another player minimizing the effects of  $v$  and  $\tilde{x}(0)$  on  $\tilde{x}(t)$ .

$$\min_{\substack{k \in [k_1, k_2] \\ \gamma \in [\gamma_1, \gamma_2]}} \max_{\substack{\tilde{x}(0) \\ v}} \frac{E \int_0^{t_f} \tilde{x}^T Q \tilde{x} dx}{E \int_0^{t_f} v^T v dx} \quad (3.15)$$

A robust synthetic gene network based on dynamic game theory and fuzzy interpolation of local linearized linear systems can thus be efficiently designed [52].

(iii) When the desired output of the synthetic gene network is not a steady state but a time-varying signal like a specific binary or periodic signal, and the network structure becomes complex, the above systematic analysis methods are difficult to apply. Because natural selection of traits suited for environmental change is an important evolutionary mechanism, kinetic parameters of synthetic gene networks can be tuned by genetic algorithms (GAs) or evolution algorithms (EAs) to optimally track desired biological functions [54,58]. Based on the evolutionary network algorithm, the kinetic parameters may be tuned to maximize the fitness to some desired phenotype selected by natural selection. Consider the optimal tracking design of a synthetic gene network by network evolution shown in Figure 9. Let the tracking error be defined as

$$e(t) = y_d(t) - y(t)$$

The design purpose is then to tune the design parameter  $k_i$  using evolutionary (or genetic) algorithms such that the network can achieve the following optimal tracking

$$\min_{k_i \in [k_i^L, k_i^U]} E \int_0^{T_p} e^T(t) e(t) dt \quad (3.16)$$

where  $T_p$  denotes the present time. Let

$$J(k) = E \int_0^{T_p} e^T(t) e(t) dt \quad (3.17)$$

Let the fitness function  $F(k)$  be defined as

$$F(k) = \frac{1}{J(k)} \quad (3.18)$$

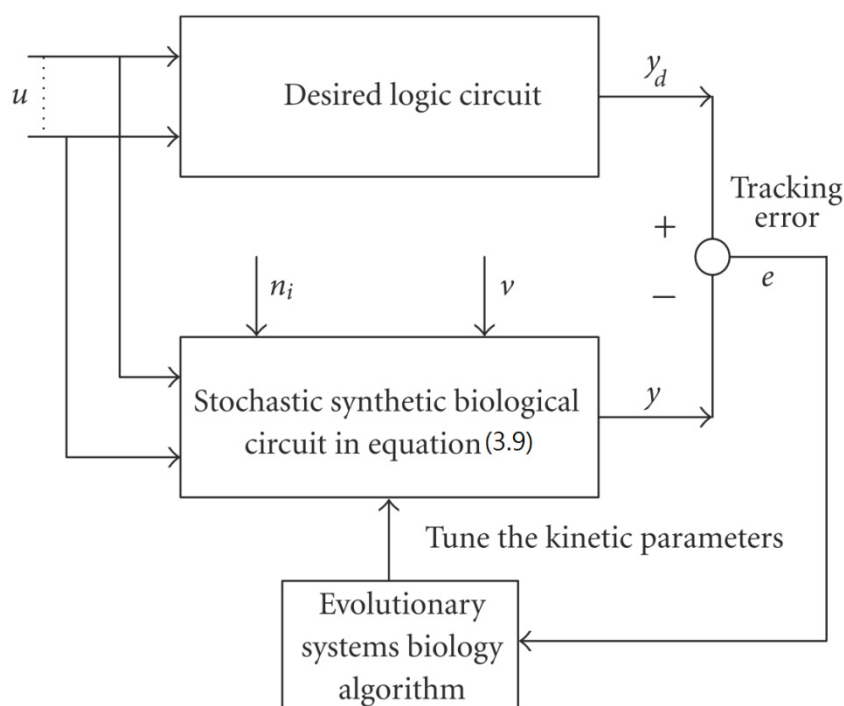
Adapting a parameter vector (chromosome)  $k_i \in [k_i^L, k_i^U]$  by EA or GA to minimize  $J(k)$  for the desired network behavior tracking of the synthetic gene network is equivalent to maximizing the fitness function in Equation (3.18) to meet the natural selection. A robust biological network design with a desired output behavior  $y_d(t)$  is therefore equivalent to as solution to the following fitness maximization problem using an evolutionary network method [54,58]

$$F(k^*) = \max_{k \in [k^L, k^U]} F(k) \quad (3.19)$$

The evolutionary network algorithm (or evolutionary systems biology algorithm [58]) is employed to solve the above fitness maximization problem via genetic operators such as selection, crossover, and mutation. It mimics natural selection in an evolutionary process to tune the kinetic parameter network of the synthetic gene network to an environmental change. A GA-based algorithm with binary coding of the chromosome has been proposed for the design of robust synthetic genetic oscillators with prescribed amplitude, period, and phase [54]. The oscillator is intended to allow protein concentrations to track desired periodic reference signals under intrinsic and environmental noises. Based on an evolutionary systems biology algorithm that encodes each chromosome in a real-valued vector, the design

parameters of target gene circuits can evolve to specific values in order to robustly track a desired biological function in spite of such interferences [58].

**Figure 9.** Block diagram of the optimal tracking scheme for synthetic biological circuit design using an evolutionary systems biology algorithm. Based on a network algorithm mimicking natural selection in an evolutionary process, the design parameters  $k$  of a synthetic biological circuit are tuned to minimize the tracking error between the desired logic circuit and the stochastic synthetic biological circuit, and to achieve the desired behavior tracking.



### 3.2. Robust Synthetic Gene Network Design via Library-Based Search Methods

Over the past decade, synthetic biology has made significant progress in designing biological parts. Even synthetic gene networks may have been constructed using a variety of biological components to achieve desired behaviors. A limitation on the development of complex synthetic gene networks intended to track specific reference trajectories is the lack of an efficient method for selecting suitable biological parts from libraries; experimental data in promoter libraries cannot be directly used for selecting adequate promoter parts. Current promoter libraries therefore need to be redefined based on promoter activity. This would allow development of library-based search methods.

A dynamic model can be used in the indirect evaluation of the activities of promoter parts to help redefine existing promoter libraries. In the TetR-regulated promoter library shown in Figure 10,  $x(c, t)$  and  $X(c, t)$  respectively denote the concentrations of the mRNA and protein of gene *yegfp* (which is used to measure protein expression). The dynamic model of the promoter-regulation gene part is constructed with [56,57]

$$\begin{aligned}\dot{x}(c, t) &= P_{TetR}(c, r) - \beta x(c, t) \\ \dot{X}(c, t) &= \alpha x(c, t) - \gamma_{yEGFP} X(c, t)\end{aligned}\quad (3.20)$$

where  $\beta$  and  $\gamma_{yEGFP}$  denote the degradation rates of mRNA and protein yEGFP, respectively, and  $\alpha$  denotes the translation rate. The promoter regulation function  $P_{TetR}(c, r)$ , which is dependent on repressor activity  $\gamma$  and the choice of promoter  $c$ , has the form

$$P_{TetR}(c, r) = c_r + \frac{c_s - c_r}{1 + (r/K_{TetR})^{n_{TetR}}} = c_r + (c_s - c_r) H_{TetR}(r), c = [c_r, c_s] \in Lib_{TetR} \quad (3.21)$$

where the promoter  $c$  has two promoter activities  $c_r$  and  $c_s$  (the minimum and maximum values of the promoter regulation function  $P_{TetR}(c, r)$ , respectively) for the TetR-regulated promoter library  $Lib_{TetR}$ . That is,  $c = (c_r, c_s) \in Lib_{TetR}$ .  $K_{TetR}$  and  $n_{TetR}$  denote the TetR–DNA binding affinity and binding cooperativity of regulatory protein TetR and DNA, respectively.  $H_{TetR}(r)$  is a Hill function capturing the effect of a regulatory protein.

Based on the estimated promoter activities  $c = (c_r, c_s)$  via maximum and minimum values of the steady state of protein expression data, some promoter libraries can be redefined in such a way that they can be efficiently selected from the design of the synthetic gene network (Table 2) [56,57]. Since a synthetic gene network always consists of a set of promoter-regulation gene circuits (Figure 10), the design of complex synthetic gene network addresses how to select a set of promoters from the corresponding promoter libraries that have promoter activities adequate for achieving the design specifications. A well-known gene circuit topology, the simple toggle switch, is shown for illustration purposes in Figure 11. The toggle switch has two distinct stable states and can be reversibly switched between them by changing the inducers ATc and IPTG. Let  $x_1(c_1, t)$ ,  $x_2(c_2, t)$ , and  $x_3(c_3, t)$  denote the concentrations of mRNAs *tetR*, *lacI*, and *yegfp*, respectively; and let  $X_1(c_1, t)$ ,  $X_2(c_2, t)$ , and  $X_3(c_3, t)$  denote the concentrations of proteins TetR, LacI, and yEGFP, respectively. The dynamic model of the toggle switch gene network in Figure 11 is then modeled as

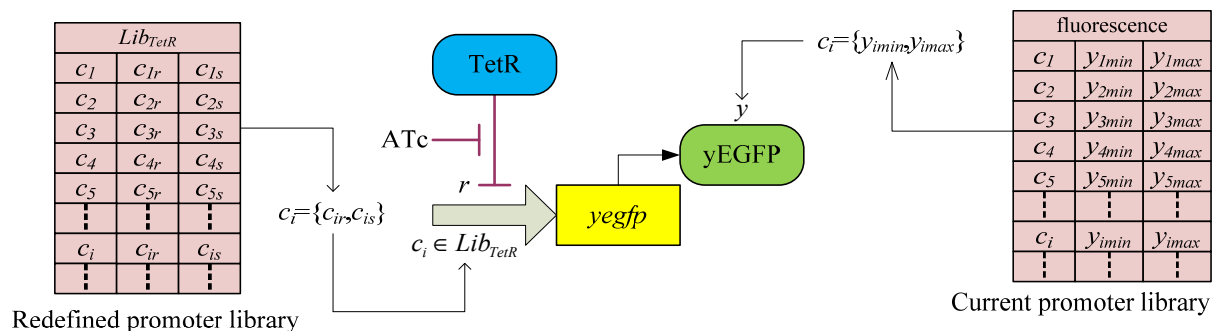
$$\begin{cases} \dot{x}_1(c_1, t) = p_{LacI}(c_1, r_1) - \beta x_1(c_1, t) \\ \dot{X}_1(c_1, t) = \alpha x_1(c_1, t) - \gamma_1 X_1(c_1, t) \\ \dot{x}_2(c_2, t) = p_{TetR}(c_2, r_2) - \beta x_2(c_2, t) \\ \dot{X}_2(c_2, t) = \alpha x_2(c_2, t) - \gamma_2 X_2(c_2, t) \\ \dot{x}_3(c_3, t) = p_{TetR}(c_3, r_3) - \beta x_3(c_3, t) \\ \dot{X}_3(c_3, t) = \alpha x_3(c_3, t) - \gamma_3 X_3(c_3, t) \\ y(c, t) = X_3(c_3, t) \\ c = [c_1, c_2, c_3], c_1 \in Lib_{LacI}, c_2, c_3 \in Lib_{TetR} \end{cases} \quad (3.22)$$

The promoter regulation functions  $p_{LacI}(c_1, r_1)$ ,  $p_{TetR}(c_2, r_2)$ , and  $p_{TetR}(c_3, r_3)$  are dependent on the selection of promoters  $c_1$ ,  $c_2$ , and  $c_3$  from the corresponding promoter libraries in Table 2. The output of interest  $y(c, t)$  is dependent of the selected promoter set  $c = [c_1, c_2, c_3]$  with adequate promoter activities from the corresponding promoter libraries in Table 2. This dynamic toggle switch gene network model consists of three interactive dynamic models of promoter-regulation gene parts, as shown in Equation (3.20).

