ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

F. Cacace A. Germani P. Palumbo

A STUDY ON OBSERVER-BASED ALGORITHMS TO INFER INFORMATION FROM GENE EXPRESSION DATA

R. 13, 2012

Filippo Cacace – Università Campus Biomedico di Roma, Via Alvaro del Portillo, 21 - 00128 Roma, Italy. f.cacace@unicampus.it.

Alfredo Germani – Dipartimento di Ingegneria Elettrica e dell'Informazione, Università degli Studi dell'Aquila, Italy. alfredo.germani@univaq.it.

Pasquale Palumbo – Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Consiglio Nazionale delle Ricerche (IASI-CNR), Viale manzoni 30, 00185 Rome, Italy. pasquale.palumbo@iasi.cnr.it.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.cnr.it
URL: http://www.iasi.cnr.it

Abstract

In this paper a mathematical tool is presented, to estimate unknown variables of transcription networks, according to a set of measurements of the transcriptional activity of promoters. The approach is based on the use of the mathematical model of the network under investigation and on the state-reconstruction technique known as "state-observer", borrowed from the control theory. To this aim, besides the general case, the network motif of the Multi-Output Feed-Forward Loop (MO-FFL) will be investigated in details. Simulations show the effectiveness of the proposed approach in a wide range of possible critical frameworks, such as only one target gene measurements, non-smooth input perturbations, noisy measurements and model parameter uncertainties.

Key words: State-observer; Nonlinear systems; Systems Biology.

A STUDY ON OBSERVER-BASED ALGORITHMS TO INFER INFORMATION FROM GENE EXPRESSION DATA

FILIPPO CACACE*, ALFREDO GERMANI†, AND PASQUALE PALUMBO‡

Abstract. In this paper a mathematical tool is presented, to estimate unknown variables of transcription networks, according to a set of measurements of the transcriptional activity of promoters. The approach is based on the use of the mathematical model of the network under investigation and on the state-reconstruction technique known as 'state-observer', borrowed from the control theory. To this aim, besides the general case, the network motif of the Multi-Output Feed-Forward Loop (MO-FFL) will be investigated in details. Simulations show the effectiveness of the proposed approach in a wide range of possible critical frameworks, such as only one target gene measurements, non-smooth input perturbations, noisy measurements and model parameter uncertainties.

Key words. State-observer, Nonlinear systems, Systems Biology

1. Introduction. The cell is the basic unit of life able to display a wide range of autonomous functions in a flexible and robust way [21, 14]. The genome of a cell actively contributes to the modulation of the timing and amount of proteins in response to internal and environmental signals. Such appropriate behavior is the result of many factors acting on different hierarchical levels and time scales, including the regulation exerted by a family of proteins, called *transcription factors*, able to selectively bind to DNA regions and activate or repress the expression of downstream target genes in a coordinated fashion. Roughly speaking, we may say that gene expression is to a large extent regulated at the level of mRNA, whose abundance is regulated by transcription factors.

One of the main goals of systems biology is to provide computational support for the formulation of new biological hypotheses on the biochemical mechanisms underlying the observed cell behavior and their experimental validation [13]. In fact, the integration of computational modeling, system analysis and quantitative experiments, has proved to be very successful in providing the field of molecular biology with new paradigms and insights [22, 3, 11].

Modern techniques, such as real-time polymerase chain reaction (RT-PCR, [18]) or the use of a Green Fluorescent Protein (GFP) as a reporter gene [24], allow to obtain measurements at short sampling intervals, providing, as a matter of fact, larger time series and better temporal resolution. More recently, the development of "deep sequencing" techniques [23] has provided ground breaking methods, featured by large throughput and almost total precision. These new techniques allow to quantify the expression level with a precision up to a few molecules, and they have very low measurement noise. It has to be stressed, however, that these techniques, as well as the widely established micro-arrays, share the drawback of only measuring the mRNA concentrations, instead of the corresponding proteins. Indeed, the common problem of these experimental approaches is the difficulty of measuring the actual protein concentration level during the living cell activity.

*Università Campus Biomedico di Roma, Via Álvaro del Portillo, 21 - 00128 Roma, Italy, f.cacace@unicampus.it

†Dipartimento di Ingegneria Elettrica e dell'Informazione, Universita degli Studi dell'Aquila, Italy, alfredo.germani@univaq.it

‡Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", IASI-CNR, Viale Manzoni 30 00185 Roma, Italy, pasquale.palumbo@iasi.cnr.it

The improvements in acquiring reliable data motivates the use of more sophisticated mathematical tools of data processing to infer information from the biological network. In this context we evaluate the use of a state-reconstruction technique known as 'state-observer', borrowed from the control theory. Given a model of a system, generally expressed through a set of Ordinary Differential Equations (ODEs), a state observer allows to track the time course of the "hidden variables" of the system by measuring the "visible" variables. This capability can be of the utmost utility for the "reverse engineering" approach of deducing the structure of a Genetic Transcription Network (GTN) from available data. In this framework, one crucial purpose of the present contribution will be to estimate the time course of transcription factors levels by means of the indirect measurements on their corresponding mRNA.

The observer-based approach is an interesting possibility to state estimation problems when a dynamical model of the system under investigation is available. Although devised for real-time estimation problems, its robustness may reveal to be successful also for off-line state variable estimations, parameters identification, and tracking of systems with complex or highly nonlinear dynamics. It has to be stressed that, from this point of view, the observer may be seen as complementary to other statistical approaches which make use as well of the model equations such as in [12], where the MCMC method has been cleverly applied to estimate the unknown parameters by suitably exploiting the ODE of the HIV model with time-varying parameters.

The focus of the paper is mainly methodological, hence we consider with some details the mathematical conditions under which the biological problems of interest can be solved by means of observer-based approaches. However, some preliminary remarks may clarify the issues related to the application of observer-based techniques to real biological scenarios:

- Observers are designed for deterministic systems, but biological data are usually noisy. The use of observer-based techniques in GTNs have been motivated by the following reasons. From one hand, high-frequency measurements with a very low level of noise are now available (see, e.g. [18, 23]). From the other hand, new results have been published in the control system society on the use of observers for stochastic systems [Ahrens & Khalil, 2009],[6]. These results show the robustness of this approach also in the case of noisy systems. Moreover, we have tested the efficacy of our approach on simulated noisy measurements.
- We consider continuous-time systems of ODEs, but real measurements are discrete, thus a discretized version of the system and of the observer is needed. We do not deal with this issue here, but this discretization is possible by means of recently proposed techniques [5], that are well suited for the large discretization intervals found in biological scenarios. This approach has already been applied in a biological framework [4]. The mathematical conditions of observability can be equivalently stated on continuous- and discrete-time systems, but the formulation of the model is more intuitive in a time-continuous framework.

The paper is organized as follows. Basic notations and mathematical modeling of gene transcription networks by means of ordinary, nonlinear differential equations are described in Section 2. An outline of the structure and theory of the adopted state observer for nonlinear systems is contained in Section 3. Section 4 deals with the application of the observer to different scenarios. Simulation results are reported in Section 5. Conclusions follow.

2. TRANSCRIPTION NETWORKS AND PROBLEM SETTING. Transcription networks describe the transcriptional regulation of genes. A way to depict these networks is to use graphs, where each node represents either a protein or a gene that encodes for a protein, and arrows refer to *which* transcription factor regulates *which* gene. In this paper we refer to biological variables using capital letters for proteins and plain letters for the corresponding gene/gene product. Then, $x_i(t)$ is the concentration of the mRNA of the gene x_i that encodes for protein X_i. According to the usual meaning, the notation $X \to y$ denotes that the transcription factor X is an activator of gene y, whereas $X \dashv y$ denotes that X is a repressor for y. Often, the intermediate gene is omitted, and we can write $X \to Y$ to denote that the transcription factor X binds the promoter of a gene that encodes for Y, or respectively, $X \dashv Y$ in the case of a repressor.

