SYNTHETIC BIOLOGY

Accurate characterization of dynamic microbial gene expression and growth rate profiles

Gonzalo Vidal [ID][1,2], Carolus Vitalis [ID][1], Macarena Muñoz Silva[1], Carlos Castillo-Passi [ID][1,3,4], Guillermo Yáñez Feliú [ID][2,5], Fernán Federici [ID][1,6], and Timothy J. Rudge [ID][2,*]

[1]Institute for Biological and Medical Engineering, Schools of Engineering, Biology and Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Interdisciplinary Computing and Complex BioSystems (ICOS) Research Group, School of Computing, Newcastle University, Newcastle Upon Tyne, UK
[3]School of Biomedical Engineering and Imaging Sciences, King's College London, St Thomas' Hospital, London, UK
[4]Millennium Institute for Intelligent Healthcare Engineering (iHEALTH), Santiago, Chile
[5]Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile
[6]ANID – Millennium Science Initiative Program, Millennium Institute for Integrative Biology (iBio) & FONDAP Center for Genome Regulation, Santiago, Chile
*Corresponding authors: E-mail: tim.rudge@ncl.ac.uk

Abstract

Genetic circuits are subject to variability due to cellular and compositional contexts. Cells face changing internal states and environments, the cellular context, to which they sense and respond by changing their gene expression and growth rates. Furthermore, each gene in a genetic circuit operates in a compositional context of genes which may interact with each other and the host cell in complex ways. The context of genetic circuits can, therefore, change gene expression and growth rates, and measuring their dynamics is essential to understanding natural and synthetic regulatory networks that give rise to functional phenotypes. However, reconstruction of microbial gene expression and growth rate profiles from typical noisy measurements of cell populations is difficult due to the effects of noise at low cell densities among other factors. We present here a method for the estimation of dynamic microbial gene expression rates and growth rates from noisy measurement data. Compared to the current state-of-the-art, our method significantly reduced the mean squared error of reconstructions from simulated data of growth and gene expression rates, improving the estimation of timing and magnitude of relevant shapes of profiles. We applied our method to characterize a triple-reporter plasmid library combining multiple transcription units in different compositional and cellular contexts in *Escherichia coli*. Our analysis reveals cellular and compositional context effects on microbial growth and gene expression rate dynamics and suggests a method for the dynamic ratiometric characterization of constitutive promoters relative to an *in vivo* reference.

Key words: Inverse problem; characterization; dynamical systems; web application; gene expression

Graphical Abstract
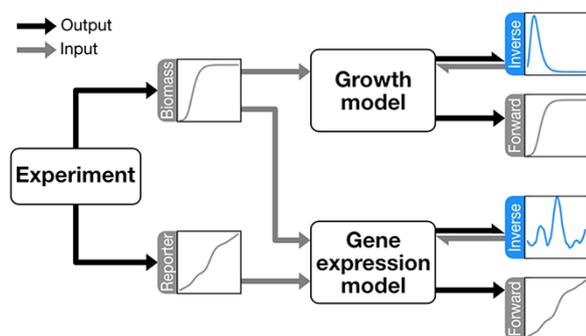
1. Introduction

Gene expression and growth are subject to variation due to changes in environmental and internal conditions, which may be divided into cellular and compositional contexts.

The cellular context is related to the cell or chassis itself such as the strain and to operational conditions such as media and carbon source. Microbial populations themselves create intrinsically dynamic conditions which can be separated into distinct phases of the growth cycle. An initial lag phase of adaptation is followed by exponential growth and finally a minimal growth stationary phase (1). These transitions in growth phases are driven by cell internal changes caused by depletion

of nutrients, accumulation of waste products or due to biochemical and physical signals (1). These signals can be part of the environment that the cell senses and responds to and can be experimentally determined.

Cellular context changes the metabolic activity of the cell and leads to distinct phenotypes (2). The effects of changes in external factors on growth and gene expression levels have been extensively studied (3–5) and correlations between them established. For example, ribosome and RNA polymerase are positively correlated with peak growth rate in different media (6). However, changes in the dynamics of gene expression due to cellular context are less clear (7), since different genes are known to be expressed differentially in each growth phase. This is at least partly due to the regulation of distinct sigma factor RNA polymerase sub-units, such as σ_{70} which peaks in exponential growth phase (8), as well as potentially gene-specific transcription factors, ribosome numbers and other translation factors (9).

The compositional context is related to the composition or sequence of the inserted DNA and can be seen at the transcription unit (TU), plasmid or genomic level. Synthetic biologists separate DNA into standard parts with determined sequences and function (10). A composition of parts that enables transcription is a TU, and the transcribed RNA can be then translated into proteins. A TU capable of gene expression can be assembled with the parts promoter (Pro), ribosome-binding site (RBS), coding sequence (CDS) and terminator (Ter) in that exact composition order. The behavior of these parts varies depending on the surrounding sequences. While the magnitude of gene expression from a TU is known to be influenced by its surrounding sequences, orientation or its position relative to other parts in a plasmid or in the genome, their effect on dynamics is not well understood (11–14).

Measuring and analyzing the dynamics of gene expression are thus fundamental to understanding cellular regulation by both natural and synthetic gene networks in the face of different cellular and compositional contexts. Typical experiments to measure gene expression rates in bacteria and other microorganisms utilize fluorescent reporters to track the expression levels of lineages of cells (7,15–17). The total biomass of these lineages is also tracked, typically using optical density measurements or colony size (18, 19). The genes measured are often fusions of promoters of interest with a downstream fluorescent reporter, and their expression rate profiles are taken to be indicative of the transcription rate of the promoter (20, 7). Promoters have been characterized relative to a standard promoter (20–22) that is measured under the same conditions as the promoter of interest. This approach has been used under steady-state conditions but has not been applied to dynamics. Thus, there is a need for methods to characterize the dynamics of gene expression and growth as phenotypic parameters.