**Table 2.** Redefined TetR- and LacI-regulated promoter libraries. The redefined TetR- and LacI-regulated promoter libraries (*i.e.*,  $Lib_{TetR}$  and  $Lib_{LacI}$ ) comprise different promoters (*i.e.*,  $T_k$  and  $L_k$ ,  $k = 0, \dots, 20$ ) with their corresponding activities of  $c_s$  and  $c_r$  obtained from previous libraries of experimental data.

TetR-regulated promoter library ( $Lib_{TetR}$ )			LacI-regulated promoter library ( $Lib_{LacI}$ )		
Promoter	Promoter activity		Promoter	Promoter activity	
	$c_s$	$c_r$		$c_s$	$c_r$
$T_0$	2121	0.1724	$L_0$	1657.5	0.3018
$T_1$	1604	0.7576	$L_1$	923.97	0.2567
$T_2$	1376.6	0.1936	$L_2$	860.87	0.2244
$T_3$	1169.8	0.4672	$L_3$	674.92	1.9189
$T_4$	974.52	0.0753	$L_4$	651.58	1.1680
$T_5$	942.77	0.2281	$L_5$	570.07	3.5062
$T_6$	967.17	0.1493	$L_6$	527.83	0.5497
$T_7$	738.57	0.0702	$L_7$	323.45	0.1248
$T_8$	641.74	0.7135	$L_8$	327.77	0.1772
$T_9$	564.24	0.2620	$L_9$	309.74	0.5439
$T_{10}$	501.35	0.0756	$L_{10}$	298.35	0.1146
$T_{11}$	469.35	0.0788	$L_{11}$	250.16	0.1326
$T_{12}$	466.16	0.1636	$L_{12}$	248.03	0.1171
$T_{13}$	356.88	0.0927	$L_{13}$	239.32	0.1010
$T_{14}$	348.95	0.1483	$L_{14}$	190.2	0.0959
$T_{15}$	274.79	0.1067	$L_{15}$	163.84	0.4813
$T_{16}$	250.04	0.0857	$L_{16}$	166.42	0.0989
$T_{17}$	188.77	0.1366	$L_{17}$	131.63	0.1190
$T_{18}$	119.57	0.0753	$L_{18}$	108.96	0.0903
$T_{19}$	111.57	0.1185	$L_{19}$	101.89	0.0982
$T_{20}$	70.909	0.1606	$L_{20}$	85.673	0.2174

**Figure 10.** Single schematic diagram of the synthetic promoter-regulation gene circuit. The existing TetR-regulated promoter library contains the minimum and maximum values of fluorescence  $[y_{imin}, y_{imax}]$  corresponding to with and without TetR (repressor) binding. Based on the promoter regulation function (3.21) and these values, the promoter library is redefined for the design of synthetic gene networks (Table 2).





The multiobjective design approach for the  $H_2/H_\infty$  synthetic gene network based on promoter libraries selects an adequate promoter set  $c = [c_1, c_2, c_3]$  from corresponding promoter libraries such that the following two design objectives are achieved simultaneously [56]:

(i)  $H_\infty$  desired noise attenuation level  $\rho_d$ :

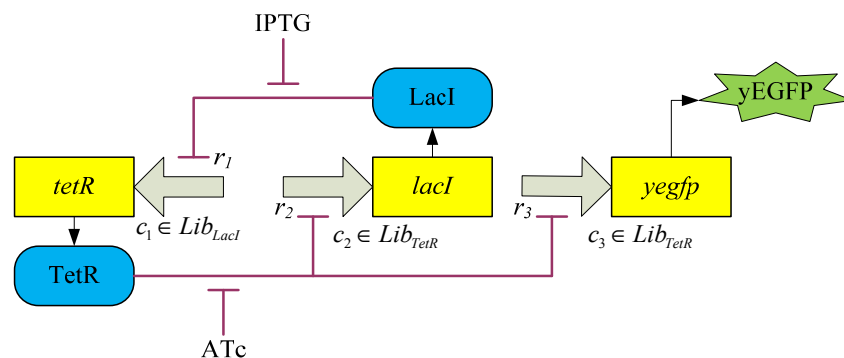
$$\frac{E \int_0^{t_f} (y(c, t) - y_r(t))^T Q (y(c, t) - y_r(t)) dt}{E \int_0^{t_f} v^T(t) v(t) dt} \leq \rho_d^2 \text{ for } c \in Lib_j, j = 1, \dots, m \quad (3.23)$$

(ii)  $H_2$  optimal reference tracking:

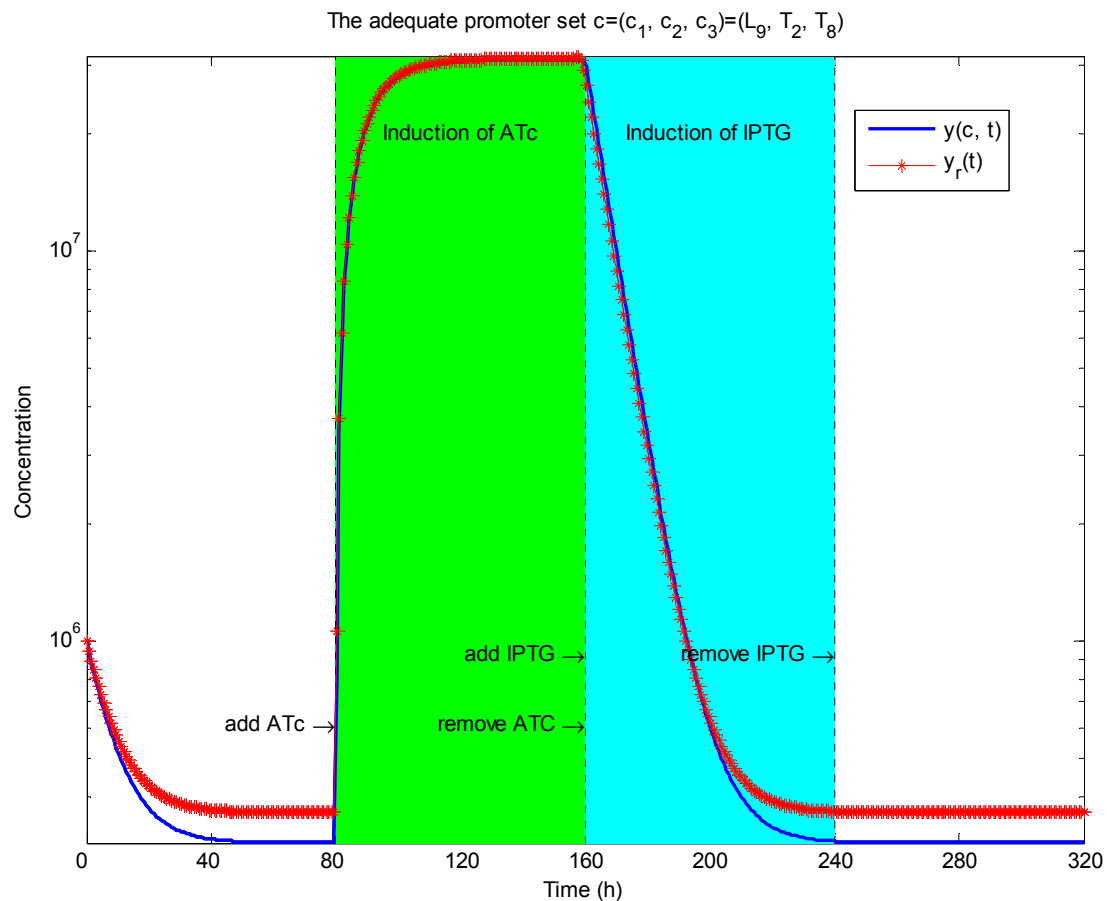
$$\min_{\substack{c \in Lib_j \\ j=1, \dots, m}} E \int_0^{t_f} (y(c, t) - y_r(t))^T Q (y(c, t) - y_r(t)) dt \quad (3.24)$$

By solving an optimization problem with two constraints [56], adequate promoters can be selected from the corresponding libraries in Table 2 to achieve the two design objectives in Equations (3.23) and (3.24). Based on the synthetic toggle switch (Figure 11) and dynamic model (equation (3.22)) with intrinsic parameter variation and external disturbance, the adequate promoter set  $c = [c_1, c_2, c_3]$   $[L_9, T_2, L_8]$  is selected from promoter libraries in Table 2 to achieve the multi-objective  $H_2/H_\infty$  reference tracking specified in Equations (3.23) and (3.24). Simulation results with  $v(t) = 10 \times [n_1 \dots n_6]^T$  are shown in Figure 12.

**Figure 11. Synthetic gene circuit topology: simple toggle switch.** The regulatory protein TetR, which is induced by ATc, inhibits the transcription of *lacI* by binding promoter  $c_2$ . TetR also inhibits transcription of *yegfp* by binding promoter  $c_3$  to repress the expression of the fluorescent protein yEGFP. The protein LacI, which is induced by the inducer IPTG, inhibits the transcription of *tetR* by binding promoter  $c_1$ . The gene circuit has two distinct stable states, and can reversibly switch between them by changing the inducers ATc and IPTG. If an adequate promoter set  $c = [c_1, c_2, c_3]$  is selected from corresponding promoter libraries, then yEGFP can be used to track the desired behaviors generated by a reference model. In the reference model,  $c_1$  is selected from the LacI-regulated promoter library, and  $c_2$  and  $c_3$  are selected from the TetR-regulated promoter library in Table 2 (i.e.,  $c_1 \in Lib_{LacI}$ ,  $c_2, c_3 \in Lib_{TetR}$ ).



**Figure 12. Simulation of toggle switch.** By solving the LMI-constrained optimization problem of the  $H_2/H_\infty$  design objective Equations (3.23) and (3.24) for the synthetic gene network in Figure 11 through the library searching method, an adequate promoter set  $c = [c_1, c_2, c_3] = [L_9, T_2, L_8]$  is selected from the corresponding promoter libraries. The inducer ATc is added to the synthetic gene network at 80–160 hours to induce the gene network, and then the inducer IPTG is added at 160–240 hours. The output  $y(c, t)$  clearly produces a robust track with the desired reference output  $y_r(t)$ .