Transcription networks can be modeled by means of nonlinear Ordinary Differential Equations [2, 9, 20]. Let us assume to have a network of N genes/proteins. The dynamics of a gene product x_i, $i = 1, \dots, N$, regulated by transcription factors $X_1, \dots, X_N$ and by an external input u_i, and of the corresponding protein X_i, is usually modeled as [9]

$$\begin{aligned} \frac{dx_i(t)}{dt} &= -\lambda_{x_i} x_i(t) + \varphi_i(X_1, \dots, X_n) + u_i(t), \\ \frac{dX_i(t)}{dt} &= -\lambda_{X_i} X_i(t) + p_{X_i} x_i(t), \end{aligned} \tag{2.1}$$

Here, λ_{x_i} and λ_{X_i} are the degradation rates of, respectively, x_i and X_i; p_{X_i} is the translation rate from x_i to X_i, and $\varphi_i(X_1, \dots, X_n)$ summarizes the transcriptional control, possibly depending by all the proteins of the network. Notice that both degradations as well as the protein translation are modeled by linear terms (with respect to the state variables), the nonlinear part being concentrated in the term $\varphi_i(X_1, \dots, X_n)$. This function may assume different forms depending on how the transcription factors X_j, $j = 1, \dots, N$, interact with the promoter of x_i and between them. A common assumption is that the different transcription factors regulating gene x_i are independent: in this case, function $\varphi_i(\cdot)$ may be written as the product of the activation/repression effects of the transcription factors $X_1, \dots, X_N$, [17]:

$$\varphi_i(X_1, \dots, X_n) = V_i^0 \prod_{X_j \in A_i} \varphi_{a_i}(X_j) \prod_{X_j \in R_i} \varphi_{r_i}(X_j), \tag{2.2}$$

where V_i^0 is the basal transcription rate of gene x_i (i.e. when there is no action of activators and inhibitors), A_i and R_i are the set of activators and repressors regulating gene x_i, and functions φ_{a_i} and φ_{r_i} are usually modeled as S-shaped Hill functions of the form:

$$\varphi_{a_i}(X_j) = 1 + (\bar{V}_{ij} - 1)\frac{X_j^{\nu_{ij}}(t)}{X_j^{\nu_{ij}}(t) + K_{ij}^{\nu_{ij}}}, \qquad \varphi_{r_i}(X_j) = \frac{K_{ij}^{\nu_{ij}}}{X_j^{\nu_{ij}}(t) + K_{ij}^{\nu_{ij}}}, \tag{2.3}$$

where $\bar{V}_{ij} > 1$ is the V_i^0 multiplicative factor giving the maximal transcription rate exerted by X_j, K_{ij} is the threshold of X_j around which the function switches from low to high level or viceversa, and ν_{ij} governs the steepness of the transition.

A common simplification of model (2.1) descends from the different time-scales in the dynamics of genes and proteins in a transcription network. For example, when external signals rapidly activate existing proteins that are transcription factors, they

cause an almost immediate change in the transcription rate of the target mRNA. Since protein translation is a slower process, it is possible to express the protein dynamics by assuming that mRNA concentration has reached the steady state (see, e.g. [1]). In this case, the x_i equation in (2.1) may be treated as an algebraic constraint (since $dx_i/dt = 0$), leading to the following dynamics for the protein kinetics:

$$\frac{dX_i(t)}{dt} = -\lambda_{X_i} X_i(t) + \frac{p_{X_i}}{\lambda_{x_i}} \varphi_i(X_1, \ldots, X_n) + \frac{p_{X_i}}{\lambda_{x_i}} u_i(t), \qquad i = 1, \ldots, N. \tag{2.4}$$

In spite of the simplifications adopted, model (2.4) is very often found in the literature (e.g. [1, 17, 20]).

Based on the above mentioned mathematical models, this paper investigates the problem of inferring the time course of some (possibly all) state variables by means of a reduced set of measurements. From a biological point of view we can distinguish between a couple of significative frameworks. The first concerns a model described by eqs.(2.1), where the only measurements are the mRNA concentrations (state variables x_i, $i = 1, \ldots, N$), and we want to estimate the time-course of proteins (state variables X_i, $i = 1, \ldots, N$). As previously mentioned in the Introduction, this is a very common case, since protein evolutions are usually achieved by suitably smoothing the corresponding mRNA measurements [19]. The second framework concerns the simplified model described by eqs.(2.4), according to which we will assume that some gene products (i.e. some state variables) are not measured at all, and will be estimated by means of other measurements.

Both cases will be approached by using a nonlinear state observer [7, 8], whose theory is briefly recap in the following section. Motivations for using such an observer rely on its computational simplicity and wide versatility, since it applies to multiple-input/multiple-output systems; moreover the observer convergence can be proved under general conditions that are usually satisfied.

Besides investigating the problem in a very general setting, the application to a particular case of *network motifs* will be treated in details. Network motifs are major biological features of gene networks [2, 15]: they are small subset of basic building-block regulatory "circuits", able to perform a large variety of biological functions alone or in combinations with other circuits, depending on the specific environmental and internal conditions. Among the most frequently encountered network motifs, in many organisms, are the *Feed-Forward Loops* (FFLs) [2]. The FFL motif is composed of a transcription factor X that regulates another transcription factor Y and both X and Y regulates a gene z or a number of target genes z_i, $i = 1, \ldots, m$ (*Multi Output Feed-Forward Loop Motif*, MO-FFL), as shown in Figure 2.1. Furthermore, gene X is usually assumed to be part of a negative autoregulation loop. Examples of FFLs can be found in the regulatory circuit of the *lac* operon [16].

According to eqs.(2.1) and to the network topology of Fig.2.1 with transcription factors X and Y regulating the target genes z_i by means of the product of Hill

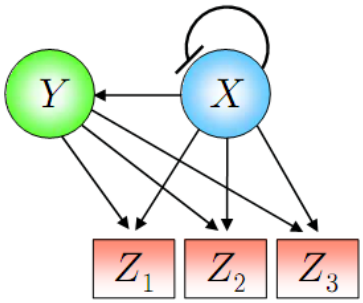

FIG. 2.1. *Three target genes MO-FFL motif with a negative autoregulation loop for gene X.*

functions like (2.2-2.3), we can write the following equations:

$$
\begin{aligned}
&\dot{x}(t) = -\lambda_x x(t) + \varphi_x(X) + u(t)\\
&\dot{X}(t) = -\lambda_X X(t) + p_X x(t)\\
&\dot{y}(t) = -\lambda_y y(t) + \varphi_y(X)\\
&\dot{Y}(t) = -\lambda_Y Y(t) + p_Y y(t)\\
&\dot{z}_i(t) = -\lambda_{z_i} z_i(t) + \varphi_{z_i}(X,Y), \qquad i = 1,\dots,m\\
&\dot{Z}_i(t) = -\lambda_{Z_i} Z_i(t) + p_{Z_i} z_i(t)
\end{aligned}
\tag{2.5}
$$

with:

$$
\varphi_x(X) = V_x^0 \frac{K_{xx}^{\nu_{xx}}}{X^{\nu_{xx}}(t) + K_{xx}^{\nu_{xx}}}, \quad \varphi_y(X) = V_y^0 \left(1 + (\bar{V}_{yx} - 1)\frac{X^{\nu_{yx}}(t)}{X^{\nu_{yx}}(t) + K_{yx}^{\nu_{yx}}}\right) \tag{2.6}
$$

and

$$
\varphi_{z_i}(X,Y) = V_{z_i}^0 \left(1 + (\bar{V}_{z_i x} - 1)\frac{X^{\nu_{z_i x}}(t)}{X^{\nu_{z_i x}}(t) + K_{z_i x}^{\nu_{z_i x}}}\right)\left(1 + (\bar{V}_{z_i y} - 1)\frac{Y^{\nu_{z_i y}}(t)}{Y^{\nu_{z_i y}}(t) + K_{z_i y}^{\nu_{z_i y}}}\right). \tag{2.7}
$$

Perturbations occur in the system by means of the master gene x, whose equation is therefore endowed with the input $u(t)$.

In this setting, the approach that we pursue is to estimate the time course of the master transcription factors X and Y by using only the mRNA measurements x, y and z_i, $i = 1,\dots,m$.

On the other hand, according to the simplifying assumptions leading to eqs.(2.4), the above system (2.5) reduces to:

$$
\begin{aligned}
&\frac{dX(t)}{dt} = -\lambda_X X(t) + \frac{p_X}{\lambda_x}\varphi_x(X) + \frac{p_X}{\lambda_x}u(t),\\
&\frac{dY(t)}{dt} = -\lambda_Y Y(t) + \frac{p_Y}{\lambda_y}\varphi_y(X)\\
&\frac{dZ_i(t)}{dt} = -\lambda_{Z_i} Z_i(t) + \frac{p_{Z_i}}{\lambda_{z_i}}\varphi_{z_i}(X,Y), \qquad i = 1,\dots,m
\end{aligned}
\tag{2.8}
$$

In this case, the observer-based algorithm helps to reconstruct the time-course of the gene products X and Y by only measuring the target genes Z_i.

3. NONLINEAR STATE OBSERVERS. In this Section we briefly provide some details on the nonlinear observer used in the next section. Convergence results are presented under a bounded-input bounded-state (BIBS) stability property that always holds for biological systems. More specifically, we deal with semiglobal observers, that ensure convergence to zero of the observation error when the state is confined to a bounded set.

3.1. Single-input/single-output case. Given a nonlinear system

$$\begin{aligned} \dot{x}(t) &= f\big(x(t)\big) + g\big(x(t)\big)u(t), \qquad x(0) = x_0, \\ y(t) &= h\big(x(t)\big), \end{aligned} \tag{3.1}$$

with the state $x(t) \in \Omega \subseteq \mathbb{R}^n$, the input $u(t)$ and the output $y(t)$ both scalar, h a $C^k(\Omega)$ scalar function, f a $C^k(\Omega)$ vector field, with k an integer that allows all differentiations needed, the *observability map* $\Phi(x)$ is defined as:

$$\Phi(x) = \begin{bmatrix} h(x) \\ L_f h(x) \\ \vdots \\ L_f^{n-1} h(x) \end{bmatrix}, \tag{3.2}$$

where $L_f^j h(x)$ is the j-th order Lie derivative defined as:

$$L_f^j h(x) = L_f\big(L_f^{j-1} h(x)\big), \qquad \text{and} \qquad L_f h(x) = \sum_{i=1}^{n} \frac{\partial h(x)}{\partial x_i} f_i(x). \tag{3.3}$$

The *observation relative degree* associated to the triple $\big(f(x), g(x), h(x)\big)$ is defined as the integer r such that:

$$\begin{aligned} &L_g L_f^s h(x) = 0, \qquad s = 0, 1, \ldots, r-2, \\ &L_g L_f^{r-1} h(x) \neq 0. \end{aligned} \tag{3.4}$$

Theorem 1, [7]. Consider a BIBS stable system (3.1) with no forcing inputs (i.e. $u(t) \equiv 0$) and assume that:

i. $\Phi(x)$ is a diffeomorphism;
ii. $L_f^n h\big(\Phi^{-1}(z)\big)$ is uniformly Lipschitz in $\Phi(\Omega)$.