Reconstructing dynamic microbial gene expression and growth rate profiles from data is difficult because, particularly at low biomass, measurements suffer from significant noise. Typical methods involve data smoothing and differentiation of the resulting signal—herein referred to as the indirect method (23–26). We consider two data smoothing filters; an anti-causal zero-phase digital filter (27) and the Savitzky–Golay polynomial filter (28). The indirect methods are sensitive to noise leading to the development of a more robust method based on the linear inversion of differential Equation models (29)—herein referred to as the direct method. Inverse problems present an approach to infer the values of parameters or functions on which measurements depend (30).

The basic requirements to solve them are appropriate measurements and a mathematical model of the process that generates them. Two difficulties in inverse problems are that it cannot determine correlated parameters and that most problems of interest are ill-posed. An ill-posed problem is one that does not satisfy one or more of the well-posed properties: a solution exists, it is unique and its behavior changes continuously with the initial conditions. However, regularization methods, basis transformations and constraints can often be used to transform ill-posed into similar well-posed problems.

We present a method for the reconstruction of gene expression and growth rate profiles using inverse problems that achieves several times lower error than the direct and indirect methods. Our approach more accurately reproduced features of dynamic gene expression and growth rate profiles, including the lag phase and peak rates. Using this method, we characterized the dynamics of a collection of synthetic TUs in different cellular and compositional contexts relative to an *in vivo* reference, revealing uncertainty due to the genetic composition and external environmental factors. Our results suggest an approach to ratiometric characterization of the dynamics of gene expression.

2. Results
2.1 Gene Expression and Growth Dynamics

The rate at which a protein is synthesized by a genetic network or circuit varies over time, giving rise to dynamic gene expression rate profiles, which may include rich behaviors such as bistability (31) and oscillations (32). To see the importance of measuring these gene expression rate profiles, consider a genetic circuit composed of regulatory proteins. In the typical case of short half-life mRNAs, we may assume quasi-steady state and,

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \phi_i(p_0, p_1, ..., p_{N-1}, t) - \gamma_i p_i - \mu(t)p_i, \tag{1}$$

for a genetic circuit with N proteins at concentrations p_i, with degradation rates γ_i, in cells growing at rate $\mu(t)$. The functions ϕ_i define the interactions in the circuit by mapping the protein concentrations to protein synthesis or gene expression rates. The dependence on time is due to variation in the cellular context such as growth phase transitions, which cause observable changes in the growth rate profile $\mu(t)$. It is, therefore, essential for the analysis of genetic circuit operation to estimate the gene expression rate profiles ϕ_i and growth rate profile μ, allowing the parameterization of models such as those in Equation 1.

The simplest genetic circuits consist of constitutive TUs which do not regulate each other, in which case $\phi_i(t)$ is only a function of time. To model the growth and reporter gene expression measurements from such a circuit in a population of cells, we use the following equations:

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \mu(t)B, \tag{2}$$

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = B(t)\phi_i(t) - \gamma_i y_i, \tag{3}$$

where B is a measure of sample biomass, $\mu(t)$ is the instantaneous relative growth rate, y_i is the intensity of reporter, $\phi_i(t)$ is the instantaneous expression rate and γ_i is the reporter degradation rate for TU i. Here we have assumed that reporter intensity $y_i = Bp_i$, with p_i the protein concentration per biomass. While this is a reasonable assumption in constant conditions (33), this relationship may not always hold leading to inaccuracy in the

reconstructed profiles, which is a general problem with gene expression reporters. More complex measurement models might be constructed, given sufficient information about the transformation of gene expression level into reporter intensity. We leave this for future work. From these equations we wish to accurately and robustly estimate the growth rate $\mu(t)$ and the expression rates $\phi_i(t)$, given an estimate of the reporter degradation rates γ_i. Typically, reporter proteins are stable and so in the following we will use $\gamma_i = 0$ (34) (see Supporting Information, Figure S1 for the effect of underestimating γ_i).

Note that the expression rate $\phi(t)$ is different from the rate of change of fluorescence dy/dt and from the rate of change of fluorescence concentration $d(y/B)/dt$, which both depend on protein degradation and dilution due to growth (and may be negative). What is being estimated in this work is the underlying fluorescent protein synthesis rate $\phi(t)$ (strictly positive), which is independent of dilution and degradation processes and which we assume is proportional to the underlying cellular gene expression rate (see Discussion).

2.2 Dynamics Estimation as an Inverse Problem

Reconstructing the functions $\mu(t)$ and $\phi(t)$ represents an inverse problem, which is underdetermined and ill posed (30). In order to reduce the dimensionality of the inverse problem, we exploit prior knowledge of the functions $\mu(t)$ and $\phi(t)$ to construct a simple basis as follows. Expression rates and growth rates may be reasonably assumed to be strictly positive and smooth on typical timescales of transcription and translation. We propose the following approximation, given a function $f(t)$ that meets our assumptions,

$$f(t) \approx \sum_{k=0}^{n-1} \hat{f}_k G_k(t), \qquad (4)$$

with

$$G_k(t) = \exp\left(\frac{-(t-k\Delta)^2}{2\Delta}\right), \qquad (5)$$

which represents a sum of n Gaussian curves G_k with weight $\hat{f}_k$, variance Δ and regularly spaced over time t at intervals Δ. Here Δ determines the timescale of variation or smoothness of the representation of the function $f(t)$. Since here we consider bulk culture experiments, gene expression bursting and noise in growth cannot be observed, and the relevant timescales are the gene expression reporter half-life and the culture doubling time, typically <1 h. Choosing Δ greater than the sampling interval of the data makes the system over-determined and regularized in the sense that it is constrained to be smooth. In the following we choose $\Delta = 1$ h, the typical timescale of protein synthesis, which is larger than the usual sampling interval of 10–15 min (see Methods). In Figure 1A and B we can see examples of reconstructed growth rate and gene expression rate profiles in turquoise and the function to be fitted in dashed black. The effect of this approximate Gaussian basis can be seen from the dependence on Δ of the maximum slope of the basis functions $G_k(t)$, which scales as $1/\sqrt{\Delta}$. This means that the sharpest change in expression rate and growth rate profiles that can be reconstructed by our method is determined by Δ (see Supporting Information, Figure S2 for effect of Δ on reconstruction error).