**Remark 4:** Collective rhythms of GRNs, especially the synchronization of dynamic cells mediated by intercellular communication, have become a subject of considerable interest to biologists and theoreticians [90]. Synchronization of a population of synthetic genetic oscillators is an important consideration in practical applications, because a population distributed over different host cells needs to exploit molecular phenomena in a simultaneous manner in order to function as a biological entity. However, this synchronization of synthetic gene networks in different host cells may be corrupted by intrinsic kinetic parameter fluctuations and extrinsic environmental molecular noise. Therefore, robust synchronization is an important design topic in nonlinear stochastic coupled synthetic genetic oscillators with intrinsic kinetic parametric fluctuations and extrinsic molecular noise. A systems biology approach indicates [59,60] that if the synchronization robustness criterion is greater than or equal to the sum of the intrinsic robustness and extrinsic robustness, then the stochastic coupled synthetic oscillators can be robustly synchronized in spite of intrinsic parameter fluctuation and extrinsic noise. If the criterion for synchronization robustness is violated, then an external control scheme can be designed to improve

robustness by adding inducers to the coupled synthetic genetic network. These robust synchronization criteria and control methods are useful for a population of coupled synthetic networks with emergent synchronization behavior, especially for multicellular engineered gene networks [60].

#### 4. Systems Metabolic Engineering

Systems metabolic engineering aims to amplify or delete specific genes in metabolic pathways to perform metabolic engineering within a systems biology framework. Regulatory gene networks, metabolic networks, and other cellular networks can thus be engineered in an integrated system manner. Systems biological analysis via large-scale genome-wide analyses and computational bioinformatic tools can allow the rapid evaluation of the global physiology of a cell with respect to various cellular regulations. These include transcriptional and translational regulation, as well as metabolic engineering distribution [37,61–63]. Results of this systems biology approach can be used to predict targets for a metabolic engineering approach within the host strain. Furthermore, the integration of high-throughput, large-scale, genome-wide analyses with *in silico* simulation results might provide additional information on cellular status at various hierarchical levels from genome to fluxome. Strain selection for strain improvement by systems metabolic engineering is divided into several phases [37]: (i) A base strain is allowed to develop. (ii) The base strain is further engineered via synthetic biological methods, based on results obtained from high-throughput genome-wide bioinformatic data and systems biology computational analyses. (iii) The performance of this preliminary production strain is evaluated in actual fermentation. (iv) The results are then fed back into further strain development until a superior strain showing the desired performance is derived.

By combining the results of multiple genome-wide analyses and computational analyses, systems biology may allow us to reach an unprecedented level of understanding of cellular physiology and characteristics [72], which can subsequently be used in a systems synthetic biology framework to design systems metabolic engineering strategies, for example, for strain improvement. However, novel computational methods such as large-scale data mining, multidimensional data integration, and data-driven network inference and deep curation schemes still need to be developed and applied to the integration and interpretation of heterogeneous large-scale genome-wide bioinformatic data, which are closely interconnected by complex regulatory and metabolic pathways [37,91]. Several robust biochemical circuit designs have recently been proposed to improve the network robustness of metabolic pathways. The proposed design schemes provide a systems biology method with potential applications in systems synthetic circuit design in systems metabolic engineering and systems drug design. Broadly, a metabolic network is a collection of enzymatic reactions that process cellular and intercellular metabolites. In systems metabolic engineering, the rates of reactions or fluxes, which correspond directly to changes in concentrations of substrates, enzymes, factors, or products, are often measured. Such systematic changes in concentrations can be expressed in terms of dynamic differential equations. The following S-system representation has been used for the last three decades as an efficient model for describing a dynamic metabolic network [7,27]

$$\begin{aligned}
\dot{X}_1 &= \alpha_1 \prod_{j=1}^{n+m} X_j^{g_{1j}} - \beta_1 \prod_{j=1}^{n+m} X_j^{h_{1j}} \\
&\vdots \\
\dot{X}_i &= \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} = V_i - V_{-i}, \quad i=1, \dots, n \\
&\vdots \\
\dot{X}_n &= \alpha_n \prod_{j=1}^{n+m} X_j^{g_{nj}} - \beta_n \prod_{j=1}^{n+m} X_j^{h_{nj}}
\end{aligned} \tag{4.1}$$

where  $X_1, \dots, X_{n+m}$  are metabolites, such as substrates, enzymes, factors, or products of a biochemical network, in which  $X_1, \dots, X_n$  denote  $n$  dependent variables and  $X_{n+1}, \dots, X_{n+m}$  denote the independent variables. In a metabolic network, intermediate metabolites and products are dependent variables, whereas substrates and enzymes are independent variables. The rate of change in  $X_i$ ,  $\dot{X}_i$  is equal to the difference between two terms, one for production or accumulation, and the other for degradation or clearance. Each term is the product of a positive rate constant  $\alpha_i$  or  $\beta_i$  and all dependent and independent variables that directly affect production or degradation, respectively. Each variable  $X_j$  is raised to the power of a kinetic parameter  $g_{ij}$  and  $h_{ij}$ , which represents an activating effect of  $X_j$  on  $X_i$  when its value is positive, and an inhibitive effect when its value is negative.  $V_i$  and  $V_{-i}$  represent aggregate flux into and out of the  $X_i$  pool.

Construction of the S-system representation of a metabolic network and estimation of its kinetic parameters from experimental data are described in [7] and the literature references therein. The nonlinear parameter estimation problem of S-systems has recently been solved efficiently by evolution optimization methods [92]. It is generally difficult to study the network robustness or sensitivity of a nonlinear system such as Equation (4.1). Fortunately, many important characteristics of an S-system at or close to the steady state can be analyzed by using simple algebraic methods. Since most metabolic networks in nature operate near the steady state, at which inputs and outputs are almost balanced, the following focuses on the network robustness of metabolic networks at the steady state [7,27].

#### 4.1. Robust Biochemical Circuit Design in Metabolic Networks

Consider the steady state of the metabolic network in Equation (4.1),

$$\alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} = \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}, \quad i=1, 2, \dots, n \tag{4.2}$$

Taking the logarithm on both sides of Equation (4.2) and making some rearrangements,

$$\sum_{j=1}^n (g_{ij} - h_{ij}) \ln X_j = \ln \beta_i - \ln \alpha_i - \sum_{j=n+1}^{n+m} (g_{ij} - h_{ij}) \ln X_j, \quad i=1, 2, \dots, n \tag{4.3}$$

Introducing new variables and coefficients,

$$y_j = \ln X_j, \quad a_{ij} = g_{ij} - h_{ij}, \quad b_i = \ln \left( \frac{\beta_i}{\alpha_i} \right) \tag{4.4}$$

the steady state of the metabolic network is obtained as

$$A_D Y_D = b - A_I Y_I \quad (4.5)$$

where

$$Y_D = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad Y_I = \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{n+m} \end{bmatrix}, \quad A_D = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \quad A_I = \begin{bmatrix} a_{1,n+1} & \cdots & a_{1,n+m} \\ \vdots & \ddots & \vdots \\ a_{n,n+1} & \cdots & a_{n,n+m} \end{bmatrix}$$

In Equation (4.5),  $A_D$  denotes the system matrix of the interactions between dependent variables  $Y_D$ , and  $A_I$  indicates the interactions between dependent variables  $Y_D$  and independent variables  $Y_I$ . In the nominal parameter case, it is assumed that the inverse of  $A_D$  exists so that  $Y_D$  can be solved uniquely, *i.e.*, the metabolic network results in only one steady state (phenotype). Therefore, the steady state of the biochemical system is given by

$$Y_D = A_D^{-1}(b - A_I Y_I) \quad (4.6)$$

Suppose that parameter variations due to mutation, thermal changes, or disease can alter the kinetic properties of the steady state of a metabolic network as follows [27]

$$(A_D + \Delta A_D)(Y_D + \Delta Y_D) = (b + \Delta b) - (A_I + \Delta A_I)(Y_I + \Delta Y_I) \quad (4.7)$$

Parameter perturbations are defined as

$$\Delta A_D = \begin{bmatrix} \Delta a_{11} & \cdots & \Delta a_{1n} \\ \vdots & \Delta a_{ij} & \vdots \\ \Delta a_{n1} & \cdots & \Delta a_{nn} \end{bmatrix} = \begin{bmatrix} \Delta g_{11} - \Delta h_{11} & \cdots & \Delta g_{1n} - \Delta h_{1n} \\ \vdots & \Delta g_{ij} - \Delta h_{ij} & \vdots \\ \Delta g_{n1} - \Delta h_{n1} & \cdots & \Delta g_{nn} - \Delta h_{nn} \end{bmatrix}$$

$$\Delta b = \begin{bmatrix} \Delta b_1 \\ \vdots \\ \Delta b_n \end{bmatrix}, \quad \Delta A_I = \begin{bmatrix} \Delta g_{1,n+1} - \Delta h_{1,n+1} & \cdots & \Delta g_{1,n+m} - \Delta h_{1,n+m} \\ \vdots & \Delta g_{i,n+j} - \Delta h_{i,n+j} & \vdots \\ \Delta g_{n,n+1} - \Delta h_{n,n+1} & \cdots & \Delta g_{n,n+m} - \Delta h_{n,n+m} \end{bmatrix}$$

where  $\Delta A_D$  denotes perturbations due to kinetic parameter variations,  $\Delta b$  denotes perturbations due to rate constant variations, and  $\Delta A_I$  denotes perturbations due to kinetic parameter variations between dependent and independent variables.  $\Delta A_D$  can influence the existence of the steady state of the metabolic network.

From Equation (4.7), we derive

$$A_D (I + A_D^{-1} \Delta A_D) (Y_D + \Delta Y_D) = (b + \Delta b) - (A_I + \Delta A_I) Y_I \quad (4.8)$$

If the following robustness condition holds [27]

$$\|A_D^{-1} \Delta A_D\|_2 < 1 \quad (4.9)$$

then the singular values of  $I + A_D^{-1} \Delta A_D$  are free of zero and the inverse  $(I + A_D^{-1} \Delta A_D)^{-1}$  exists. Therefore, the steady state of the perturbative metabolic network in Equation (4.8) is uniquely solved: as

$$Y_D + \Delta Y_D = (I + A_D^{-1} \Delta A_D)^{-1} A_D^{-1} [(b + \Delta b) - (A_I + \Delta A_I) Y_I] \quad (4.10)$$

The above analysis implies that if the robustness condition in Equation (4.9) holds, then the steady state of a metabolic system is preserved under parameter variations  $\Delta A_D$ , *i.e.*,  $Y_D + \Delta Y_D$  in Equation (4.10) has a small difference  $\Delta Y_D$  from the nominal in Equation (4.6) under small perturbation. However, if the condition in Equation (4.9) does not hold, then individual values of  $I + A_D^{-1}\Delta A_D$  may be zero, the inverse  $(I + A_D^{-1}\Delta A_D)^{-1}$  may not exist, and the steady state  $Y_D + \Delta Y_D$  may cease to exist under the parameter perturbation  $\Delta A_D$ . As an example, consider the singular value decomposition:

$$A_D = U^T \Sigma V = \sum_{i=1}^n \sigma_i u_i^T v_i \quad (4.11)$$

where  $\sigma_i$  denotes the  $i$ -th singular value and  $u_i, v_i \in R^n$  denote the corresponding left and right singular vectors, respectively. Therefore, if a parameter variation is specified as follows:

$$\Delta A_D = -\sigma_i u_i^T v_i, \quad i = 1, 2, \dots, n \quad (4.12)$$

then

$$A_D + \Delta A_D = U^T \begin{bmatrix} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_{i-1} & & & & \\ & & & 0 & & & \\ & & & & \sigma_{i+1} & & \\ & 0 & & & & \ddots & \\ & & & & & & \sigma_n \end{bmatrix} V \quad (4.13)$$

Obviously, the inverse  $(A_D^{-1}\Delta A_D)^{-1}$  or  $(I + A_D^{-1}\Delta A_D)^{-1}$  does not exist under the parameter perturbation in Equation (4.12).