Then there exists a gain vector $K \in \mathbb{R}^{n\times 1}$ such that the solution of the system equations

$$\dot{\hat{x}} = f(\hat{x}) + Q^{-1}\big(\hat{x}\big) K\big(y - h(\hat{x})\big), \qquad \hat{x}(0) = \bar{x}, \tag{3.5}$$

with $Q(x) = d\Phi/dx$ the observability matrix, exponentially converges to $x(t)$, whatever is the initial state estimate $\bar{x} \in \Omega$. □

Theorem 2, [7]. Consider a BIBS stable system (3.1) with uniformly bounded input u and assume that:

i. $\Phi(x)$ is a diffeomorphism;
ii. $L_f^n h\big(\Phi^{-1}(z)\big)$ and $L_g L_f^{n-1} h\big(\Phi^{-1}(z)\big)$ are uniformly Lipschitz in $\Phi(\Omega)$;
iii. the observation relative degree associated to the triple $\big(f(x), g(x), h(x)\big)$ is n;

Then there exists a gain vector $K \in \mathbb{R}^{n\times 1}$ such that the solution of the system equations

$$\dot{\hat{x}} = f(\hat{x}) + g(\hat{x})u + Q^{-1}\left(\hat{x}\right)K\big(y - h(\hat{x})\big), \qquad \hat{x}(0) = \bar{x}, \tag{3.6}$$

exponentially converges to $x(t)$, whatever is the initial observed state $\bar{x} \in \Omega$.

More in details, the gain matrix K may be designed in order to assign the observer error dynamics (see [7]).

It is easy to show that under the hypothesis on the relative degree (Theorem 2, item iii) the observability map $\Phi(x)$ yields a vector whose components are the output $y(t)$ together with its $n-1$ time derivatives. When the relative degree hypothesis is not satisfied the mapping $\Phi(x)$ can be modified in order to still satisfy the above property, but in this case $\Phi(x)$ may depend on the input $u(t)$ and its time derivatives. This makes more difficult to verify the convergence property of the observer. Moreover, the time derivatives of $u(t)$ are generally not available in practical situations.

As for the implementation of the observers (3.5-3.6), it can be noticed that

- the implementation does not require the knowledge of the inverse of the observability map $\Phi^{-1}(z)$, or that of $L_f^n h(\Phi^{-1}(z))$ and $L_g L_f^{n-1}(\Phi^{-1}(z))$;
- the inverse of matrix Q needs not to be computed. Indeed, name $v = Q^{-1}K$. What is required is actually the computation of vector v which comes from the solution of the linear system $Qv = K$, whose computational burden is smaller than the one involved in the inversion of matrix Q.

3.2. Multiple-input/multiple-output case. Given a nonlinear system

$$\begin{aligned} \dot{x}(t) &= f\big(x(t)\big) + g\big(x(t)\big)u(t), \qquad x(0) = x_0, \\ y(t) &= h\big(x(t)\big), \end{aligned} \tag{3.7}$$

with $x(t) \in \Omega \subseteq \mathbb{R}^n$, $u(t) \in U \subseteq \mathbb{R}^p$, $y(t) \in \mathbb{R}^q$, h a $C^k(\Omega)$ vector function, f a $C^k(\Omega)$ vector field, and $g(x) = [g_1(x), \ldots, g_p(x)]$ a matrix whose columns are $C^k(\Omega)$ vector fields, with k an integer that allows all differentiations needed, an observability map from the output to the state variables can be defined as follows (see [8] for more details). Let $\bar{s} = \{s_1, \ldots, s_q\}$ be a multi-index such that $\sum_{j=1}^{q} s_j = n$, and let

$$\Phi_j^{s_j}(x) = \begin{bmatrix} h_j(x) & L_f h_j(x) & \cdots & L_f^{s_j-1} h_j(x) \end{bmatrix}^T \tag{3.8}$$

be the vector function from $\mathbb{R}^n$ to $\mathbb{R}^{s_j}$ obtained by taking the j-th component of the output and its first $s_j - 1$ Lie derivatives. Consider the following square map

$$\Phi_{\bar{s}}(x) = \begin{bmatrix} \Phi_1^{s_1 T}(x) & \cdots & \Phi_q^{s_q T}(x) \end{bmatrix}^T. \tag{3.9}$$

Denoting $Y_{\bar{s}}$ the vector of output derivatives $Y_{\bar{s}} = [y_1 \ \cdots \ y_1^{(s_1-1)} \ \cdots \ y_q \ \cdots \ y_q^{(s_q-1)} \ ]^T$, if $u(t) \equiv 0$ it is $\Phi_{\bar{s}}(x(t)) = Y_{\bar{s}}(t)$.

If $\Phi_{\bar{s}}(x)$ is a diffeomorphism in an open set that contains Ω, it is possible to reconstruct the state x from the knowledge of vector $Y_{\bar{s}}$. In this case, the Jacobian associated to the observability map

$$Q_{\bar{s}}(x) = \frac{\partial \Phi_{\bar{s}}(x)}{\partial x} \tag{3.10}$$

is nonsingular in Ω, and the inverse map of $z = \Phi_{\bar{s}}(x)$ exists in $\Phi_{\bar{s}}(\Omega)$. Although in general such map is difficult to obtain, its Jacobian can be easily computed as $Q_{\bar{s}}^{-1}(x)$.

In analogy with the definition of the scalar case, for each component h_j of the output function we define the observation relative degree r_j as the smallest integer such that $L_g L_f^{r_j-1} h_j(x) \neq 0$.

Theorem 3, [8]. Consider a BIBS stable system (3.7) with uniformly bounded input u. For a given choice of the multi-index $\bar{s}$ assume that:

i. $\Phi_{\bar{s}}(x)$ is a diffeomorphism;
ii. the functions $L_f^{s_j} h_j\big(\Phi_{\bar{s}}^{-1}(z)\big)$ are uniformly Lipschitz in $\Phi_{\bar{s}}(\Omega)$ for any $j = 1, \ldots, q$;
iii. the observations relative degrees are such that $s_j \leq r_j$, $j = 1, \ldots, q$;
iv. when $s_j = r_j$, the functions $L_g L_f^{s_j-1} h_j\big(\Phi_{\bar{s}}^{-1}(z)\big)$ are uniformly Lipschitz in $\Phi_{\bar{s}}(\Omega)$;

Then there exists a gain matrix $K \in \mathbb{R}^{n\times q}$ such that the solution of the system equations

$$\dot{\hat{x}} = f(\hat{x}) + g(\hat{x})u + Q_{\bar{s}}^{-1}\big(\hat{x}\big) K\big(y - h(\hat{x})\big), \qquad \hat{x}(0) = \bar{x}, \tag{3.11}$$

exponentially converges to $x(t)$, whatever is the initial observed state $\bar{x} \in \Omega^b$.

The structure of K can be chosen block diagonal, where the j-th block, $j = 1, \ldots, q$ is a gain vector $s_j \times 1$, see [8] for more details.

4. OBSERVERS FOR TRANSCRIPTION NETWORKS. In this section the problem concerning *a priori* observability will be investigated, by checking the invertibility of the observability map. Two main experimental frameworks will be considered.