The model given in Equations 2 and 3 combined with the approximation of Equations 4 and 5 represents the forward
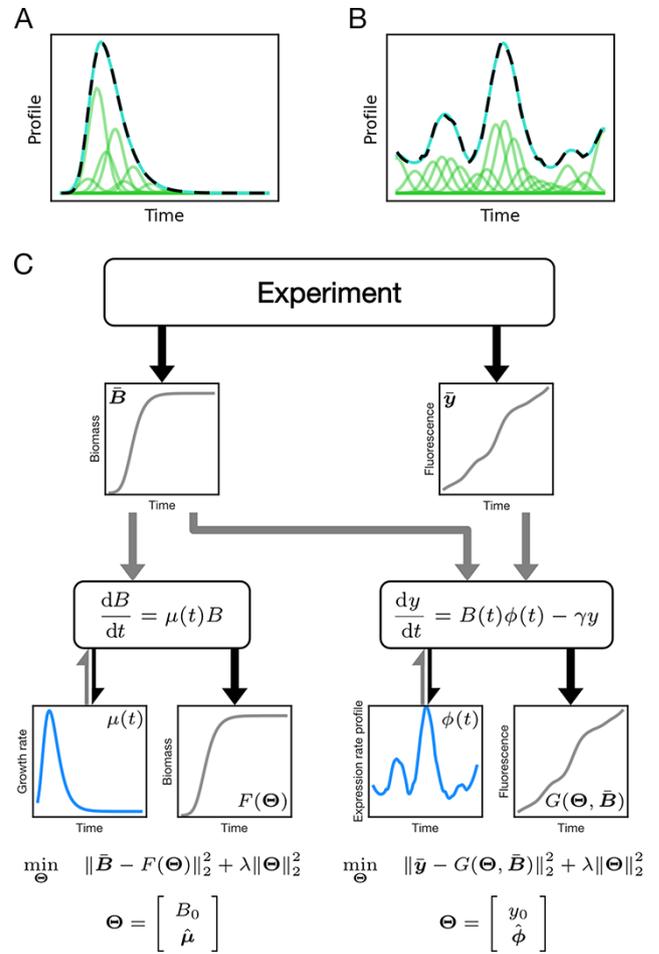

Figure 1. Algorithm overview. A, B Growth and gene expression rates approximation using Gaussian basis. The light green curves are individual Gaussian curves that compose the basis. The turquoise line represents the sum of the Gaussian basis. The black dashed line represents the growth rate profile (A) or gene expression rate profile (B) to be fitted. (C) Inverse problem algorithm diagram. Once an experiment is performed, biomass data ($\bar{B}$) and fluorescence data ($\bar{y}$) are collected. These data are input to the models, where the biomass model requires only biomass as input and the fluorescence model requires both biomass and fluorescence. Finally, growth rate ($\mu(t)$) and expression rate ($\phi(t)$) profiles are generated from the biomass and fluorescence models, respectively. The gray arrows indicate inputs and black arrows indicate outputs.

models of the inverse problems for the reconstruction of $\mu(t) \approx \sum_k \hat{\mu}_k G_k(t)$ and $\phi(t) \approx \sum_k \hat{\phi}_k G_k(t)$. In practice, the measurements used to estimate B and y are discrete, will contain background signal and are subject to noise. The background signals B' and y' are typically estimated by measuring appropriate control samples containing no cells (B') and cells with no reporter expression (y') (see Methods). After subtracting these background measurements, we are left with the noisy estimates $\bar{B}$ and $\bar{y}$. We then wish to parameterize the forward models given by Equations 2 and 3 such that $\|y-\bar{y}\|_2^2$ and $\|B-\bar{B}\|_2^2$ are minimized.

Note that while this approach uses a smooth basis to represent the reconstructed profiles, it does not filter or smooth the input data as in the indirect method. Here we approximate the growth and gene expression rate profiles using a superposition of Gaussian functions to represent a continuous function as a discrete vector of parameters. The measurements and the estimated parameters are inputs of the forward model.

The forward model then generates simulated measurements using different estimated parameters, and we compute the ones that minimize the difference between the measurements and the model.

2.3 Accurate Estimation of Growth Rate Dynamics

The inverse problem for the reconstruction of the growth rate $\mu(t)$ can be stated as,

$$\min_{\Theta} \quad \|\bar{B} - F(\Theta)\|_2^2 + \lambda\|\Theta\|_2^2 \tag{6}$$

with,

$$\Theta = \begin{bmatrix} B_0 \\ \hat{\mu} \end{bmatrix}. \tag{7}$$

We minimize the difference between noisy estimates of Biomass $\bar{B}$ and the result of the forward model $F(\Theta)$. Since this problem is ill-posed we regularize using a Tikhonov penalty term λ. Θ is the vector of parameters to optimize, containing the initial biomass B_0 and the Gaussian basis weights $\hat{\mu}_k$ to represent $\mu(t) \approx \sum_{k=0}^{n-1} \hat{\mu}_k G_k(t)$. The hyperparameter λ was chosen to minimize error within a reasonable range of values (see Supporting Information, Figure S2).