**Remark 5:** The parameter perturbations in the direction of singular vectors like Equation (4.12) are the network's points of fragility. The robustness prevents this kind of parameter variation and guarantees the existence of the steady state of the metabolic networks. When unexpected perturbations like Equation (4.12) are encountered, a catastrophic failure of the network follows. Robust circuit design is a necessary fail-safe mechanism in such situation. For example, the trehalose pathway in yeast consists of only a few metabolites that form a substrate cycle. It is governed by a surprisingly complex control system that is composed of several inhibiting or activating signaling mechanisms [7].

Equivalently, the network robustness of the metabolic network in Equation (4.9) can be rewritten as a more intuitive criterion for network robustness:

$$A_D^{-1} \Delta A_D \Delta A_D^T A_D^{-T} < I \quad \text{or} \quad \Delta A_D \Delta A_D^T < A_D A_D^T \quad (4.14)$$

That is,  $A_D A_D^T$  is the upper bound of  $\Delta A_D \Delta A_D^T$  without violation of robust stability at steady state. If the network robustness criterion in Equation (4.14) holds, then the steady state of the perturbative metabolic network still exists.

If a metabolic network cannot tolerate perturbations, *i.e.*, the network robustness criterion in Equation (4.14) is violated, then robust control via a biochemical circuit design is a suitable remedy. Based on robustness analysis, we develop a biochemical circuit design scheme for the robust control of metabolic

networks. Consider the robust control system design of the metabolic network (in Equation (4.1)) by specific biochemical feedback circuits in a more general metabolic network form [27]

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} \prod_{k=1}^n X_k^{f_{ik}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \prod_{k=1}^n X_k^{l_{ik}}, \quad i=1, \dots, n \quad (4.15)$$

where  $X_k^{f_{ik}}$  denotes a new biochemical control circuit with  $X_k$  regulating the production of  $X_i$  by the kinetic parameter  $f_{ik}$ .  $X_k^{l_{ik}}$  denotes a new biochemical control circuit with  $X_k$  regulating the degradation of  $X_i$  by the kinetic parameter  $l_{ik}$ . The choice of regulating objects,  $X_k$  and  $X_i$ , and the specification of the kinetic parameters,  $f_{ik}$  and  $l_{ik}$ , are designed according to the feasibility of biochemical circuit linkage to achieve a desired robustness to tolerate  $\Delta A_D$  within the prescribed range of kinetic parameter perturbations in a metabolic network. Since  $f_{ik}$  and  $l_{ik}$  represent the elasticities of the corresponding enzymes in the designed control circuits, the implementation of control circuits is heavily dependent on the elasticity specification of these enzymes.

Consider the robust control system of the metabolic network in Equation (4.15). By using a similar procedure to Equations (4.2)–(4.7) at steady state,

$$(A_D + F + \Delta A_D)(Y_D + \Delta Y_D) = (b + \Delta b) - (A_I + \Delta A_I)(Y_I + \Delta Y_I) \quad (4.16)$$

where  $f_{ij}$  and  $l_{il}$  are the kinetic parameters in  $F$  of the biochemical control circuit to be specified in Equation (4.15).

Suppose we can find some  $F$  such that the inverse of  $(A_D + F + \Delta A_D)$  exists. Equation (4.16) is equivalent then to

$$(A_D + F) \left( I + (A_D + F)^{-1} \Delta A_D \right) (Y_D + \Delta Y_D) = (b + \Delta b) - (A_I + \Delta A_I)(Y_I + \Delta Y_I) \quad (4.17)$$

Similar to Equation (4.14), a robust design scheme for the controlled metabolic system in Equation (4.17) is given by

$$\left\| (A_D + F)^{-1} \Delta A_D \right\|_2 < 1 \quad \text{or} \quad \Delta A_D \Delta A_D^T < (A_D + F)(A_D + F)^T \quad (4.18)$$

In this case of robust circuit design, the design purpose is to specify feedback circuits such that the structural stability robustness of the metabolic network is improved, thus enabling the network robustness criterion in Equation (4.18) to tolerate larger parameter perturbations  $\Delta A_D$ . The phenotype (*i.e.*, the steady state of the controlled metabolic network in Equation (4.16)) is then given by

$$Y_D + \Delta Y_D = \left( I + (A_D + F)^{-1} \Delta A_D \right)^{-1} (A_D + F)^{-1} [(b + \Delta b) - (A_I + \Delta A_I)(Y_I + \Delta Y_I)] \quad (4.19)$$

A simple example of network robustness analysis and circuit design is given in **supplementary example 2**.

#### 4.2. Multipurpose Circuit Control Design of Metabolic Networks

The robustness design above focuses on the tolerance of kinetic parameter perturbations  $\Delta A_D$ . The effects of rate constant variations  $\Delta b$  and of environmental changes or upstream regulatory changes  $\Delta Y_I$  on output variations  $\Delta Y_D$  should also be considered in the circuit design of metabolic networks to guarantee robustness against both intrinsic parameter variations and extrinsic environmental

perturbations. The sensitivity of  $\Delta Y_D$  to  $\Delta b$  in the designed metabolic network of Equation (4.16) is given by [7]

$$\frac{\Delta Y_D}{\Delta b} = (A_D + F)^{-1} \quad (4.20)$$

The sensitivity of  $\Delta Y_D$  to  $\Delta Y_I$  is given by

$$\frac{\Delta Y_D}{\Delta Y_I} = -(A_D + F)^{-1} A_I \quad (4.21)$$

It is more appealing to design a robust metabolic network with desired sensitivities to variations in rate constants and environmental signals, *i.e.*,

$$\left\| \frac{\Delta Y_D}{\Delta b} \right\|_2 < s_1, \quad \left\| \frac{\Delta Y_D}{\Delta Y_I} \right\|_2 < s_2 \quad (4.22)$$

where the upper bounds  $s_1$  and  $s_2$  are prescribed in advance by the biochemical circuit designer. From Equations (4.20)–(4.22), the equivalent sensitivity criteria for Equation (4.22) are obtained as [27]

$$\begin{aligned} I - s_1^2 (A_D + F)(A_D + F)^T &< 0 \\ A_I A_I^T - s_2^2 (A_D + F)(A_D + F)^T &< 0 \end{aligned} \quad (4.23)$$

A multipurpose circuit design approach is therefore aimed to construct a biochemical circuit  $F$  to simultaneously satisfy the design requirements of network robustness in Equation (4.18) and network sensitivity in Equation (4.23). For example, suppose the goal is to design a robust biochemical circuit  $f_{22}$  in supplementary example 2 to tolerate kinetic parameter variation  $\Delta A_D$  and satisfy the network sensitivity in Equation (4.22) or (4.23) with prescribed sensitivities of  $\|\Delta Y_D / \Delta b\|_2 < \|\Delta Y_D / \Delta b\|_{2,\text{nominal}} \equiv s_1 = \|A_D^{-1}\|_2 = 3.42$  and  $\|\Delta Y_D / \Delta Y_I\|_2 < \|\Delta Y_D / \Delta Y_I\|_{2,\text{nominal}} \equiv s_2 = \|A_D^{-1} A_I\|_2 = 2.66$ . Then  $f_{22}$  should be specified to satisfy the robust circuit design and the following inequalities:

$$\begin{aligned} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 3.42^2 \begin{bmatrix} -0.5 & -0.1 & -0.05 \\ 0.5 & -0.5 + f_{22} & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} -0.5 & 0.5 & 0 \\ -0.1 & -0.5 + f_{22} & 0.5 \\ -0.05 & 0 & -0.5 \end{bmatrix} &< 0 \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} - 2.66^2 \begin{bmatrix} -0.5 & -0.1 & -0.05 \\ 0.5 & -0.5 + f_{22} & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} -0.5 & 0.5 & 0 \\ -0.1 & -0.5 + f_{22} & 0.5 \\ -0.05 & 0 & -0.5 \end{bmatrix} &< 0 \end{aligned} \quad (4.24)$$

simultaneously. With the help of Matlab, the range of  $f_{22}$  necessary to tolerate  $\Delta A_D$  in and satisfy the desired network sensitivity in Equation (4.24) is found to be  $[-1, -0.081]$ . We choose  $f_{22} = -0.407$  as a design example, which is a negative self-regulation. It has been found to efficiently eliminate the effect of parameter variations by negative compensation. About 10% of yeast genes encoding regulators are negatively self-regulating; thus, this mechanism seems to be important for maintaining robustness in yeast [27]. The metabolic network and time responses are shown in Figure S5(d) and (e), respectively.

Similarly, suppose the goal in the design case in equation is to specify  $f_{12}$  and  $l_{22}$  to satisfy equations and (4.24), *i.e.*,



$$\begin{aligned}
& \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - s_1^2 \begin{bmatrix} -0.5 & -0.1 + f_{12} & -0.05 \\ 0.5 & -0.5 - l_{22} & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} -0.5 & 0.5 & 0 \\ -0.1 + f_{12} & -0.5 - l_{22} & 0.5 \\ -0.05 & 0 & -0.5 \end{bmatrix} < 0 \\
& \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} - s_2^2 \begin{bmatrix} -0.5 & -0.1 + f_{12} & -0.05 \\ 0.5 & -0.5 - l_{22} & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} -0.5 & 0.5 & 0 \\ -0.1 + f_{12} & -0.5 - l_{22} & 0.5 \\ -0.05 & 0 & -0.5 \end{bmatrix} < 0
\end{aligned} \tag{4.25}$$

where  $s_1$  and  $s_2$  are the same as above. Similarly, the ranges of  $f_{12}$  and  $l_{22}$  necessary to tolerate  $\Delta A_D$  in equation (4.25) and satisfy the desired sensitivity criteria in Equation (4.23) are found to be  $[-1, 0]$  and  $[0, 1]$ , respectively.  $f_{12} = -0.08$  and  $l_{22} = 0.31$  are chosen as a design example. The metabolic network and time responses are shown in Figure S5(f) and (g), respectively. Another example of a TCA cycle is given in **supplementary example 3**.