4.1. Step-wise forcing input. Consider the ODE model of a generic transcription network described by eqs.(2.1). According to the following more compact notation

$$\chi(t) = \begin{bmatrix} x_1 \\ \vdots \\ x_N \\ X_1 \\ \vdots \\ X_N \end{bmatrix} \in \mathbb{R}^{2N}, \qquad f(\chi) = \begin{bmatrix} -\lambda_{x_1}\chi_1 + \varphi_1(\chi_{N+1}, \ldots, \chi_{2N}) \\ \vdots \\ -\lambda_{x_N}\chi_N + \varphi_N(\chi_{N+1}, \ldots, \chi_{2N}) \\ p_{X_1}\chi_1 - \lambda_{X_1}\chi_{N+1} \\ \vdots \\ p_{X_N}\chi_N - \lambda_{X_N}\chi_{2N} \end{bmatrix} \tag{4.1}$$

and assuming only one forcing input (i.e., without loss of generality: $u_i \equiv 0$ for $i > 1$), system (2.1) can be written as:

$$\dot{\chi}(t) = f\big(\chi(t)\big) + gu(t), \qquad g = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{R}^{2N}. \tag{4.2}$$

By assuming to measure the mRNA (i.e. state variables χ_1, ..., χ_N) we have the following measurement equations:

$$\xi_i(t) = h_i\big(\chi(t)\big), \qquad h_i(\chi) = \chi_i, \qquad i = 1, \ldots, N. \tag{4.3}$$

As it easily comes from the system equations, if the perturbation input is set equal to zero, the system approaches the only equilibrium point (the steady-state, actually). One common experiment is given by a simple stepwise perturbation of the steady-state of the system, that is we assume to have $u(t) \equiv \bar{u} \neq 0$ for $t \geq 0$: from a mathematical point of view, this means to restate the dynamics as:

$$\dot{\chi}(t) = \tilde{f}\big(\chi(t)\big), \qquad \tilde{f}(\chi) = f(\chi) + g\bar{u}. \tag{4.4}$$

Then, according to the multi-output case, the following observability map can be computed:

$$\Phi(\chi) = \begin{bmatrix} h_1(\chi) \\ L_{\tilde{f}} h_1(\chi) \\ \vdots \\ h_N(\chi) \\ L_{\tilde{f}} h_N(\chi) \end{bmatrix} = \begin{bmatrix} \chi_1 \\ -\lambda_{x_1}\chi_1 + \varphi_1(\chi_{N+1}, \dots, \chi_{2N}) + \bar{u} \\ \vdots \\ \chi_N \\ -\lambda_{x_N}\chi_N + \varphi_N(\chi_{N+1}, \dots, \chi_{2N}) \end{bmatrix} \tag{4.5}$$

A first result is that the observability map (4.5) is invertible iff its Jacobian:

$$\frac{d\Phi}{d\chi} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ -\lambda_{x_1} & 0 & \cdots & 0 & \varphi_1^{(1)} & \cdots & \varphi_1^{(N)} \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -\lambda_{x_N} & \varphi_N^{(1)} & \cdots & \varphi_N^{(N)} \end{bmatrix} \qquad \varphi_i^{(j)} = \frac{\partial \varphi_i}{\partial \chi_{N+j}} \tag{4.6}$$

is invertible, that is iff matrix:

$$\begin{bmatrix} \varphi_1^{(1)} & \cdots & \varphi_1^{(N)} \\ \vdots & \ddots & \vdots \\ \varphi_N^{(1)} & \cdots & \varphi_N^{(N)} \end{bmatrix} \qquad \text{is invertible} \tag{4.7}$$

Notice that, according to (4.7), a necessary condition is that each node needs to work both as a regulated gene and as a transcription factor, otherwise matrix (4.7) has a null row/column: in this case not all the protein expressed by the network can be observed. For instance, if we consider the case of a MO-FFL (eqs.(2.5)), proteins Z_i cannot be observed, since they do not regulate any gene of the network. Then, if we apply the observer theory to a MO-FFL, a reasonable problem to investigate consists of estimating the time course of the master proteins X and Y by using measurements from mRNAs, that is

$$\xi_i(t) = h_i\big(\chi(t)\big), \qquad h_i(\chi) = \chi_i, \qquad i = 1, \dots, m+2 \tag{4.8}$$

with $\chi = \begin{bmatrix} x & y & z_1 & \cdots & z_m & X & Y \end{bmatrix}^T \in \mathbb{R}^{4+m}$. For instance, let us suppose to have only one target gene z (i.e. $m = 1$). Then, we can write eqs.(2.5)), in the same compact form of (4.4) with:

$$f(\chi) = \begin{bmatrix} -\lambda_x\chi_1 + \varphi_x(\chi_4) \\ -\lambda_y\chi_2 + \varphi_y(\chi_4) \\ -\lambda_z\chi_3 + \varphi_z(\chi_4, \chi_5) \\ -\lambda_X\chi_4 + p_X\chi_1 \\ -\lambda_Y\chi_5 + p_Y\chi_2 \end{bmatrix}. \tag{4.9}$$

If we choose the observability map as:

$$\Phi(\chi) = \begin{bmatrix} h_1(\chi) \\ L_{\tilde{f}} h_1(\chi) \\ h_2(\chi) \\ h_3(\chi) \\ L_{\tilde{f}} h_3(\chi) \end{bmatrix} = \begin{bmatrix} \chi_1 \\ -\lambda_x\chi_1 + \varphi_x(\chi_4) + \bar{u} \\ \chi_2 \\ \chi_3 \\ -\lambda_z\chi_3 + \varphi_z(\chi_4, \chi_5) \end{bmatrix} \tag{4.10}$$

then $\Phi(\chi)$ is a diffeomorphism if its Jacobian:

$$\frac{d\Phi}{d\chi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\lambda_x & 0 & 0 & \varphi_x^{(X)} & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\lambda_z & \varphi_z^{(X)} & \varphi_z^{(Y)} \end{bmatrix}, \qquad \begin{aligned} \varphi_x^{(X)} &= \frac{d\varphi_x}{dX}, \\ \varphi_z^{(X)} &= \frac{\partial \varphi_z}{\partial X}, \qquad \varphi_z^{(Y)} = \frac{\partial \varphi_z}{\partial Y} \end{aligned} \tag{4.11}$$

is invertible, which clearly is, since:

$$\det\left(\frac{d\Phi}{d\chi}\right) = \varphi_x^{(X)}(\chi_4)\varphi_z^{(Y)}(\chi_4, \chi_5) \neq 0, \qquad \forall \chi_4, \chi_5 > 0. \tag{4.12}$$

Then, consider to have measurements only from the target genes z_i:

$$\xi_i(t) = h_i\big(\chi(t)\big), \qquad h_i(\chi) = \chi_{i+2}, \quad i = 1, \dots, m \tag{4.13}$$

and to aim to observe not only the protein contents X and Y but also the corresponding mRNA x and y. Then, by taking into account the case of a single target gene, we have the unique choice for the observability map:

$$\Phi(\chi) = \begin{bmatrix} h(\chi) \\ L_{\tilde{f}} h(\chi) \\ \vdots \\ L_{\tilde{f}}^4 h(\chi) \end{bmatrix} \tag{4.14}$$

where, starting from the third component, we have dependence from all the five state variable, which is a necessary condition to find $\Phi(\chi)$ a diffeomorphism (cumbersome computations are not reported). By adding more target genes (i.e. more measurements) computations simplify: for instance, according to two target genes, the observability map may be written as:

$$\Phi(\chi) = \begin{bmatrix} h_1(\chi) \\ L_{\tilde{f}} h_1(\chi) \\ L_{\tilde{f}}^2 h_1(\chi) \\ h_2(\chi) \\ L_{\tilde{f}} h_2(\chi) \\ L_{\tilde{f}}^2 h_2(\chi) \end{bmatrix} \tag{4.15}$$

involving only second order Lie derivative (instead of 4th), with:

$$L_{\tilde{f}} h_i(\chi) = -\lambda_{z_i}\chi_{i+2} + \varphi_{z_i}(\chi_5, \chi_6) \tag{4.16}$$

$$\begin{aligned} L_{\tilde{f}}^2 h_i(\chi) = {} & \varphi_{z_i}^{(X)}(\chi_5, \chi_6)\big(-\lambda_X \chi_5 + p_X \chi_1\big) + \varphi_{z_i}^{(Y)}(\chi_5, \chi_6)\big(-\lambda_Y \chi_6 + p_Y \chi_2\big) \\ & -\lambda_{z_i}\big(-\lambda_{z_i}\chi_{i+2} + \varphi_{z_i}(\chi_5, \chi_6)\big) \end{aligned} \tag{4.17}$$

Thus, it comes, from further computations:

$$\begin{aligned} \det\left(\frac{d\Phi}{d\chi}\right) = p_X p_Y \Big(& \Big(\varphi_{z_1}^{(X)}(\chi_5, \chi_6)\Big)^2 \Big(\varphi_{z_2}^{(Y)}(\chi_5, \chi_6)\Big)^2 \\ & + \Big(\varphi_{z_1}^{(Y)}(\chi_5, \chi_6)\Big)^2 \Big(\varphi_{z_2}^{(X)}(\chi_5, \chi_6)\Big)^2 \Big) \end{aligned} \tag{4.18}$$

which is strictly greater than zero for any $\chi_5, \chi_6 > 0$.