This problem is a nonlinear least squares optimization, which we solve using the trust region reflective algorithm (35). We generated simulated data from Equations 2 and 3 using 100 randomly parameterized Gompertz growth models (1) and three different levels of measurement noise and characterized the growth rate using the inverse, direct and indirect methods (Supporting information Figures S3–S5). Hyperparameters for each method were optimized to minimize error within a reasonable range of values (see Supporting Information, Figure S2). We compare our results to the direct linear inversion method and show that our approach reduces mean squared error by more than 29-fold ($P < 10^{-35}$, Welch's t-test) (Figure 2A). Then we compared to the indirect method and show that the inverse method reduced the mean squared error by >2-fold ($P < 10^{-5}$, Welch's t-test) (Figure 2A). Finally, we compared to a different indirect method that uses an anti-causal zero-phase digital filter and show that the inverse method reduced the mean squared error by more than a 1000-fold (Supporting information Figure S17). The inverse method maintained the best performance at high noise levels. Furthermore, our method is more robust to noise in early biomass measurements, where the direct method overestimates the initial values and the indirect produces noisy reconstructions that usually goes below zero at the beginning (Figure 2B). The inverse method correctly reconstructed the lag phase, which
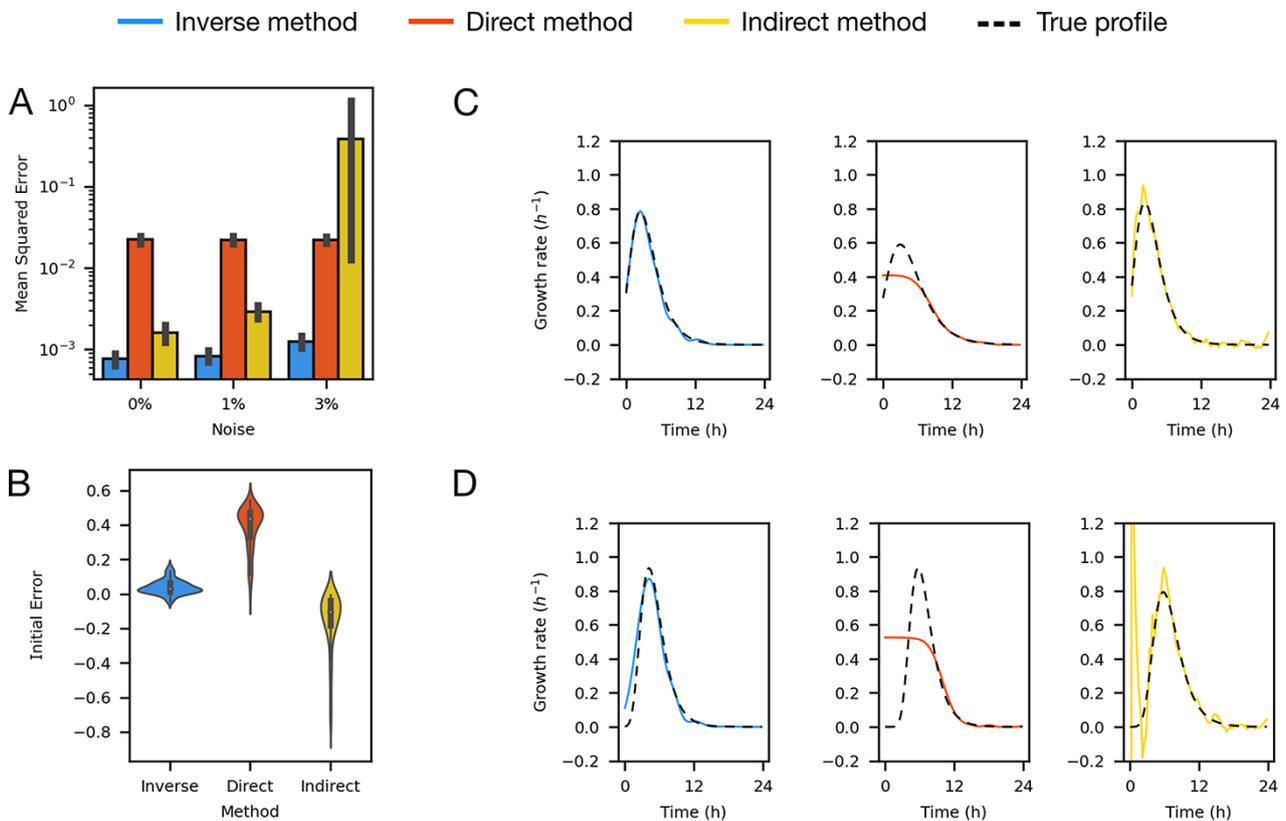

Figure 2. Method comparison on growth rate simulations. (A) The errors corresponding to the inverse, direct and indirect methods are presented, and the error of the inverse method was almost 30-fold lower than the direct method and 2-fold lower than the indirect method; error bars represent 95% confidence interval. (B) The error of the initial growth rate over all noise levels. Growth rate, due to its nature, has a peak, which is not easily identified with the direct method. Since the initial growth rate should be low, we can see that the direct method overestimates and the indirect method underestimates the initial growth rate, showing that the inverse method is more accurate at low biomass. The error calculated for growth rate values at time 0 compared to the true profile. (C) Examples of accurate growth rate reconstructions from simulated data. Gompertz profiles characterized using the inverse, direct and indirect methods (left, center and right, respectively) compared to the true profile (black dashed line). **D** Examples of inaccurate growth rate reconstructions from simulated data. Gompertz profiles characterized using the inverse, direct and indirect methods (left, center and right, respectively) compared to the true profile (black dashed line) ($n = 100$).

is missing from the linear inversion solution. Our method also reconstructed better the growth rate peak, effectively distinguishing between lag, exponential and stationary growth phases (Figure 2C and D).

2.4 Accurate Estimation of Gene Expression Rate Dynamics

In a similar way to growth rate, we wish to find the optimal parameters,

$$\Theta = \left[\begin{array}{c} y_0 \\ \hat{\phi} \end{array} \right], \tag{8}$$

where Θ is the vector of parameters to optimize, containing the initial reporter intensity y_0 and the Gaussian basis with weights $\hat{\phi}_k$ and the forward model $G(\Theta, \bar{B})$. The problem is again a nonlinear least squares optimization,

$$\min_{\Theta} \quad \|\bar{y} - G(\Theta, \bar{B})\|_2^2 + \lambda\|\Theta\|_2^2 \tag{9}$$