Given that the goal is to design a robust metabolic circuit to tolerate the perturbation  $\Delta A_D$  and to achieve the desired sensitivities  $s_1$  and  $s_2$  in Equation (4.23), the robust control circuit design problem can be reduced to specifying  $F$  to satisfy the following multipurpose control circuit design derived from Equations (4.18) and (4.23):

$$\begin{aligned}
& \Delta A_D \Delta A_D^T < (A_D + F)(A_D + F)^T \\
& I - s_1^2 (A_D + F)(A_D + F)^T < 0 \\
& A_I A_I^T - s_2^2 (A_D + F)(A_D + F)^T < 0
\end{aligned} \tag{4.26}$$

with nominal sensitivities  $s_1 = \|A_D^{-1}\|_2 = 8.3685$  and  $s_2 = \|A_D^{-1} A_I\|_2 = 7.5464$ .

If one biochemical control pathway with kinetic parameter  $f_{12}$  is designed to satisfy the multi-objective design criteria in Equation (4.26), then the range of  $f_{12}$  is found to be within  $[-0.8, -0.1]$ . If  $f_{12}$  is chosen as  $-0.2$  (Figure S6(a), blue line), then the time responses of the designed TCA cycle network shown in Figure S6(d), which match the desired properties of the proposed design method, are obtained. That is, the robust controlled biochemical network not only can tolerate  $\Delta A_D$  (to preserve its phenotype under parameter perturbations) but also retains sensitivity to environmental molecules in the nominal case. If dynamic circuit design is employed to implement the biochemical circuit  $f_{12}$ , then an enzyme capable of catalyzing the reaction  $X_2 \rightarrow X_1$  is required. Additionally, a TF ( $Z$ ) has to be found such that oxaloacetate2 ( $X_2$ ) can bind to the promoter of the enzyme's inhibitor gene, as  $f_{12}$  is negative. The concentration of  $X_2$  could therefore regulate  $X_1$  through the kinetic parameter  $f_{12}$ . The elasticity of the enzyme inhibitor's gene sequence has to be modulated then to the specified performance by rational design or directed evolution. This allows the construction of the biochemical control circuit  $f_{12}$ .

The previous section discusses robust circuit design of metabolic networks based on the steady state of an S-system model. In the following section, a stochastic dynamic model approach is taken based on the systems biology approach outlined in Section 3.

#### 4.3. Robust Control Circuit Design of Stochastic Dynamic Metabolic Networks

Based on sampled data from biochemical experiments and the delayed effect of molecular diffusion and transport in cells, the linear metabolic regulatory network can be suitably modeled by the following discrete-time dynamic system:

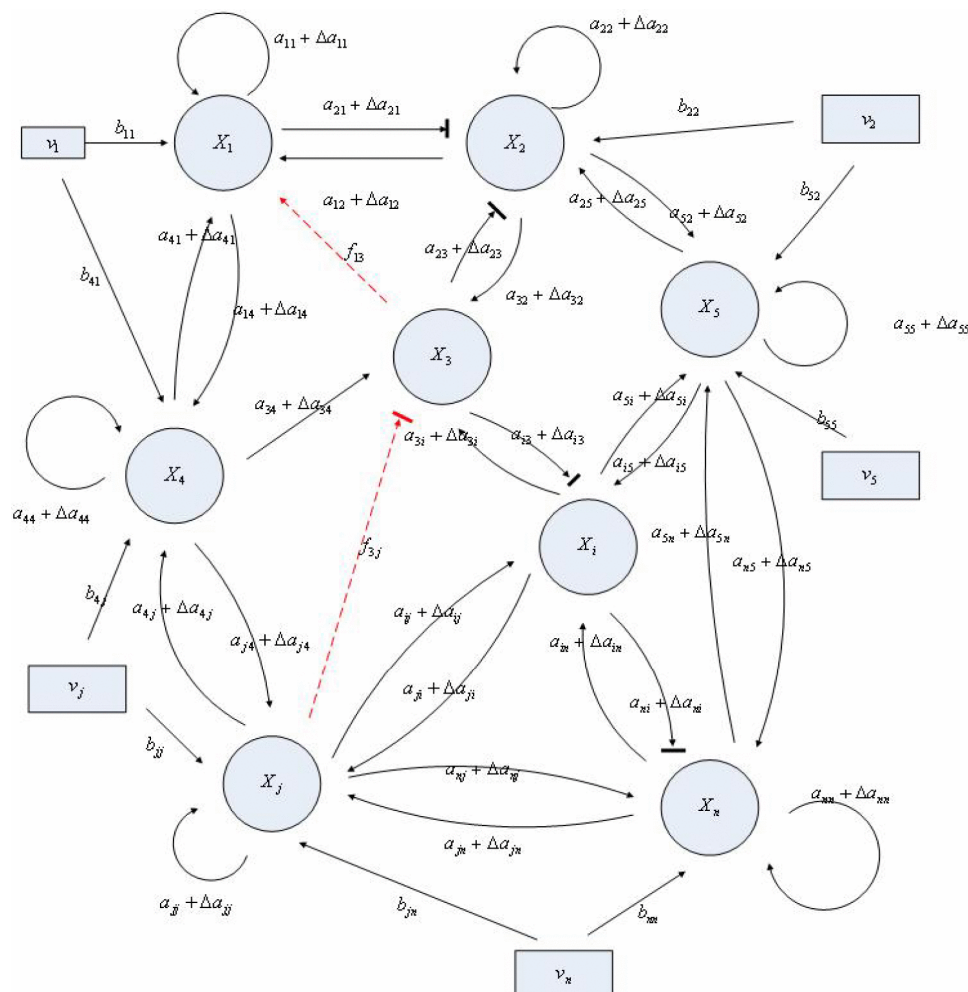
$$\begin{aligned} X(t+1) &= AX(t), \quad X(0) = X_0 \\ y(t) &= CX(t) \end{aligned} \quad (4.27)$$

where the state vector  $X(t)$  denotes the discrete-time expression vector of the molecules (mRNAs, proteins, or other chemical complexes in the biochemical regulatory network) at time  $t$ ,  $A$  denotes the stoichiometric (interactive) matrix among these molecules (see Figure 13), and  $y(t)$  denotes the metabolic output. Thus,

$$X(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_i(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & \ddots & & & \vdots \\ \vdots & & a_{ij} & & \vdots \\ \vdots & & & \ddots & \vdots \\ a_{n1} & \dots & \dots & \dots & a_{nn} \end{bmatrix}$$

where  $x_i(t)$ ,  $i = 1, \dots, n$  are molecular concentrations of mRNAs, proteins, and other complexes in the biochemical network.

**Figure 13.** Linear metabolic network of  $n$  molecules with intrinsic parameter fluctuation  $\Delta a_{ij}$  and extrinsic noise  $v_i$   $f_{ij}$  denotes the biochemical circuit design from  $x_j$  to  $x_i$  to improve network robustness stability and noise-filtering ability.



Suppose the kinetic parameters of a biochemical regulatory network of metabolic processes are affected by the following intrinsic perturbation and environmental disturbance:

$$\begin{aligned} X(t+1) &= \left( A + \sum_{i=1}^L A_i n_i(t) \right) X(t) + B_v v(t) \\ y(t) &= CX(t) \end{aligned} \quad (4.28)$$

where  $\sum_{i=1}^L A_i n_i(t)$  denotes the intrinsic parameter fluctuations due to an  $L$  random fluctuation source (e.g., thermal fluctuation, alternative splicing, molecular diffusion, *etc.*), and  $n_i(t)$  denotes the  $i$ -th random noise with the statistics  $E[n_i(t)] = 0$  and  $E[n_i^2(t)] = \sigma_i^2$ ,  $i = 1, \dots, L$ .  $v(t)$  denotes the environmental disturbance.

Based on the robust stabilization and disturbance filtering described in Section 2, the environmental disturbance attenuation level  $\rho$  of the metabolic network is denoted as

$$\frac{\|y(t)\|_2}{\|v(t)\|_2} \leq \rho \quad (4.29)$$

Following the systems biology approach in Section 2, the following robust result for a linear stochastic metabolic network is derived.

**Proposition 6** [28]:

For a linear stochastic metabolic network with intrinsic parameter fluctuation and environmental disturbance, if there exists a positive definite matrix  $P = P^T > 0$  such that the following matrix inequality holds for a desired disturbance attenuation level  $\rho$

$$A^T P A - P + \sum_{i=1}^L \sigma_i^2 A_i^T P A_i + C^T C + (A^T P B_v)^T (B_v^T P B_v - \rho^2 I)^{-1} A^T P B_v < 0 \quad (4.30)$$

then the intrinsic parameter perturbation can be tolerated and environmental disturbance can be attenuated to a level  $\rho$  in the stochastic metabolic network Equation (4.28).

The phenotype robustness criterion in Equation (4.30) could be rewritten as

$$\underbrace{\sum_{i=1}^L \sigma_i^2 A_i^T P A_i}_{\text{intrinsic robustness}} + \underbrace{C^T C + (A^T P B_v)^T (B_v^T P B_v - \rho^2 I)^{-1} A^T P B_v}_{\text{environmental robustness}} \leq \underbrace{P - A^T P A}_{\text{network robustness}} \quad (4.31)$$

**Remark 6:** (i) The physical interpretation of the phenotype robustness criterion of the metabolic network in Equation (4.31) is that if intrinsic robustness allowing tolerance of intrinsic parameter fluctuation and environmental robustness, as well as attenuation of environmental disturbance are simultaneously conferred by the network robustness, then the phenotype of the metabolic network is maintained. It can be shown that if the eigenvalues of  $A$  are closer to the origin, then the network robustness is commensurately larger.