Finally, we investigate the observability problem concerning the simplified model of a MO-FFL case (2.8). By setting the state vector $\chi = [X \;\; Y \;\; Z_1 \;\; \cdots \;\; Z_m]^T \in \mathbb{R}^{2+m}$ and the measurement equations:

$$\xi_i(t) = h_i\big(\chi(t)\big) = \chi_{i+2}(t), \qquad i = 1, \ldots, m, \tag{4.19}$$

the ODE model can be rewritten in the compact notation of (4.4) where:

$$f(\chi) = \begin{bmatrix} -\lambda_X \chi_1 + \frac{p_X}{\lambda_x}\varphi_x(\chi_1) \\ -\lambda_Y \chi_2 + \frac{p_Y}{\lambda_y}\varphi_y(\chi_1) \\ -\lambda_{Z_1} \chi_3 + \frac{p_{Z_1}}{\lambda_{z_1}}\varphi_{z_1}(\chi_1, \chi_2) \\ \vdots \\ -\lambda_{Z_m} \chi_{m+2} + \frac{p_{Z_m}}{\lambda_{z_m}}\varphi_{z_m}(\chi_1, \chi_2) \end{bmatrix}, \qquad g = \begin{bmatrix} \frac{p_X}{\lambda_x} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4.20}$$

By first considering the case of only one measurement, we have:

$$\Phi(\chi) = \begin{bmatrix} h(\chi) \\ L_{\tilde{f}} h(\chi) \\ L^2_{\tilde{f}} h(\chi) \end{bmatrix} \tag{4.21}$$

with:

$$\begin{aligned} &L_{\tilde{f}} h(\chi) = -\lambda_Z \chi_3 + \tfrac{p_Z}{\lambda_z}\varphi_z(\chi_1, \chi_2), \\ &L^2_{\tilde{f}} h(\chi) = \tfrac{p_Z}{\lambda_z}\varphi_z^{(X)}(\chi_1)\tilde{f}_1(\chi_1) + \tfrac{p_Z}{\lambda_z}\varphi_z^{(Y)}(\chi_2)\tilde{f}_2(\chi_1, \chi_2) - \lambda_Z \tilde{f}_3(\chi_1, \chi_2, \chi_3) = \psi(\chi), \end{aligned} \tag{4.22}$$

from which the observability matrix is computed:

$$\frac{d\Phi}{d\chi} = \begin{bmatrix} 0 & 0 & 1 \\ \frac{p_Z}{\lambda_z}\varphi_z^{(X)} & \frac{p_Z}{\lambda_z}\varphi_z^{(Y)} & -\lambda_Z \\ \psi^{(X)} & \psi^{(Y)} & \lambda_Z^2 \end{bmatrix}, \qquad \text{with} \qquad \psi^{(X)} = \frac{\partial \psi}{\partial X}, \quad \psi^{(Y)} = \frac{\partial \psi}{\partial Y} \tag{4.23}$$

The diffeomorphism of $\Phi(\chi)$ is related to the invertibility of the Jacobian, which is provided iff the following inequality holds true:

$$\det\left(\frac{d\Phi}{d\chi}\right) = \frac{p_Z}{\lambda_z}\Big(\varphi_z^{(X)}(\chi_1, \chi_2)\psi^{(Y)}(\chi_1, \chi_2, \chi_3) - \varphi_z^{(Y)}(\chi_1, \chi_2)\psi^{(X)}(\chi_1, \chi_2, \chi_3)\Big) \neq 0. \tag{4.24}$$

Again, in case of more target genes, we have a richer set of measurements to use. For instance, assume to have three target genes. Then, the observability map may be defined as:

$$\Phi(\chi) = \begin{bmatrix} h_1(\chi) \\ L_{\tilde{f}} h_1(\chi) \\ h_2(\chi) \\ L_{\tilde{f}} h_2(\chi) \\ h_3(\chi) \end{bmatrix} \tag{4.25}$$

from which:

$$\frac{d\Phi}{d\chi} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ \varphi_{z_1}^{(X)} & \varphi_{z_1}^{(Y)} & -\lambda_{Z_1} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \varphi_{z_2}^{(X)} & \varphi_{z_2}^{(Y)} & 0 & -\lambda_{Z_2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.26}$$

Therefore, the observability condition becomes:

$$\varphi_{z_1}^{(X)}(\chi_1,\chi_2)\varphi_{z_2}^{(Y)}(\chi_1,\chi_2) - \varphi_{z_2}^{(X)}(\chi_1,\chi_2)\varphi_{z_1}^{(Y)}(\chi_1,\chi_2) \neq 0, \tag{4.27}$$

which is easier to compute and to deal with, with respect to (4.24). In fact, it clearly comes out that an observability map like (4.25) is a diffeomorphism only if φ_{z_1} is different than φ_{z_2}, which is easy to grasp, since if φ_{z_1} was equal to φ_{z_2}, the transcription factors expressed by master genes X and Y would exert the same influence to Z_1 and Z_2.

4.2. Generic forcing input. A more interesting case is when the perturbation input is actually time-varying, for instance a piecewise constant input with many possible time instants of switch, or a periodic harmonic signal. In these cases, it is important to look for the relative degree associated to a given output function $h_i(\chi)$, which is defined by (3.4). If we consider the general model (4.1-4.3), we have relative degree equal to 1 for $h_1(\chi) = \chi_1$ since $L_g h_1(\chi) = 1 \neq 0$, while different output functions may have different relative degrees according to the way the input u affects the output function derivatives. As a matter of fact, the observability map needs to be designed according to the relative degrees of the output functions. For instance, in case of a single target FFL, model equations (4.2) with function f as in (4.9) and measurements given by (4.8) ($m = 1$), the output functions $h_2(\chi) = \chi_2$ and $h_3(\chi) = \chi_3$ have both relative degree 3, since:

$$L_g h_2(\chi) = 0 \tag{4.28}$$

$$L_f h_2(\chi) = -\lambda_y \chi_2 + \varphi_y(\chi_4) \qquad \Longrightarrow \qquad L_g L_f h_2(\chi) = 0 \tag{4.29}$$

$$L_f^2 h_2(\chi) = -\lambda_y\big(-\lambda_y\chi_2 + \varphi_y(\chi_4)\big) + \varphi_y^{(X)}(\chi_4)\big(-\lambda_X\chi_4 + p_X\chi_1\big)$$

$$\Longrightarrow \qquad L_g L_f^2 h_2(\chi) = p_X \varphi_y^{(X)}(\chi_4) \neq 0 \tag{4.30}$$

and

$$L_g h_3(\chi) = 0 \tag{4.31}$$

$$L_f h_3(\chi) = -\lambda_z \chi_3 + \varphi_z(\chi_4,\chi_5) \qquad \Longrightarrow \qquad L_g L_f h_3(\chi) = 0 \tag{4.32}$$

$$L_f^2 h_3(\chi) = -\lambda_z\big(-\lambda_z\chi_3 + \varphi_z(\chi_4,\chi_5)\big) + \varphi_z^{(X)}(\chi_4,\chi_5)\big(-\lambda_X\chi_4 + p_X\chi_1\big)$$

$$+\varphi_z^{(Y)}(\chi_4,\chi_5)\big(-\lambda_Y\chi_5 + p_Y\chi_2\big) \quad \Longrightarrow \quad L_g L_f^2 h_3(\chi) = p_X \varphi_z^{(X)}(\chi_4,\chi_5) \neq 0 \tag{4.33}$$

Then, a reasonable choice for the observability map is (recall that Z is not observable):

$$\Phi(\chi) = \begin{bmatrix} h_1(\chi) \\ h_2(\chi) \\ L_f h_2(\chi) \\ h_3(\chi) \\ L_f h_3(\chi) \end{bmatrix} = \begin{bmatrix} \chi_1 \\ \chi_2 \\ -\lambda_y\chi_2 + \varphi_y(\chi_4) \\ \chi_3 \\ -\lambda_z\chi_3 + \varphi_z(\chi_4,\chi_5) \end{bmatrix} \tag{4.34}$$

which is a diffeomorphism since (computations similar to the ones which provided eq.(4.12):

$$\det\left(\frac{d\Phi}{d\chi}\right) = \varphi_y^{(X)}(\chi_4)\varphi_z^{(Y)}(\chi_4,\chi_5) \neq 0, \qquad \forall \chi_4,\chi_5 > 0. \tag{4.35}$$

On the other hand, in case of measurements coming only from the target genes z_i (i.e. as in (4.13)), we need at least a couple of outputs to ensure a full relative degree and an invertible observability map

$$\Phi(\chi) = \begin{bmatrix} h_1(\chi) \\ L_f h_1(\chi) \\ L_f^2 h_1(\chi) \\ h_2(\chi) \\ L_f h_2(\chi) \\ L_f^2 h_2(\chi) \end{bmatrix} \tag{4.36}$$

since we have seen that $h_i(\chi) = \chi_{i+2}$ has relative degree equal to 3.

It has to be stressed that, in case of output functions which do not ensure a full relative degree, a way to cope with this problem is to include the input $u(t)$, and (possibly) its time derivatives, in the observability map (time-varying, actually). Such a case is specially useful when the relative degree is close to be full. This case occurs when investigating the observability of a simplified model of the single output FFL (model equations as in (4.2), with functions f, g given by (4.20) and measurement equation (4.19), $m = 1$). In this case we have a relative degree equal to 2, since:

$$\begin{aligned} &L_g h(\chi) = 0 \\ &L_f h(\chi) = -\lambda_Z \chi_3 + \frac{p_Z}{\lambda_z}\varphi_z(\chi_1,\chi_2) \quad \Longrightarrow \quad L_g L_f h(\chi) = \frac{p_Z}{\lambda_z}\varphi_z^{(X)}(\chi_1,\chi_2) \neq 0, \end{aligned} \tag{4.37}$$

Then, we can design the observability map as:

$$\Phi(\chi,t) = \begin{bmatrix} h(\chi) \\ L_{f+gu} h(\chi) \\ L_{f+gu}^2 h(\chi) \end{bmatrix} = \begin{bmatrix} h(\chi) \\ L_f h(\chi) \\ L_f^2 h(\chi) + L_g L_f h(\chi) u(t) \end{bmatrix} \tag{4.38}$$

Besides the usual conditions on the diffeomorphism of Φ and the fact that $L_{f+gu}^2\big(\Phi^{-1}(\cdot)\big)$ and $L_{f+gu}^3\big(\Phi^{-1}(\cdot)\big)$ are uniformly Lipschitz, a further sufficient condition is required to ensure the convergence of the observer error, that is the time derivative of the input to be uniformly bounded. Note that this last condition seems to exclude the possibility to apply the observer to piecewise constant switching perturbations; nevertheless what really happens in practical cases (as it will be shown also in the next Section, dedicated to simulations) is that, by applying the observer, each time a switch occurs, the observed state may be drifted far from the real state (to which it was clamped), towards which it converges again in a reasonable time depending on the convergence rate.