which we solve using the same numerical procedure as for growth rate. To test this approach, we generated 100 random gene expression rate profiles from smoothed lognormal random walks, with random Gompertz models for the biomass, and three measurement noise levels (Supporting information Figures S6–S8) (see Methods). Again, we compared the inverse method to the direct linear inversion method and show that the mean squared error is more than 4-fold lower ($P < 10^{-10}$, Welch's *t*-test) and close to 3-fold lower than the indirect method ($P < 10^{-7}$ Welch's *t*-test) (Figure 3A). Finally, we compared to a different indirect method that uses an anti-causal zero-phase digital filter and show that the inverse method reduced the mean squared error by more than a 100-fold (Supporting information Figure S17). The gene expression simulations are not constrained to start with low values or to have an initial peak which make the three methods have similar initial errors (Figure 3B). We find that the direct method does not correctly reconstruct early peaks in gene expression rate profiles and that the indirect method produces extremely noisy solutions (Figure 3C and D). The inverse method is the best performer capturing the shape of the growth rate profile, although this method suffers with sharp changes due to its dependence with Δ (Figure 3C and D). The inverse method requires knowledge of the timescale of the process that you want to measure to set the hyperparameter Δ (see Supporting information).

2.5 Characterizing Growth and Gene Expression Rate Dynamics in *Escherichia coli*

Using the inverse, direct and indirect methods we reconstructed growth and gene expression dynamics from experimental data of *Escherichia coli* carrying a synthetic triple TU plasmid pAAA (Figure 4A and B). This plasmid contains three TUs with the
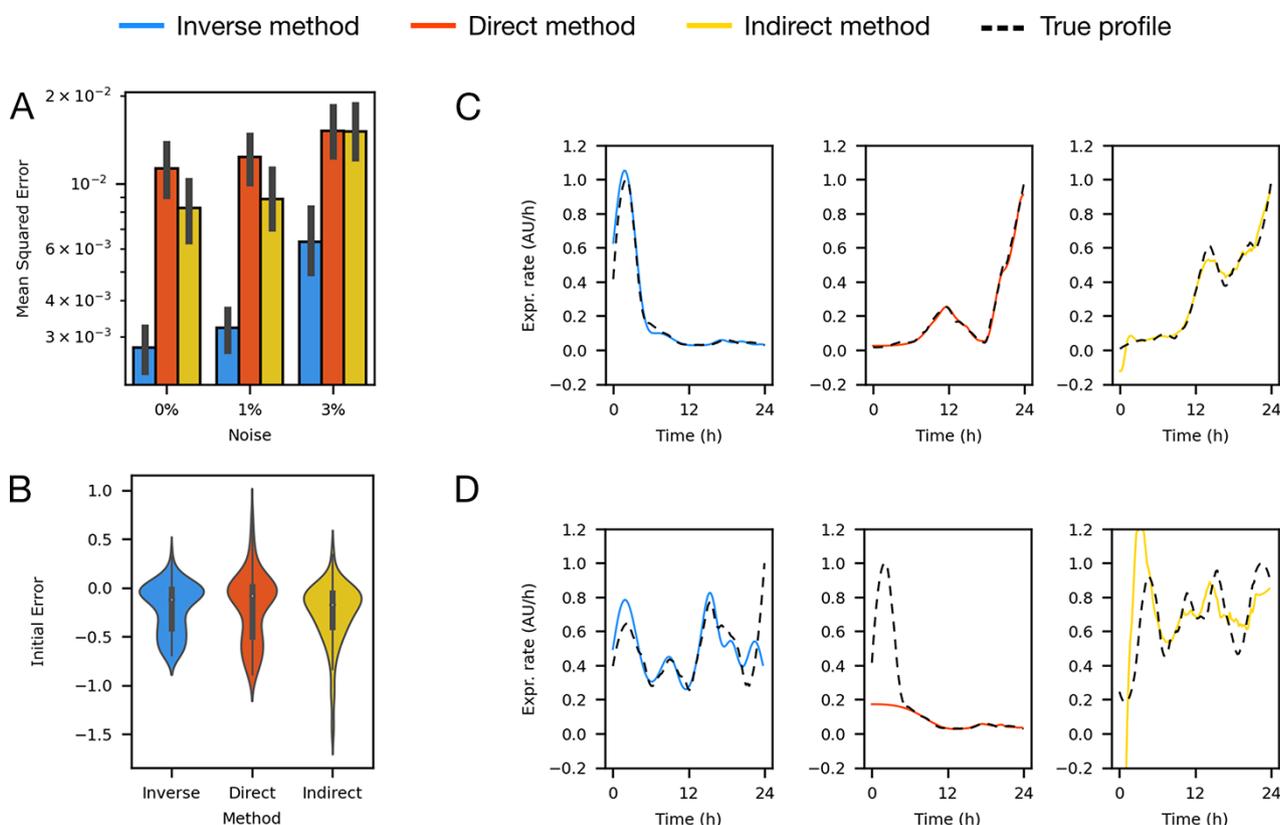

Figure 3. Method comparison on gene expression rate simulations. (A) The errors corresponding to the inverse, direct and indirect methods are presented; the inverse method has been shown to be 4-fold better than the direct method and close to 3-fold better than the indirect method, error bars represent 95% confidence interval. (B) The error at initial time. Since the initial gene expression rate simulations were not restricted to be low at the beginning, the three methods show similar errors. The error calculated for expression rate values at time 0 compared to the true profile. (C) Representative examples of accurate gene expression reconstructions from simulated data. Random profiles characterized using the inverse, direct and indirect methods (left, center and right, respectively) compared to the true profile (black dashed line). (D) Representative examples of inaccurate gene expression rate reconstructions from simulated data. Random profiles characterized using the inverse, direct and indirect methods (left, center and right, respectively) compared to the true profile (black dashed line) ($n = 100$).