(ii) To solve the phenotype robustness criterion in Equation (4.30), it can be transformed into the following equivalent LMI:

$$\begin{bmatrix} A^T P A - P + C^T C + \sum_{i=1}^L \sigma_i^2 A_i^T P A_i & A^T P B_v \\ B_v^T P A & B_v^T P B_v - \rho^2 I \end{bmatrix} \leq 0 \quad (4.32)$$

According to its definition, the disturbance attenuation (filtering) ability  $\rho_0$  (i.e., the minimum  $\rho$ ) can be obtained by solving the following constrained optimization problem:

$$\begin{aligned} \rho_0 &= \min_P \rho \\ \text{subject to } P &> 0, \text{ and LMI in (4.32)} \end{aligned} \quad (4.33)$$

If the metabolic network in Equation (4.28) cannot achieve a prescribed disturbance attenuation ability  $\rho_P$  (i.e.,  $\rho_P < \rho_0$ ) specified for therapeutic or biotechnological purposes, then a biochemical circuit design using state feedback is necessary:

$$\begin{aligned} X(t+1) &= \left( A + F + \sum_{i=1}^L A_i n_i(t) \right) X(t) + B_v v(t) \\ y(t) &= CX(t) \end{aligned} \quad (4.34)$$

The phenotype robustness criterion in Equation (4.31) is then modified as

$$\underbrace{\sum_{i=1}^L \sigma_i^2 A_i^T P A_i}_{\text{intrinsic robustness}} + \underbrace{C^T C + (A^T P B_v)^T (B_v^T P B_v - \rho_P^2 I)^{-1} A^T P B_v}_{\text{environmental robustness}} \leq \underbrace{P - (A + F)^T P (A + F)}_{\text{network robustness}} \quad (4.35)$$

The negative feedback circuit  $F$  could shift the eigenvalues of  $A$  to the origin, thereby improving the network robustness of the metabolic network on the right-hand side of Equation (4.35) and achieving the prescribed attenuation level  $\rho_P < \rho_0$ . In general, the interactions of a biochemical regulatory network in metabolic processes are nonlinear in real biosystems. In this situation, a nonlinear biochemical regulatory network of metabolic pathways under intrinsic stochastic parameter perturbation and environmental disturbance can be represented based on the stochastic dynamic model of systems biology in Section 2:

$$\begin{aligned} X(t+1) &= f(x) + \sum_{i=1}^L f_i(x) n_i(t) + B_v v(t) \\ y(t) &= CX(t) \end{aligned} \quad (4.36)$$

In the following, the robust stability and filtering ability  $\rho_0$  on  $v(k)$  at  $y(k)$  in the stochastic nonlinear metabolic network in Equation (4.36) is discussed.

**Proposition 7** [28]:

If the following matrix inequality holds for a Lyapunov function  $V(x(k)) > 0$  and a noise attenuation level  $\rho$

$$\begin{aligned} &\left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T (f(X(t)) - X(t)) + \frac{1}{2} \left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T \left( \frac{\partial V(X(t))}{\partial X(t)} \right) + \frac{1}{2} \sum_{i=1}^L f_i^T(X(t)) \\ &\times \frac{\partial^2 V(X(t))}{\partial X^2(t)} f_i(X(t)) + \frac{1}{2\rho^2} \left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T B_v B_v^T \left( \frac{\partial V(X(t))}{\partial X(t)} \right) + \frac{1}{2} X^T(t) C^T C X(t) \leq 0 \end{aligned} \quad (4.37)$$

then the effect of environmental disturbance  $v(k)$  on  $y(k)$  is less than  $\rho$ , i.e., the robust disturbance attenuation level  $\rho$  is achieved for the nonlinear metabolic network in Equation (4.36).

The filtering ability  $\rho_0$  of the nonlinear stochastic metabolic network in Equation (4.36) that attenuates environmental disturbance can be obtained by solving the following constrained optimization problem:

$$\rho_0 = \min_{V(X)>0} \rho$$

subject to HJI in (4.37)

(4.38)

Following the robustness result discussed in Section 2, the phenotype robustness criterion for the nonlinear stochastic metabolic network in Equation (4.36) is given as

$$\begin{aligned} & \underbrace{\frac{1}{2} \sum_{i=1}^L f_i^T(X) \frac{\partial^2 V(X(t))}{\partial X^2(t)} f_i(X)}_{\text{intrinsic robustness}} \\ & + \underbrace{\frac{1}{2} \left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T \left( \frac{\partial V(X(t))}{\partial X(t)} \right) + \frac{1}{2\rho^2} \left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T B_v B_v^T \left( \frac{\partial V(X(t))}{\partial X(t)} \right) + \frac{1}{2} X^T(t) C^T C X(t)}_{\text{environmental robustness}} \\ & \leq - \underbrace{\left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T (f(X(t)) - X(t))}_{\text{network robustness}} \end{aligned} \quad (4.39)$$

Suppose the network robustness of the metabolic network cannot confer enough intrinsic robustness and environmental robustness to maintain the phenotype of the metabolic network. In this situation, some negative feedback loops  $Fg(X(t))$  should be implemented to improve network robustness as follows:

$$\begin{aligned} X(t+1) &= f(X(t)) + Fg(X(t)) + \sum_{i=1}^L f_i(X(t))n_i(t) + B_v v(t) \\ y(t) &= CX(t) \end{aligned} \quad (4.40)$$

In this case, the phenotype robustness criterion in Equation (4.39) should be modified as follows:

$$\begin{aligned} & \underbrace{\frac{1}{2} \sum_{i=1}^L f_i^T(X) \frac{\partial^2 V(X(t))}{\partial X^2(t)} f_i(X)}_{\text{intrinsic robustness}} \\ & + \underbrace{\frac{1}{2} \left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T \left( \frac{\partial V(X(t))}{\partial X(t)} \right) + \frac{1}{2\rho^2} \left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T B_v B_v^T \left( \frac{\partial V(X(t))}{\partial X(t)} \right) + \frac{1}{2} X^T(t) C^T C X(t)}_{\text{environmental robustness}} \\ & \leq - \underbrace{\left( \frac{\partial V(X(t))}{\partial X(t)} \right)^T (f(X(t)) + Fg(X(t)) - X(t))}_{\text{network robustness}} \end{aligned} \quad (4.41)$$

That is, the negative feedback loops  $Fg(X)$  are employed to improve network robustness to tolerate more fluctuations of the intrinsic parameter and to filter more environmental disturbances.

It is generally very difficult to solve the HJI in Equations (4.37), (4.39), and (4.41) for robust chemical circuit design of metabolic networks. Based on the global linearization method in Equation (3.12) [79,81] or fuzzy interpolation methods [83,93], the nonlinear stochastic metabolic network can be approximated by interpolating M local metabolic networks as follows:

$$\begin{aligned}
 X(t+1) &= \sum_{j=1}^M \alpha_j(X(t)) \left( A_j X(t) + FG_j X(t) + \sum_{i=1}^L A_{ij} X(t) n_i(t) \right) + B_v v(t) \\
 y(t) &= CX(t)
 \end{aligned}
 \quad (4.42)$$

where  $f(X)$ ,  $g(X)$ , and  $f_i(X)$  are approximated by  $\sum_{j=1}^M \alpha_j(X) A_j X(t)$ ,  $\sum_{j=1}^M G_j X(t)$ , and  $\sum_{i=1}^M A_{ij} X(t)$ , respectively. In this situation, the following robust chemical circuit design for a metabolic network with a prescribed noise attenuation level  $\rho_P$  is derived.

**Proposition 8:**

Suppose the negative feedback loop  $F$  is designed for the metabolic network in Equation (4.40) or (4.42), such that the following LMIs have a positive solution  $P > 0$ :

$$\begin{bmatrix}
 -P + \sum_{i=1}^L A_{ij}^T P A_{ij} + C^T C & (A_j + FG_j)^T P & (A_j + FG_j)^T P B_v \\
 P(A_j + FG_j) & -P & 0 \\
 B_v^T P(A_j + FG_j) & 0 & B_v^T P B_v - \rho_P^2 I
 \end{bmatrix}, \text{ for } j=1, 2, \dots, M \quad (4.43)$$

Thus, the prescribed disturbance attenuation level  $\rho_P$  is achieved by the biochemical circuit design.

If the optimal robust filtering design is employed for the nonlinear stochastic metabolic network in Equation (4.40) or (4.42), then the control matrix  $F$  is specified to solve the following constrained optimization problem:

$$\begin{aligned}
 \rho_0 &= \min_F \rho_P \\
 \text{subject to } &P > 0 \text{ and LMIs in (4.43)}
 \end{aligned}
 \quad (4.44)$$

where  $\rho_0$  in Equation (4.44) is the disturbance (noise)-filtering ability of the optimally controlled metabolic network in Equation (4.40) or (4.42) (see **supplementary example 4**).

This robust biochemical circuit design example in a metabolic network illustrates that the proposed systematic method of biochemical circuit design produces metabolic networks that not only tolerate more fluctuations of the intrinsic random parameter but also filter more environmental disturbances. Systems metabolic engineering design is thus capable of improving the network robustness of metabolic networks and efficiently attenuating the effect of environmental noises. This approach may serve as a basis for drug designs against genetic perturbations, pathological environmental cues (such as infectious agents or chemical carcinogens), or both.

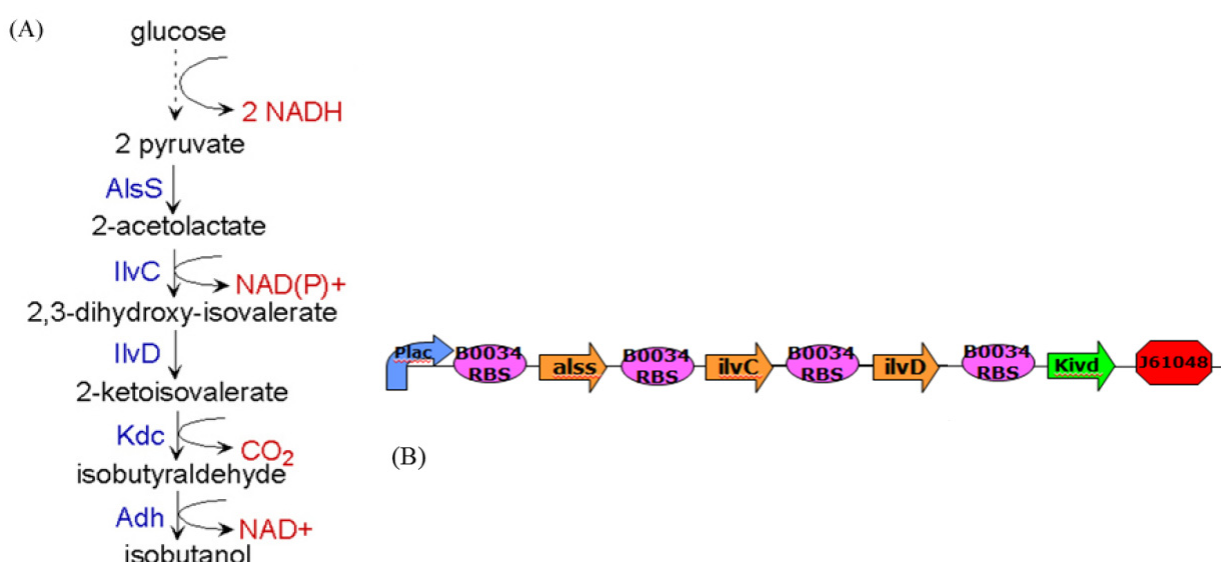
Network robustness is a systematic property that allows a metabolic network to maintain its biochemical function or to generate biochemical products despite external disturbance and intrinsic parametric fluctuation. It is from a class of fundamental and ubiquitously observed systems-level phenomena that cannot be understood by observing individual components only. Thus, a study of network robustness at one level is simply a study of why, when, and how metabolic networks function properly or otherwise [47,80,94]. If network robustness is not large enough to obtain sufficient intrinsic robustness to tolerate random parameter fluctuations and environmental robustness to filter environmental molecular noises, then a robust biochemical circuit design is necessary to improve robustness such that the network can maintain its function or phenotype. The proposed robust circuit design principles could potentially be used for robust biosynthetic network design with applications in drug design, gene therapy, and metabolic engineering. Future focus may be on robust chemical circuit design for such

applications or construction of other pathways by nanotechnology and metabolic engineering [28]. If network robustness, intrinsic robustness, and environmental robustness are taken into consideration in metabolic network engineering, such networks would function more reliably and efficiently.