However, the case of one target gene is the only one in which the hypothesis of full relative degree does not hold for the above mentioned model: it is enough to add just one more target gene to overcome the lack of relative degree of a single output function. Consequently, with the conditions $m > 1$ and bounded input $u(t)$ it is possible to build an observer which ensures convergence to zero of the estimation error.

TABLE 5.1
Parameter values of a FFL with one target gene, eqs.(2.5-2.7)

$\lambda_x = 0.2$	$V_x^0 = 1.1$	$K_{xx} = 0.9$	$\nu_{xx} = 2.0$	
$\lambda_y = 0.4$	$V_y^0 = 0.9$	$\bar{V}_{yx} = 1.5$	$K_{yx} = 1.0$	$\nu_{yx} = 2.0$
$\lambda_X = 0.05$	$p_X = 1.2$	$\lambda_Y = 0.08$	$p_Y = 1.5$	
$\lambda_z = 0.1$	$V_z^0 = 1.03$	$\bar{V}_{zx} = 1.1$	$K_{zx} = 1.1$	$\nu_{zx} = 3.0$
$\bar{V}_{zy} = 1.2$	$K_{zy} = 1.2$	$\nu_{zy} = 4.0$		

5. SIMULATION RESULTS. This section proposes some simulations showing the good performances of the observer-based approach to infer information from available transcript measurements, as well as its robustness with respect to measurement and parameter uncertainties. All simulations refer to the FFL network motif.

5.1. Single-Output FFL: protein estimation from noisy mRNA measures. First results concern the case of a FFL with only one target gene, $m = 1$, modeled as in (2.5-2.7), with parameters given by Table 5.1, and a constant forcing input $u(t) \equiv 1$. Measurements are given by mRNAs (i.e. state variables x, y, z), eq.(4.8). In order to stress the efficacy of the observer-based methodology, pseudo-random noisy measurements have been considered by suitably modifying the output equations (4.8) as:

$$\xi_i(t) = \chi_i(t) + \mathcal{N}_i(t), \tag{5.1}$$

where the disturbance $\mathcal{N}_i(t)$ is modeled by the following sum of harmonics:

$$\mathcal{N}_i(t) = \sum_{j=1}^{k} H_{ij} \sin\left(2\pi f_{ij} t + \alpha_{ij}\right). \tag{5.2}$$

Simulations are reported on a time range of 50 time units, concerning disturbances (5.2) with $k = 7$ high-frequency harmonics (high-frequency with respect to the half-life of the system dynamics); indeed, according to half-lives spanning from about 2 to 20 time units, we have chosen harmonics with periods spanning from 1/50 to 1/5000 time units:

$$f_{ij} \in \{50, \quad 100 \quad 200, \quad 500, \quad 1000, \quad 2000, \quad 5000\} \tag{5.3}$$

Parameters H_{ij} and α_{ij} are assigned to guarantee a level of uncertainties compatible with the one reported in the literature. For instance, in [18], a high level of accuracy and reproducibility ($< 2.5\%$ variation) is reported for real-time PCR of a target gene transcript. Therefore the pseudo-random noise parameters H_{ij}, α_{ij} have been set to have uncertainties providing a signal-to-noise ratio never definitely below 5% (see Fig.5.1).

The chosen observability map is that of eq.(4.10). The real initial conditions for the protein content are $X(0) = 50$, $Y(0) = 10$ and estimated initial protein content are set to 20% of their real value. As it can be seen by Figs.5.2-5.3, the observer allows to estimate both X and Y protein content in a quite accurate way, despite to the non-negligible noises affecting the output measurements.

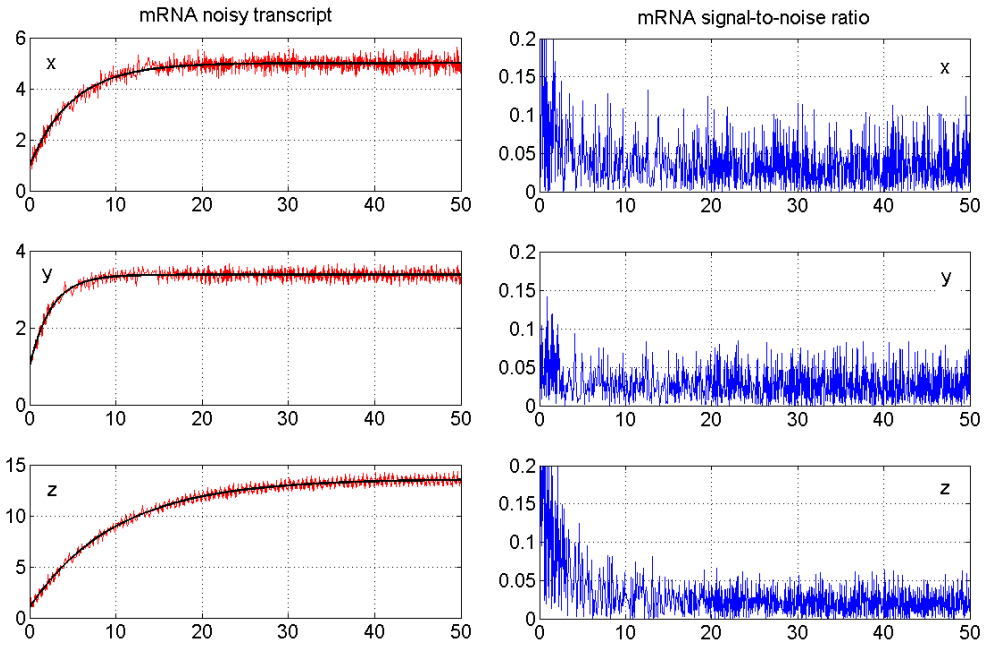

FIG. 5.1. *mRNA noisy measurements and the corresponding signal-to-noise ratio.*

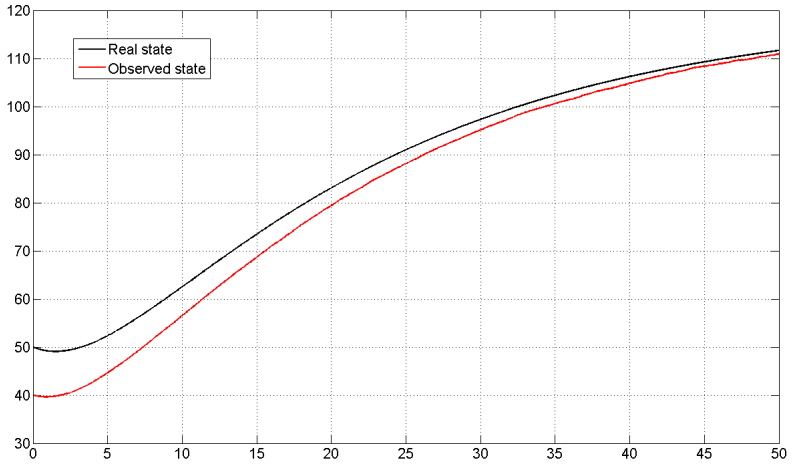

FIG. 5.2. *X protein content, single target FFL case.*

5.2. Single-Output FFL: simplified model with time-varying input. A second set of simulations has been carried out according to the simplified model of a single-output FFL (2.8), whose model parameters are reported in Table 5.2. In this case, we consider also an exogenous piecewise constant input, which is a square wave with lower level equal to zero, upper level equal to 2 and a duty cycle of 50% on a 10 time units period. The state vector is $\chi = [X \quad Y \quad Z]^T$ and measurements come from Z, according to (4.19), $m = 1$. As previously shown, this is the case of a relative degree lower than the system dimension; nevertheless, by choosing a time-varying

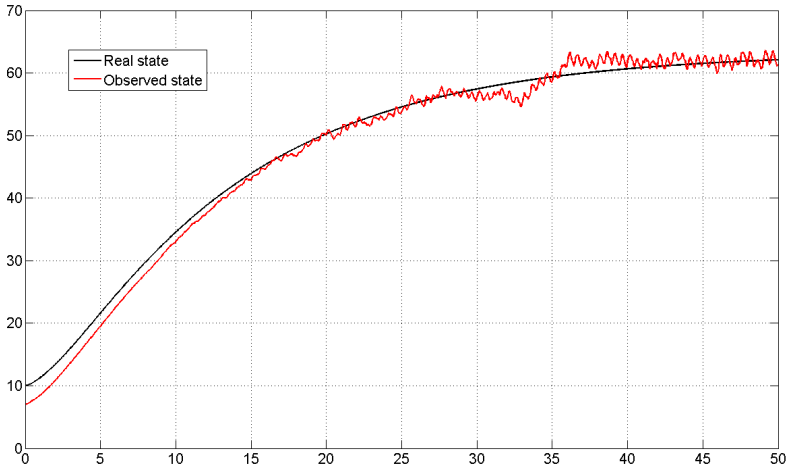

FIG. 5.3. *Y protein content, single target FFL case.*

TABLE 5.2
Parameter values of a FFL with one target gene, eqs.(2.8)