same synthetic σ_{70} constitutive promoter J23101 (36) in different plasmid compositional contexts determined by its position in the plasmid and in different TU compositional contexts determined by different sets of promoter downstream elements RBS-CDS-Terminator (Figure 4A). Each CDS encoded a different fluorescent protein as reporter. Assays were performed using a 96-well microplate reader to measure fluorescence in each reporter channel as a proxy for protein concentration and optical density as a proxy for biomass (see Methods). To assess the effects of cellular context on growth and gene expression dynamics we measured two strains of *E. coli* carrying the pAAA plasmid, growing on two different carbon sources (26) (Figure 4B–E) (see Supporting information, Figure S9 for an example of the raw fluorescence and OD 600 data).
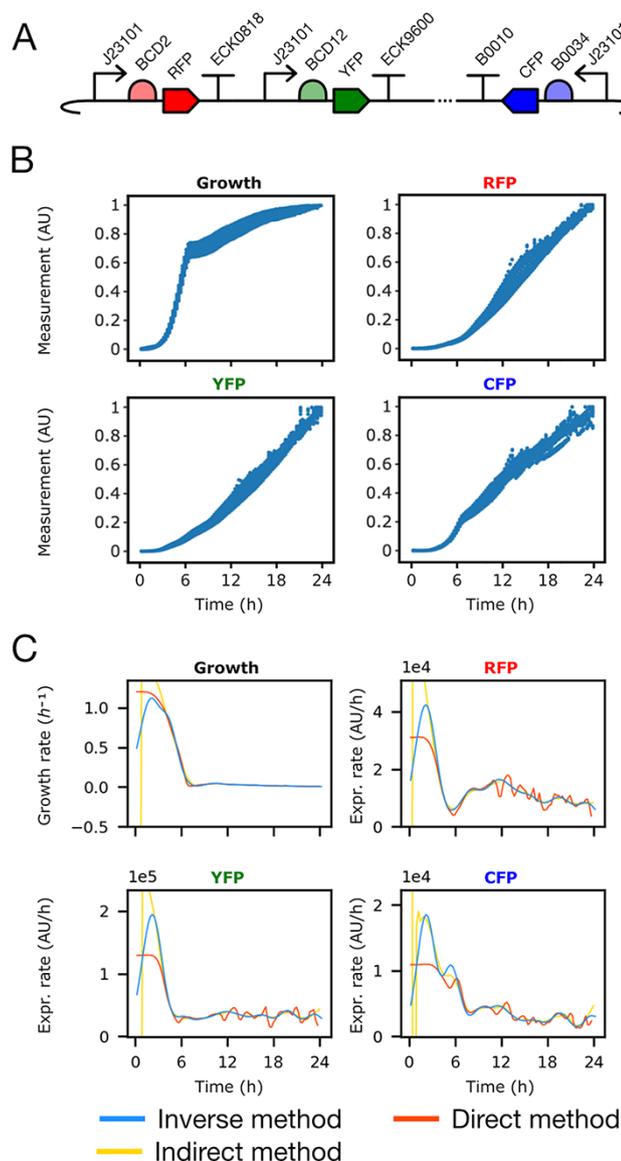
Figure 4. Gene expression dynamics shape reconstructed from experimental data is different for transcription units with the same promoter. (A) Plasmid pAAA synthetic biology open language (SBOL) visual diagram. (B) Raw experimental data of optic density (OD) and fluorescence. (C) Growth and gene expression rates reconstructed with the inverse, direct and indirect methods (blue, red and yellow lines, respectively) on strain MG1655z1 in media M9-glucose ($n = 30$).

The inverse method captured the shape of the growth and gene expression rate profiles in a smooth way, consistent with population averages on the timescale of protein synthesis, while the direct and indirect methods produced noisy solutions. Our method reconstructed a peak in growth rate consistent with the transition between lag, exponential and stationary phases, as in the Gompertz growth model (37). Furthermore, it captured a peak in the expression rate coincident with the growth rate peak, which is consistent with promoter dependence on σ_{70} and the abundance of this factor, as well as ribosomes, during peak growth (Figure 4B–E) (6, 9).

The gene expression rate profiles were different for TUs in the same plasmid under the control of the same promoter due to both different cellular and compositional contexts (Figure 4B–E). While in all cases peak gene expression rate coincided with peak growth rate, in some contexts multiple peaks were observed. The cellular context (strain and carbon source) as well as the compositional context (promoter downstream elements and position and orientation in the plasmid) clearly change gene expression and growth dynamics, leading to different peak timing and overall shape. All growth rates characterized using the inverse method exhibited clear lag, exponential and stationary phases, which are not apparent with the direct method (Figure 4B–E). The timing of growth phase transitions was different in each cellular context (Figure 4B–E).

2.6 Characterization of Gene Expression Rate Dynamics Relative to *in vivo* Reference

Gene expression magnitude has been characterized relative to a standard *in vivo* reference containing promoter J23101, in order to normalize for cellular context (22). Plasmid pAAA provides such a reference, with three TUs containing the same J23101 promoter in different compositional contexts. We hypothesized that this reference plasmid could be used to characterize the dynamics of gene expression in a standardized fashion. Each TU in the pAAA plasmid presents a standard reference for a particular compositional context—the promoter downstream elements, position and orientation in the plasmid. We then wish to characterize TUs with arbitrary promoters relative to these reference TUs, allowing us to describe their dynamics in a concise way. In order to compare gene expression rate dynamics from different experiments we synchronized profiles and normalized each one by subtracting its mean and dividing by its standard deviation. The timing of growth phase transitions is variable due to differences in initial conditions and experimental variability, which leads to differences in the timing of gene expression rate profiles. In order to correct for these differences, we used the reconstructed growth rate peak time t_0 to synchronize the expression rate profiles, shifting time to $\tau = t - t_0$, such that $\tau = 0$ is the time of peak growth rate.