With the exception of implementing feedback biochemical circuits in metabolic networks, systems synthetic biology techniques detailed in Section 3 could be combined with systems metabolic engineering approaches to engineer complete metabolic pathways or networks to produce biochemical products that could not be produced in the host cell. As an example, *E. coli* does not have non-fermentative pathways for isobutanol (see Figure 14(A)). Suppose a synthetic pathway in *E. coli* is required for the production of butanol as biofuel. In this application, it is necessary to engineer transcription and translation genes to produce the enzymes AlsS, IlvC, IlvD, and Kdc (KIVD), which are necessary to catalyze the isobutanol metabolic pathway (Figure 14(B)). The last enzyme of the metabolic pathway (Adh) is already present in *E. coli*. Promoters and RBSs with different regulatory abilities can be selected from the corresponding libraries. With adequate computation and simulation following the systems biology approach (see Section 2), a synthetic metabolic pathway could be designed and engineered (see Section 3) to achieve high-yield and high-specificity metabolic production of isobutanol.

**Figure 14.** Engineered synthetic metabolic pathway for isobutanol production in *E. coli*.

(A) Schematic representation of engineered isobutanol production pathway. (B) Engineered synthetic genetic circuit to generate the enzymes necessary for pathway in (A) for isobutanol production in *E. coli*.



## 5. Future Challenges in Systems Biology

Systems biology approaches to synthetic biology and metabolic engineering are built on molecular and genetic findings, and results of studies in omics fields such as genomics, proteomics, and metabolomics. The main concerns faced by systems biology are the complexity and dynamic character of biological systems, the vast quantities of biological data, and the fragmented nature of biological knowledge. These fragments of information need to be integrated with nonlinear stochastic dynamic

models at different system levels by suitable computational tools before they can be applied to system synthetic biology and systems metabolic engineering.

The first challenges of systems biology are how to enrich omics data (e.g., updating and improving protein microarray, ChIP-chip databases, and pathway database for systematic studies discussed in chapters 2–4) and develop more powerful computational tools for sophisticated data handling, advanced modeling, integrated analysis, and knowledge integration. A systems biology approach based on the integration of these enriched omics data and advanced computational models will enable us to predict the behaviors of biological systems more precisely. This will increase our understanding of the underlying molecular mechanisms and our ability to efficiently predict the effects of designed genetic circuits in metabolic engineering as well as the impact of perturbations on biological systems in drug treatment.

The second challenge is how to increase the capability of researchers to navigate and relate various data and knowledge resources using the integrated platform of bioinformatics, systems synthetic biology, and system metabolic engineering to enable innovations. Connecting genomics, molecular networks, and physiology will provide us with deeper understanding of how individual differences in the genome affect physiological processes through alterations in molecular networks. Therefore, predictive and preventive medicine based on network-based biomarkers inevitably lead to personalized medicine that may revolutionize healthcare in the future [7]. The third challenge of systems biology is how to integrate bioinformatics, gene circuit design, and metabolic engineering technologies for systems drug design in future predictive and preventative medicine.

Drastically increasing oil consumption, exhaustion of natural resources, and global warming are worldwide concerns at present. They have spurred research in the areas of bioenergy and biorefining, as well as microbial production of bulk chemicals and materials. Systems biology will play an important role in these undertakings, collaborating with systems synthetic biology and systems metabolic engineering to generate bio-based products to replace some, if not most, currently used chemicals and materials. The fourth challenge of systems biology is how to make these bio-based processes competitive.

## Supplementary Materials

Supplementary materials can be accessed at: <http://www.mdpi.com/2073-4409/2/4/635/S1>.

## Acknowledgments

The work was supported by the National Science Council of Taiwan, R.O.C. [NSC 102-2745-E-007-001-ASP].

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Joyce, A.R.; Palsson, B.O. The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell. Bio.* **2006**, *7*, 198–210.



2. Ideker, T.; Bafna, V.; Lemberger, T. Integrating scientific cultures. *Mol. Syst. Biol.* **2007**, *3*, 105–106.
3. Chuang, H.Y.; Lee, E.; Liu, Y.T.; Lee, D.; Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **2007**, *3*, 140–157.
4. Oti, M.; Snel, B.; Huynen, M.A.; Brunner, H.G. Predicting disease genes using protein-protein interactions. *J. Med. Genet.* **2006**, *43*, 691–698.
5. Peri, S.; Navarro, J.D.; Amanchy, R.; Kristiansen, T.Z.; Jonnalagadda, C.K.; Surendranath, V.; Niranjan, V.; Muthusamy, B.; Gandhi, T.K.B.; Gronborg, M.; *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **2003**, *13*, 2363–2371.
6. Zhang, S.H.; Jin, G.X.; Zhang, X.S.; Chen, L.N. Discovering functions and revealing mechanisms at molecular level from biological networks. *Proteomics* **2007**, *7*, 2856–2869.
7. Voit, E.O. *Computational Analysis of Biochemical Systems: a Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press: New York, NY, USA, 2000.
8. Chen, B.S.; Wu, C.C. On the calculation of signal transduction ability of signaling transduction pathways in intracellular communication: systematic approach. *Bioinformatics* **2012**, *28*, 1604–1611.
9. Kitano, H. Opinion—Cancer as a robust system: Implications for anticancer therapy. *Nat. Rev. Cancer* **2004**, *4*, 227–235.
10. Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F.J.; Doyle, J. Robustness of cellular functions. *Cell* **2004**, *118*, 675–685.
11. Storey, J.D.; Xiao, W.Z.; Leek, J.T.; Tompkins, R.G.; Davis, R.W. Significance analysis of time course microarray experiments. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12837–12842.
12. Cherry, J.M.; Adler, C.; Ball, C.; Chervitz, S.A.; Dwight, S.S.; Hester, E.T.; Jia, Y.K.; Juvik, G.; Roe, T.; Schroeder, M.; *et al.* SGD, *Saccharomyces* genome database. *Nucleic Acids Res.* **1998**, *26*, 73–79.
13. Gasch, A.P.; Spellman, P.T.; Kao, C.M.; Carmel-Harel, O.; Eisen, M.B.; Storz, G.; Botstein, D.; Brown, P.O. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell.* **2000**, *11*, 4241–4257.
14. Harbison, C.T.; Gordon, D.B.; Lee, T.I.; Rinaldi, N.J.; Macisaac, K.D.; Danford, T.W.; Hannett, N.M.; Tagne, J.B.; Reynolds, D.B.; Yoo, J.; *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **2004**, *431*, 99–104.
15. Teixeira, M.C.; Monteiro, P.; Jain, P.; Tenreiro, S.; Fernandes, A.R.; Mira, N.P.; Alenquer, M.; Freitas, A.T.; Oliveira, A.L.; Sa-Correia, I. The YEASTRACT database, a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2006**, *34*, D446–D451.
16. Stark, C.; Breitkreutz, B.J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID, a general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539.
17. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; *et al.* Gene ontology, Tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **2000**, *25*, 25–29.
18. Chen, H.C.; Lee, H.C.; Lin, T.Y.; Li, W.H.; Chen, B.S. Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle. *Bioinformatics* **2004**, *20*, 1914–1927.

19. Chen, B.S.; Wang, Y.C.; Wu, W.S.; Li, W.H. A new measure of the robustness of biochemical networks. *Bioinformatics* **2005**, *21*, 2698–2705.
20. Chang, W.C.; Li, C.W.; Chen, B.S. Quantitative inference of dynamic regulatory pathways via microarray data. *BMC Bioinf.* **2005**, *6*, 44–62.
21. Lin, L.H.; Lee, H.C.; Li, W.H.; Chen, B.S. Dynamic modeling of cis-regulatory circuits and gene expression prediction via cross-gene identification. *BMC Bioinf.* **2005**, *6*, 258–274.
22. Chang, Y.H.; Wang, Y.C.; Chen, B.S. Identification of transcription factor cooperativity via stochastic system model. *Bioinformatics* **2006**, *22*, 2276–2282.
23. Wu, W.S.; Li, W.H.; Chen, B.S. Computational reconstruction of transcriptional regulatory modules of the yeast cell cycle. *BMC Bioinf.* **2006**, *7*, 421–435.
24. Wu, W.S.; Li, W.H.; Chen, B.S. Identifying regulatory targets of cell cycle transcription factors using gene expression and ChIP-chip data. *BMC Bioinf.* **2007**, *8*, 188–205.
25. Chen, B.S.; Wang, Y.C. On the attenuation and amplification of molecular noise in genetic regulatory networks. *BMC Bioinf.* **2006**, *7*, 52–65.
26. Chen, B.-S.; Li, C.-W. Analysing microarray data in drug discovery using systems biology. *Expert Opin. Drug Discovery* **2007**, *2*, 755–768.
27. Chen, B.S.; Wu, W.S.; Wang, Y.C.; Li, W.H. On the Robust Circuit Design Schemes of Biochemical Networks, Steady-State Approach. *IEEE T Biomed. Circ. S* **2007**, *1*, 91–104.
28. Chen, B.S.; Chen, P.W. Robust Engineered Circuit Design Principles for Stochastic Biochemical Networks With Parameter Uncertainties and Disturbances. *IEEE T Biomed. Circ. S* **2008**, *2*, 114–132.
29. Chen, B.S.; Lin, Y.P. A unifying mathematical framework for genetic robustness, environmental robustness, network robustness and their trade-off on phenotype robustness in biological networks Part I, Gene regulatory networks in systems and evolutionary biology. *Evol. Bioinform* **2013**, *9*, 43–68.
30. Chen, B.S.; Lin, Y.P. A Unifying mathematical framework for genetic robustness, environmental robustness, Network robustness and their tradeoff on phenotype robustness in biological networks Part II, Ecological networks. *Evol. Bioinform* **2013**, *9*, 69–85.
31. Wang, Y.C.; Huang, S.H.; Lan, C.Y.; Chen, B.S. Prediction of phenotype-associated genes via a cellular network approach, a candida albicans infection case study. *PloS One* **2012**, *7*, e35339.
32. Kuo, Z.Y.; Chuang, Y.J.; Chao, C.C.; Liu, F.C.; Lan, C.Y.; Chen, B.S. Identification of infection- and defense-related genes via a dynamic host-pathogen interaction network using a candida albicans -zebrafish infection model. *J. Innate Immun.* **2013**, *5*, 137–152.
33. Tu, C.T.; Chen, B.S. New measurement methods of network robustness and response ability via microarray data. *PloS One* **2013**, *8*, e55230.
34. Tu, C.T.; Chen, B.S. On the Increase in Network Robustness and Decrease in Network Response Ability During the Aging Process, A Systems Biology Approach via Microarray Data. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2013**, *10*, 468–480.
35. Csete, M.E.; Doyle, J.C. Reverse engineering of biological complexity. *Science* **2002**, *295*, 1664–1669.