$K_{xx} = 0.9$	$\nu_{xx} = 2.0$	$\lambda_X = 0.2$	$\lambda_Y = 0.4$	$p_X/\lambda_x = 1$
$\bar{V}_{yx} = 1.5$	$K_{yx} = 1.0$	$\nu_{yx} = 2.0$	$\lambda_Z = 0.1$	
$\bar{V}_{zx} = 1.1$	$K_{zx} = 1.1$	$\nu_{zx} = 3.0$	$p_X V_x^0/\lambda_x = 1.1$	
$\bar{V}_{zy} = 1.2$	$K_{zy} = 1.2$	$\nu_{zy} = 4.0$	$p_Y V_y^0/\lambda_y = 0.9$	$p_Z V_z^0/\lambda_z = 1.03$

observability map as in (4.38), we still obtain very good results, as it comes out from Fig.s 5.4-5.5 (initial conditions: $X(0) = 1$, $Y(0) = 1$ and $\hat{X}(0) = 0.5$, $\hat{Y}(0) = 2.8$).

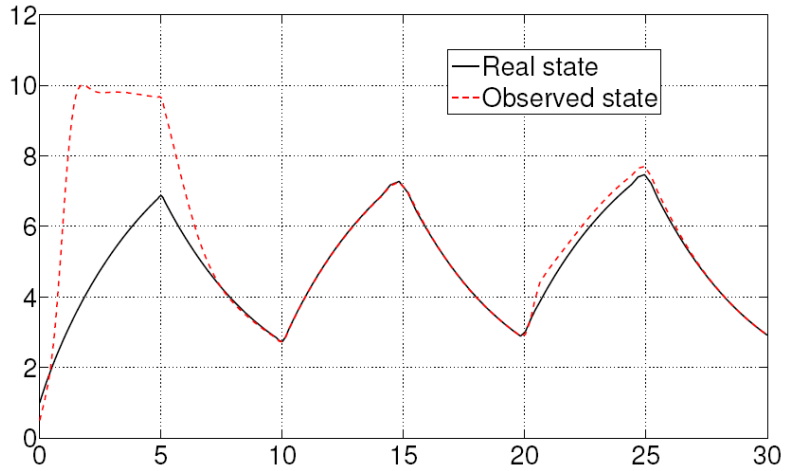

FIG. 5.4. *Gene x transcript evolution: square wave input, the SISO case.*

5.3. Multi-Output FFL: simplified model with uncertainties. The good performances of the proposed observer are theoretically constrained to the knowledge

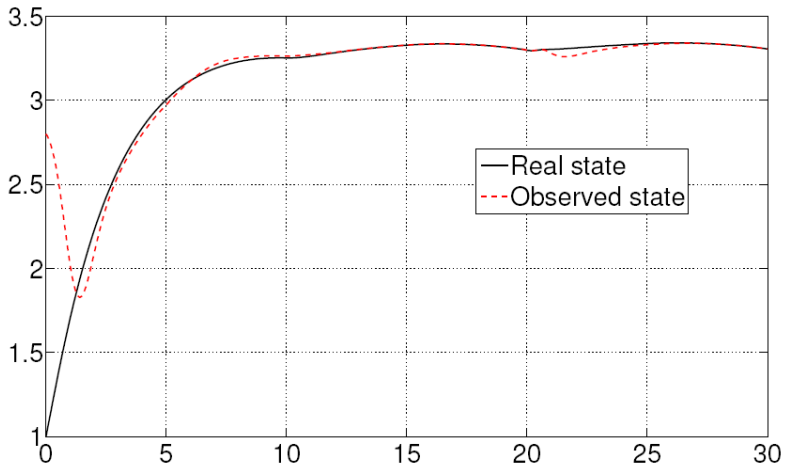

FIG. 5.5. *Gene y transcript evolution: square wave input, the SISO case.*

TABLE 5.3
Parameter values of a FFL with 3 target genes, eqs.(2.8), as originally stated

$\lambda_X = 0.05$	$p_X V_x^0/\lambda_x = 0.1056$	$K_{xx} = 0.9$	$\nu_{xx} = 1.0$	$p_X/\lambda_x = 1$
$\lambda_Y = 0.05$	$p_Y V_y^0/\lambda_y = 0.0327$	$\bar{V}_{yx} = 5.7793$	$K_{yx} = 2.0$	$\nu_{yx} = 3.0$
$\lambda_{Z_1} = 0.08$	$p_{Z_1} V_{z_1}^0/\lambda_{z_1} = 0.0506$	$\bar{V}_{z_1x} = 3.6886$	$K_{z_1x} = 2.1$	$\nu_{z_1x} = 3.0$
$\lambda_{Z_2} = 0.02$	$p_{Z_2} V_{z_2}^0/\lambda_{z_2} = 0.0146$	$\bar{V}_{z_2x} = 2.7743$	$K_{z_2x} = 4.5$	$\nu_{z_1x} = 3.5$
$\lambda_{Z_3} = 2$	$p_{Z_3} V_{z_3}^0/\lambda_{z_3} = 1.9925$	$\bar{V}_{z_3x} = 3.0924$	$K_{z_3x} = 9.99$	$\nu_{z_3x} = 5.1$
$\bar{V}_{z_1y} = 4.8$	$K_{z_1y} = 1.8750$	$K_{z_2y} = 2.1250$	$\bar{V}_{z_2y} = 5.1$	$\nu_{z_1y} = 4.2$
$\nu_{z_2y} = 3.1$	$\bar{V}_{z_3y} = 6.3$	$K_{z_3y} = 3.7375$	$\nu_{z_3y} = 5.5$	

of both the model equations and the model parameters. The following set of simulations is therefore devoted to stress the methodology by adding some uncertainties above the parameter estimates. We refer to a recent paper [10], where an identification procedure has been considered to validate a mathematical model (among a small set of candidates) by means of some set of measurements. The approach provides the model parameter estimates as a by product. In that case a three-target-genes FFL was considered, whose model equations are the ones of the simplified model (2.8), and the model parameters were set equal to the ones reported in Table 5.3. The identification procedure provided the model parameters reported in Table 5.4, according to the measurements from the target genes Z_i.

Notice that most parameters are very well estimated, with an error percentage lower than 6% (higher percentage error for ν_{z_1x}, equal to 5.81%). Then, we pursue to estimate the time course of genes X and Y by way of Z_i measurements: to this aim we compute the observer parameters according to the estimated parameters of Table 5.4 (instead of the real ones of Table 5.3), thus introducing a light (but reasonable) uncertainty in the observer equations. Moreover, we considered the case of an exogenous (known) harmonic forcing input, equal to $u(t) = 0.1 + 0.1\sin(2\pi t/40)$. In this case we have full relative degree, since we have more than one target gene. Initial conditions are $X(0) = 1$, $Y(0) = 1$ and $\hat{X}(0) = 0.8$, $\hat{Y}(0) = 1.2$. Despite the uncertainties, we still obtain very good results, as it comes out from Fig.s 5.6-5.7.

TABLE 5.4
Estimated parameter values of a FFL with 3 target genes, eqs.(2.8). Real values in Table 3.

$\lambda_X = 0.0501$	$p_X V_x^0/\lambda_x = 0.1050$	$K_{xx} = 0.9117$	$\nu_{xx} = 0.9968$	$p_X/\lambda_x = 1$
$\lambda_Y = 0.0496$	$p_Y V_y^0/\lambda_y = 0.0332$	$\bar{V}_{yx} = 5.6375$	$K_{yx} = 2.0171$	$\nu_{yx} = 3.0318$
$\lambda_{Z_1} = 0.0794$	$p_{Z_1} V_{z_1}^0/\lambda_{z_1} = 0.0489$	$\bar{V}_{z_1x} = 3.6724$	$K_{z_1x} = 2.0358$	$\nu_{z_1x} = 2.8257$
$\lambda_{Z_2} = 0.0198$	$p_{Z_2} V_{z_2}^0/\lambda_{z_2} = 0.0145$	$\bar{V}_{z_2x} = 2.7819$	$K_{z_2x} = 4.4856$	$\nu_{z_1x} = 3.5473$
$\lambda_{Z_3} = 2.0198$	$p_{Z_3} V_{z_3}^0/\lambda_{z_3} = 2.0133$	$\bar{V}_{z_3x} = 3.0748$	$K_{z_3x} = 10.0217$	$\nu_{z_3x} = 5.2261$
$\bar{V}_{z_1y} = 4.7601$	$K_{z_1y} = 1.8939$	$\nu_{z_1y} = 4.2507$	$\bar{V}_{z_2y} = 5.0889$	$K_{z_2y} = 2.1568$
$\nu_{z_2y} = 3.0738$	$\bar{V}_{z_3y} = 6.3128$	$K_{z_3y} = 3.7325$	$\nu_{z_3y} = 5.6263$	