We tested this approach on a collection of 14 combinatorial three-reporter plasmids, combining 10 different TUs which were each driven by one of seven promoters (38). Each plasmid contained three TUs producing red fluorescent protein (RFP), yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP), containing the same promoter downstream elements as the corresponding reference TU. The promoter downstream elements RBS, CDS and terminator for each reporter were maintained constant and we refer to them using the reporter name (Supporting information Tables S1 and S2). The CFP TU was maintained the same in all plasmids, to serve as a control (22, 20). This collection of plasmids presented a variety of promoters in different TU compositional contexts, that is, with different promoter downstream elements.
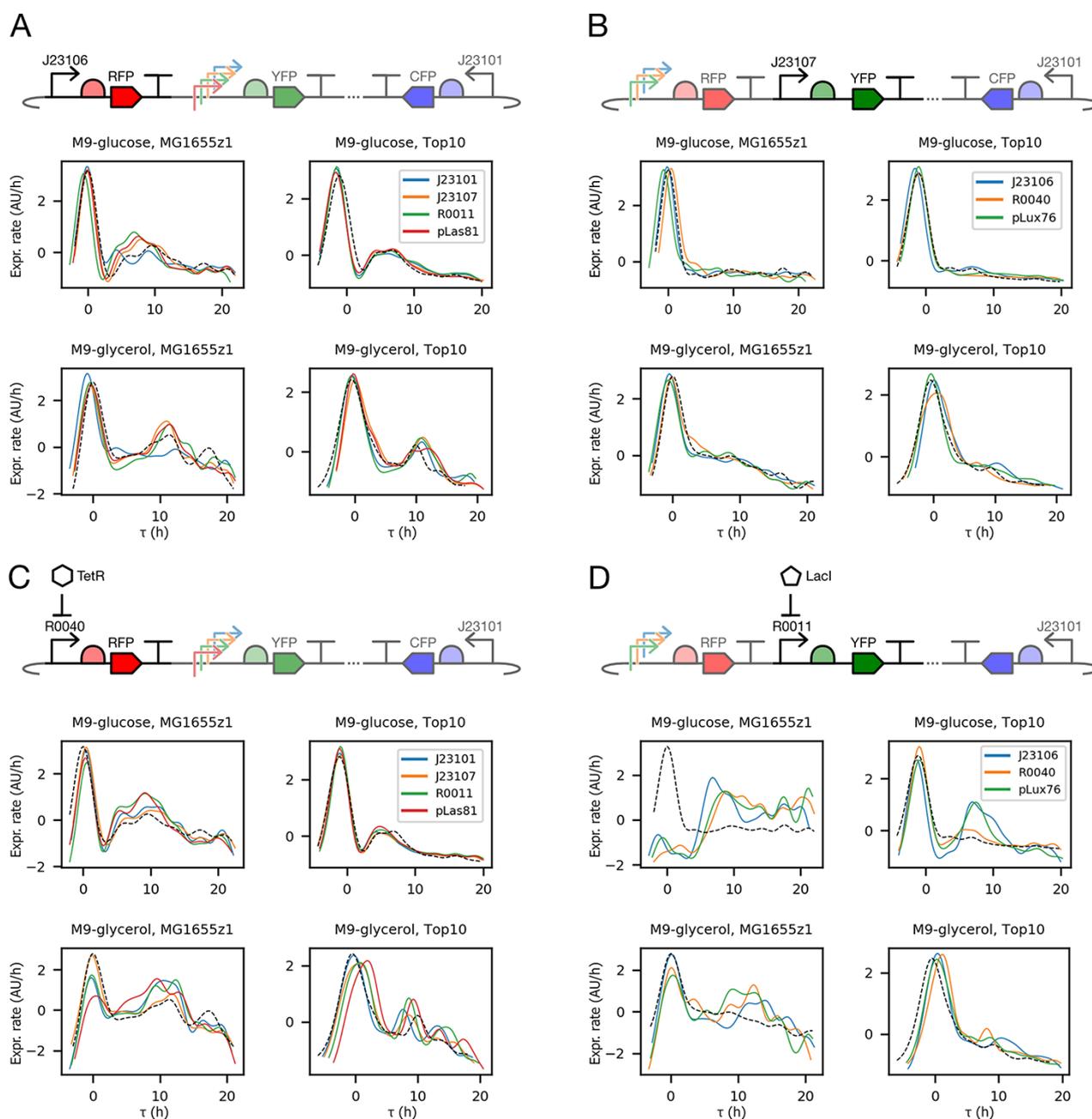
Figure 5. The gene expression profile shape reconstructed from experimental data is similar for promoters under constitutive expression. We used TUs from pAAA as reference (black dashed line) for TUs within the same conditions but with different promoters (solid lines). (A) RFP reference TU compared to RFP TU with J23106 in different downstream TU compositional contexts. (B) YFP reference TU compared to YFP TU with J23107 in different upstream TU compositional contexts. (C) RFP reference TU compared to RFP TU with R0040 basal expression in different downstream TU compositional contexts. (D) Reference YFP TU compared to YFP TU with R0011 with different upstream TUs compositional context. All profiles are shown normalized by mean and standard deviation ($n = 30$).

Each TU was assembled into multiple plasmid compositional contexts, that is, in the presence of different upstream or downstream TUs.

Three of the TUs in the collection contained a promoter from a family of constitutive promoters created by mutating a consensus sequence (36), which includes the reference TU promoter, J23101. The RFP reference TU matched the gene expression rate profile shape of the RFP TU with the constitutive promoter J23106 over different downstream TU contexts (Figure 5A). The YFP reference TU matched the gene expression profile shape of the YFP TU with the constitutive promoters J23107 (Figure 5B) and J23101

(Supporting information Figure S10B) over different upstream TU contexts. The CFP TU was consistent across all compositional contexts (Supporting information Figure S10A). These results suggest that the dynamics are not affected by the tested plasmid compositional contexts.

Two of the TUs contained promoters that are repressible by a transcription regulating protein that binds to the promoter. In the case of the TetR-responsive TU (containing promoter R0040), the RFP reference TU matched the gene expression rate profile over different downstream TU contexts and over different cellular contexts (Figure 5C). This is in spite of the fact that the genome of the

strain MG1655z1 contains a constitutive TetR gene, while Top10 does not produce the protein. This result suggests that the dynamics of the TetR-repressible TU are not affected by the action of the repressor.