36. Lee, K.H.; Park, J.H.; Kim, T.Y.; Kim, H.U.; Lee, S.Y. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol. Syst. Biol.* **2007**, *3*, 149.
37. Park, J.H.; Lee, S.Y.; Kim, T.Y.; Kim, H.U. Application of systems biology for bioprocess development. *Trends Biotechnol* **2008**, *26*, 404–412.
38. Chen, B.S.; Chang, Y.T.; Wang, Y.C. Robust H infinity-stabilization design in gene networks under stochastic molecular noises, fuzzy-interpolation approach. *IEEE Trans. Syst. Man Cybern B Cybern* **2008**, *38*, 25–42.
39. Chu, L.H.; Chen, B.S. Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets. *BMC Syst. Biol.* **2008**, *2*, 56.
40. Chen, B.S.; Chang, C.H.; Chuang, Y.J. Robust model matching control of immune systems under environmental disturbances, Dynamic game approach. *J. Theor. Biol.* **2008**, *253*, 824–837.
41. Chen, B.S.; Yang, S.K.; Lan, C.Y.; Chuang, Y.J. A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. *BMC Med. Genomics* **2008**, *1*, 46.
42. Chen, B.S.; Chen, P.W. On the estimation of robustness and filtering ability of dynamic biochemical networks under process delays, internal parametric perturbations and external disturbances. *Math. Biosci.* **2009**, *222*, 92–108.
43. Wang, Y.C.; Lan, C.Y.; Hsieh, W.P.; Murillo, L.A.; Agabian, N.; Chen, B.S. Global screening of potential *Candida albicans* biofilm-related transcription factors via network comparison. *BMC Bioinf.* **2010**, *11*, 53–73.
44. Li, C.W.; Chen, B.S. Identifying Functional Mechanisms of Gene and Protein Regulatory Networks in Response to a Broader Range of Environmental Stresses. *Comp. Funct. Genom.* **2010**, *2010*, 408705.
45. Gardner, T.S.; Cantor, C.R.; Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **2000**, *403*, 339–342.
46. Elowitz, M.B.; Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **2000**, *403*, 335–338.
47. Hasty, J.; McMillen, D.; Collins, J.J. Engineered gene circuits. *Nature* **2002**, *420*, 224–230.
48. McDaniel, R.; Weiss, R. Advances in synthetic biology, on the path from prototypes to applications. *Curr. Opin. Biotechnol.* **2005**, *16*, 476–483.
49. Endy, D. Foundations for engineering biology. *Nature* **2005**, *438*, 449–453.
50. McAdams, H.H.; Arkin, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **1999**, *15*, 65–69.
51. Batt, G.; Yordanov, B.; Weiss, R.; Belta, C. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics* **2007**, *23*, 2415–2422.
52. Chen, B.S.; Chang, C.H.; Lee, H.C. Robust synthetic biology design, stochastic game theory approach. *Bioinformatics* **2009**, *25*, 1822–1830.
53. Chen, B.S.; Wu, C.H. A systematic design method for robust synthetic biology to satisfy design specifications. *BMC Syst. Biol.* **2009**, *3*, 66–83.
54. Chen, B.S.; Chen, P.W. GA-based Design Algorithms for the Robust Synthetic Genetic Oscillators with Prescribed Amplitude, Period and Phase. *Gene Regul. Syst. Bio.* **2010**, *4*, 35–52.

55. Chen, B.S.; Wu, C.H. Robust optimal reference-tracking design method for stochastic synthetic biology systems, T-S fuzzy approach. *IEEE T. Fuzzy Syst.* **2010**, *18*, 1144–1159.
56. Wu, C.H.; Zhang, W.H.; Chen, B.S. Multiobjective H-2/H-infinity synthetic gene network design based on promoter libraries. *Math. Biosci.* **2011**, *233*, 111–125.
57. Wu, C.H.; Lee, H.C.; Chen, B.S. Robust synthetic gene network design via library-based search method. *Bioinformatics* **2011**, *27*, 2700–2706.
58. Chen, B.S.; Hsu, C.Y.; Liou, J.J. Robust design of biological circuits, Evolutionary systems biology approach. *J. Biomed. Biotechnol.* **2011**, *2011*, 30423.
59. Chen, B.S.; Lin, Y.P. A unifying mathematical framework for genetic robustness, Environmental robustness, Network robustness and their trade-offs on phenotype robustness in biological networks. Part III, Synthetic gene networks in synthetic biology. *Evol. Bioinform* **2013**, *9*, 87–109.
60. Chen, B.S.; Hsu, C.Y. Robust synchronization control scheme of a population of nonlinear stochastic synthetic genetic oscillators under intrinsic and extrinsic molecular noise via quorum sensing. *BMC Systems Biology* **2012**, *6*, 136–150.
61. Kim, P.J.; Lee, D.Y.; Kim, T.Y.; Lee, K.H.; Jeong, H.; Lee, S.Y.; Park, S. Metabolite essentiality elucidates robustness of Escherichia coli metabolism. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13638–13642.
62. Segre, D.; Vitkup, D.; Church, G.M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15112–15117.
63. Chandran, D.; Copeland, W.B.; Sleight, S.C.; Sauro, H.M. Mathematical modeling and synthetic biology. *Drug Discovery Today* **2008**, *5*, 299–309.
64. Zhou, S.; Iverson, A.G.; Grayburn, W.S. Engineering a native homoethanol pathway in Escherichia coli B for ethanol production. *Biotechnol. Lett.* **2008**, *30*, 335–342.
65. Shlomi, T.; Eisenberg, Y.; Sharan, R.; Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* **2007**, *3*, 101–107.
66. Jamshidi, N.; Palsson, B.O. Formulating genome-scale kinetic models in the post-genome era. *Mol. Syst. Biol.* **2008**, *4*, 171–186.
67. Bonneau, R.; Facciotti, M.T.; Reiss, D.J.; Schmid, A.K.; Pan, M.; Kaur, A.; Thorsson, V.; Shannon, P.; Johnson, M.H.; Bare, J.C.; *et al.* A predictive model for transcriptional control of physiology in a free living cell. *Cell* **2007**, *131*, 1354–1365.
68. Kuepfer, L.; Sauer, U.; Parrilo, P.A. Efficient classification of complete parameter regions based on semidefinite programming. *BMC Bioinf.* **2007**, *8*, 12–22.
69. Yao, C.W.; Hsu, B.D.; Chen, B.S. Constructing gene regulatory networks for long term photosynthetic light acclimation in Arabidopsis thaliana. *BMC Bioinf.* **2011**, *12*, 335.
70. Wang, Y.C.; Chen, B.S. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med. Genomics* **2011**, *4*, 2–16.
71. Wang, Y.C.; Chen, B.S. Integrated cellular network of transcription regulations and protein-protein interactions. *BMC Syst. Biol.* **2010**, *4*, 20–36.
72. Shiue, E.; Prather, K.L.J. Synthetic biology devices as tools for metabolic engineering. *Biochem. Eng. J.* **2012**, *65*, 82–89.
73. Chen, B.S.; Zhang, W.H. Stochastic H(2)/H(infinity) control with state-dependent noise. *IEEE T Automat. Contr.* **2004**, *49*, 45–57.

74. Zhang, W.H.; Chen, B.S.; Tseng, C.S. Robust H-infinity filtering for nonlinear stochastic systems. *IEEE T Signal. Proces.* **2005**, *53*, 589–598.
75. Zhang, W.H.; Chen, B.S. State feedback H(infinity) control for a class of nonlinear stochastic systems. *Siam J. Control. Optim* **2006**, *44*, 1973–1991.
76. Hughes, T.R.; Marton, M.J.; Jones, A.R.; Roberts, C.J.; Stoughton, R.; Armour, C.D.; Bennett, H.A.; Coffey, E.; Dai, H.Y.; He, Y.D.D.; *et al.* Functional discovery via a compendium of expression profiles. *Cell* **2000**, *102*, 109–126.
77. Johansson, R. *System Modeling & Identification*; Prentice-Hall, Inc.; London, UK, 1993.
78. Li, C.W.; Chu, Y.H.; Chen, B.S. Construction and clarification of dynamic gene regulatory network of cancer cell cycle via microarray data. *Cancer Informatics* **2006**, *2*, 223–241.
79. Boyd, S.P. *Linear Matrix Inequalities in System and Control Theory*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1994.
80. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **2004**, *5*, 826–837.
81. Chen, B.S.; Chen, W.H.; Wu, H.L. Robust H-2/H-infinity global linearization filter design for nonlinear stochastic systems. *IEEE T Circuits-I* **2009**, *56*, 1441–1454.
82. Chen, B.S.; Lee, T.S.; Feng, J.H. A nonlinear H-Infinity control design in robotic systems under parameter perturbation and external disturbance. *Int J. Control.* **1994**, *59*, 439–461.
83. Chen, B.S.; Tseng, C.S.; Uang, H.J. Mixed H-2/H-infinity fuzzy output feedback control design for nonlinear dynamic systems, An LMI approach. *IEEE T. Fuzzy Syst.* **2000**, *8*, 249–265.
84. Stricker, J.; Cookson, S.; Bennett, M.R.; Mather, W.H.; Tsimring, L.S.; Hasty, J. A fast, robust and tunable synthetic gene oscillator. *Nature* **2008**, *456*, 516–519.
85. Tiggles, M.; Marquez-Lago, T.T.; Stelling, J.; Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **2009**, *457*, 309–312.
86. Basu, S.; Gerchman, Y.; Collins, C.H.; Arnold, F.H.; Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **2005**, *434*, 1130–1134.
87. Tsai, T.Y.; Choi, Y.S.; Ma, W.; Pomeroy, J.R.; Tang, C.; Ferrell, J.E., Jr. Robust, Tunable biological oscillations from interlinked positive and negative feedback loops. *Science* **2008**, *321*, 126–129.
88. Alon, U. *An Introduction to Systems Biology, Design Principles of Biological Circuits*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2007.
89. Ellis, T.; Wang, X.; Collins, J.J. Diversity-based, Model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol* **2009**, *27*, 465–471.
90. Mondragon-Palomino, O.; Danino, T.; Selimkhanov, J.; Tsimring, L.; Hasty, J. Entrainment of a Population of Synthetic Genetic Oscillators. *Science* **2011**, *333*, 1315–1319.
91. Ghosh, S.; Matsuoka, Y.; Asai, Y.; Hsin, K.Y.; Kitano, H. Software for systems biology, from tools to integrated platforms. *Nat. Rev. Genet.* **2011**, *12*, 821–832.
92. Tsai, K.Y.; Wang, F.S. Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics* **2005**, *21*, 1180–1188.
93. Chen, B.S.; Tseng, C.S.; Uang, H.J. Robustness design of nonlinear dynamic systems via fuzzy linear control. *IEEE T. Fuzzy Syst.* **1999**, *7*, 571–585.

94. Kitano, H. Computational systems biology. *Nature* **2002**, *420*, 206–210.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).