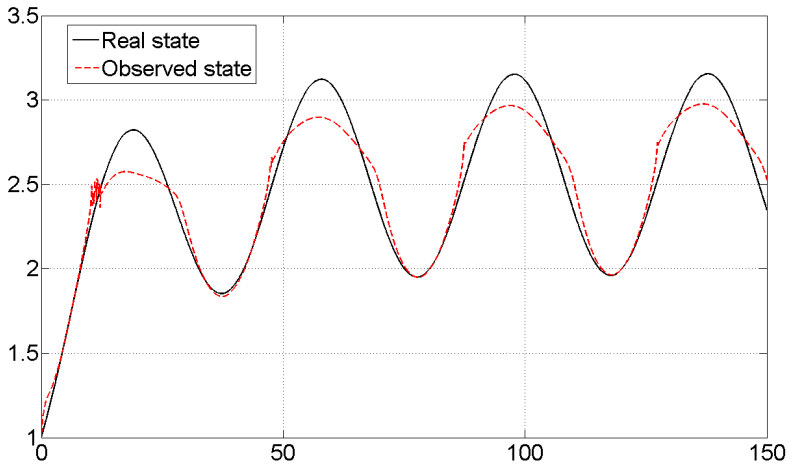

FIG. 5.6. *Gene x transcript evolution: harmonic input, the MIMO case with uncertainties.*

In order to further stress the robustness of the proposed approach, we have increased the uncertainties of a reduced set of parameters (the ones involving gene y, for instance: ν_{yx}, K_{yx}, ν_{z_1y}, ν_{z_2y}, ν_{z_2y}, K_{z_1y}, K_{z_2y}, K_{z_3y}) up to a coefficient of variation of 20%, with respect to the original values of Table 5.3. Then, a set of 1,000 *in silico* experiments has been done, each experiment related to the same MO-FFL whose parameters are the ones in Table 5.3. It clearly appears from simulations that the x gene expression is still perfectly observed (they are not reported but definitely close to the one of Fig.5.6); on the other hand the y gene expression can result more difficult to observe (indeed all the uncertain parameters are directly related to gene y). To evaluate the goodness of the gene y estimate, we have stated that a *good* estimate constrains the asymptotic error between real and estimated Y within the 20% of the real value, and a *very good* estimate constrains the asymptotic error between real and estimated Y within the 5% of the real value. Results show that according to the above mentioned uncertainties we have 98,3% of *good* estimates, and 55,7% of *very good* estimates.

6. CONCLUSIONS. The systems biology approach to the study of the gene regulatory system in living cells has suffered so far by the lack of time series experimental data that allow to reconstruct the structure of transcription networks. In the first place, genome-wide techniques like micro-arrays only allow the measurement of mRNA, or, in other words, they measure only transcription activity of pro-

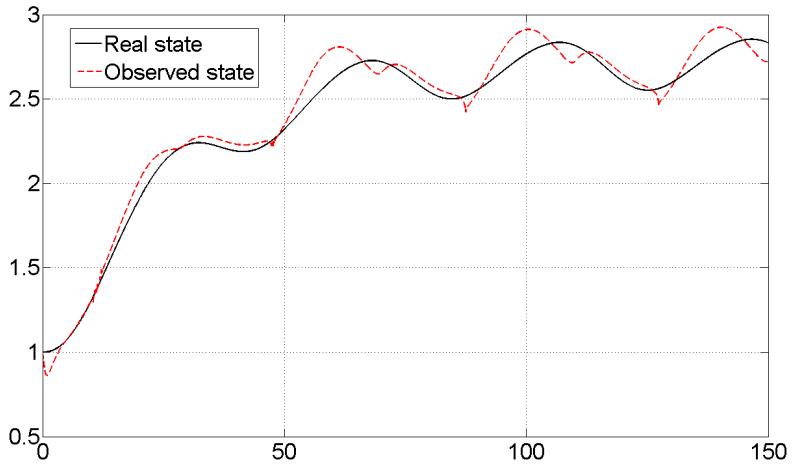

FIG. 5.7. *Gene y transcript evolution: harmonic input, the MIMO case with uncertainties.*

moters. Thus the dynamics of proteins remains largely unknown. This problem is specially important for the network analysis, since proteins are potentially regulated in every step of their synthesis process by many regulation mechanisms, including post-transcriptional regulations, such as the rate of mRNA degradation, the rate of translation, and the rate of protein degradation. For all these reasons, methods that allow the reconstruction of the protein levels are of the utmost importance in the biological analysis of cell functions.

In this paper a mathematical tool is presented, to infer information on the time-course of protein concentration of a transcription network, according to a set of measurements of the transcriptional activity of promoters. The approach is based on the use of the mathematical model of the network under investigation and on the state-reconstruction technique known as 'state-observer', borrowed from the control theory. To this aim, besides the general case, the network motif of the Multi-Output Feed-Forward Loop (MO-FFL) is investigated in details. Simulations show the effectiveness of the proposed approach in a very wide range of possible critical frameworks, such as only one target gene measurements, non-smooth input perturbations or parameter uncertainties.

7. ACKNOWLEDGEMENTS. The Authors wish to thank Prof. Lorenzo Farina for his stimulating, critical and helpful discussion on gene network topics.

REFERENCES

[Ahrens & Khalil, 2009] J.H. Ahrens and H.K. Khalil, High gain observers in the presence of measurement noise: a switched-gain approach, *Automatica*, 45 (4), 936–943, 2009.

[1] U. Alon, *An introduction to systems biology*, Chapman and Hall/CRC, (2006).

[2] U. Alon, Network motifs: theory and experimental approaches, *Nature Reviews Genetics*, **8** (2007), 450–461.

[3] D. Angeli and E.D. Sontag, Oscillations in I/O monotone systems, *IEEE Transactions on Circuits and Systems, Special Issue on Systems Biology*, **55** (2008), 166–176.

[4] F. Cacace, V. Cusimano, L. Di Paola and A. Germani, Observer-based techniques for the identification and analysis of avascular tumor growth, to be published in *Mathematical Biosciences*.

[5] F. Cacace, V. Cusimano, and A. Germani, An efficient approach to the design of observers for

continuous-time systems with discrete-time measurements, Proceedings of the 50th IEEE Conference on Decision and Control (2011), Orlando, Florida.

[6] F. Cacace, A. Germani and P. Palumbo, A state observer approach to filter stochastic nonlinear differential systems, Proceedings of the 50th IEEE Conference on Decision and Control (2011), Orlando, Florida.

[7] G. Ciccarella, M. Dalla Mora, and A. Germani, A Luemberger-like observer for nonlinear systems, *International Journal of Control*, **57** (1993), 537–556.

[8] M. Dalla Mora, A. Germani and C. Manes, Design of state observers from a drift-observability property, *IEEE Trans. Automatic Control*, **45** (2000), 1536–1540.

[9] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review, *Journal of Computational Biology*, **9** (2002), 67–103.

[10] L. Farina, A. Germani, G. Mavelli and P. Palumbo, Identification of regulatory network motifs from gene expression data, *Journal of Mathematical Modelling and Algorithms*, **9** (2010), 233–245.

[11] T. S. Gardner and J. J., Faith, Reverse-engineering transcription control networks, *Physics of Life Reviews*, **2** (2005), 65–88.

[12] Y. Huang, D. Liu and H. Wu, Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system, *Biometrics*, **62** (2006), 413–423.

[13] H. Kitano, Computational systems biology, *Nature*, **420** (2002), 206–210.

[14] H. Kitano, Towards a theory of biological robustness, *Molecular Systems Biology*, **3** (2007).

[15] T.I. Lee, N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, R.A. Young, Transcriptional regulatory networks in Saccharomyces cerevisiae, *Science*, **298** (2002), 799–804.

[16] S. Mangan, A. Zaslaver and U. Alon, The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks, *Journal of Molecular Biology*, **334** (2003), 197–204.

[17] P. Mendes, W. Sha and K. Ye, Artificial gene networks for objective comparison of analysis algorithms, *Bioinformatics*, **19** (2003), 122–129.

[18] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR, *Nucleic Acids Research*, **29** (2001), 2002–2007.

[19] M. Ronen, R. Rosenberg, B.I. Shraiman and U. Alon, Assignining numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics, *PNAS*, **99** (2002), 10555–10560.

[20] N. Soranzo, G. Bianconi and C. Altafini, Comparing association network algorithms for reverse engineering of large scale gene regulatory networks: synthetic vs real data, *Bioinformatics*, **23** (2007), 1640–1647.

[21] J. Stelling, U. Sauer, Z. Szallasi, F.J. Doyle III and J. Doyle, Robustness of cellular functions, *Cell* **118** (2004), 675–685.

[22] T.M. Yi, Y. Huang, M.I. Simon and J. Doyle, Robust perfect adaptation in bacterial chemotaxis through integral feedback control, *PNAS*, **97** (2000), 4649–4653.

[23] Z. Wang, M. Gerstein, and M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nature Genetics*, **10** (2009), 57–64.

[24] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette and U. Alon, A comprehensive library of fluorescent transcriptional reporters for Escherichia coli, *Nature Methods*, **3** (2006), 623–628.