The other repressible TU contained promoter R0011, regulated by LacI. LacI is produced by both strains and partly regulated by cAMP in Top10 (39). The YFP reference TU matched the LacI-repressible YFP TU gene expression rate profiles in M9-glycerol but did not match them in M9-glucose (Figure 5D). In MG1655z1 growing on glucose gene expression rates became negatively correlated with growth rate, and in Top10 on glucose they exhibited a second peak. These results suggest that transcription regulation can significantly affect the dynamics of gene expression compared to constitutive expression profiles.

The remaining two TUs contained promoters activated by different one-component signaling systems, measured in the absence of signal and regulator to study their basal expression. The RFP TU responsive to C6 homoserine lactone, containing promoter pLux76, was consistent with the constitutive reference TU in all contexts (Supporting information Figure S10C). However, for the YFP TU responsive to C12 homoserine lactone, containing promoter pLas81, the gene expression rate profile inverted its correlation with growth rate when downstream of the TU containing promoter R0040 and growing on glucose in strain MG1655z1. All the magnitudes of these measurements can be found in Supporting information Figure S13. These results show the uncertainty that can be introduced in gene expression rate dynamics due to changing compositional and cellular context (Supporting information Figure S10D).

3. Discussion

The accurate characterization of dynamic gene expression and growth rate profiles is essential for the characterization of genetic circuits and the inference of gene regulatory interactions in natural networks (40–42). We have demonstrated an inverse problem approach to reconstructing dynamic gene expression rate and growth rate profiles from noisy kinetic measurement data. We compared our method to the current state-of-the-art algorithms, direct linear inversion (29) and indirect smoothing and differentiation (25). Our approach reduced the mean squared error of reconstructions from simulated data of growth rate by almost 30-fold and gene expression rate by more than 4-fold with respect to the direct method.

The comparison showed that the direct method often fails to capture peaks at the beginning of the profiles. Indirect methods, even after filtering the noise, fail to reconstruct early stages of growth. This is likely because the biomass can have very low and even negative values after background correction and dividing by these values can result in the amplification of noise. This highlights that particular attention should be paid to reducing noise from experimental procedures and measurement techniques since all methods attain lower reconstruction errors with less noise. Surprisingly, our indirect method often performed better than the direct method, but our inverse problems approach improved on the indirect method by almost 3-fold for gene expression rates, and more than 2-fold for growth rate. Furthermore, we were able to reconstruct features of both growth rate and gene expression rate profiles, such as exponential phase and peak growth, that were not apparent from the direct linear inversion method nor the indirect method. The growth rate peak is an important feature captured better with

our method, allowing the synchronization of gene expression rate profiles.

While in terms of computation time our method is relatively slow, it is the most intuitive to adjust by estimating the timescale of dynamics to obtain Δ. The indirect methods require knowledge of the signal processing filters used, and the direct method requires tuning the 'insignificant value' which is rather obscure.

Using our method we showed that the dynamic form of gene expression rates, not only their magnitude, is determined both by cellular and compositional contexts (Figure 5, Supporting information Figures S10 and S13–S15). We examined two types of compositional context. Firstly, the composition of parts within a TU, not only the sequence of the promoter, determined the dynamics of gene expression rates. Secondly, gene expression rate profiles were largely independent of the context in which the TU was placed, that is the upstream and downstream TUs. In most cases, gene expression rates peaked in exponential growth phase, with some promoters exhibiting a second peak in stationary phase. This may be an effect of the abundance of different sigma factors in each growth phase and the sensitivity of the promoter to them. It is known that the binding sites for σ_{70}, most abundant in exponential growth phase, and σ_S, most abundant in stationary growth phase, are very similar (43). Therefore, a promoter with peaks in both exponential and stationary phases may be activated by both σ_{70} and σ_S to different extents and the peaks caused by variations in the sigma factor abundance in each growth phase. In some cases gene expression rate profiles inverted their correlation with growth rate, highlighting the uncertainty introduced by changing circuit composition. This uncertainty may be due to various mechanisms that modulate gene expression at sequence level such as DNA supercoiling (12).

Furthermore, our results also showed that cellular context, that is media and strain, changed dynamic gene expression and growth rate profiles. Our results suggest that while the dynamic characterization for TUs under constitutive or leaky expression relative to an *in vivo* reference could be useful, uncertainty due to compositional and cellular context must be taken into account. This highlights the need for strategies to mitigate compositional context effects, such as gene expression load (44), five prime untranslated region (45) as well as techniques to predict interactions of genetic elements at the sequence level (11,21,46).

Typically, measurements of promoter–reporter fusions are used as a proxy for transcription rates under various levels of transcription factors, external signals and other determinants of gene expression (20,21,47,48). However, applying our method to multiple reporter TUs with the same promoter, we showed that gene expression rate profiles should not be taken as indicative of intrinsic characteristics of promoters, since they are affected by both the promoter and downstream genetic elements, gene position and orientation and external factors such as carbon source and host strain (see Supporting information Figure S16).

Synthetic biology aims to design novel genetic networks or circuits from compositions of transcription units. It relies heavily on the characterization of the functions of these TUs. Fundamentally, gene expression rates must be reconstructed from noisy measurement data in a range of conditions, including concentrations of inducer chemicals. The function of the circuit may then be mathematically modeled as in Equation 1. Methods such as ours will enable such approaches for dynamical systems (32, 31) where the dynamic profile of gene expression rates is essential to the operation of the circuit.

In order to model circuit operation in this way, calibration of the fluorescence and biomass signals with respect to standard references (15, 18) or the use of relative (ratiometric) quantification (21, 20) are necessary and can be easily incorporated into our workflow. Our results show that profiles for constitutive gene expression were consistent across a range of promoters in a range of contexts suggesting that an *in vivo* reference gene may be used to infer the expression profiles of other genes in a circuit without directly measuring them. The inverse problems approach provides a framework that could be extended easily to fit more complex models of gene expression and also for regulatory parameters (e.g. Hill functions) as well as dynamic profiles, providing an accurate and flexible characterization method for synthetic biology.

4. Methods
4.1 Kinetic Gene Expression and Growth Assays
Kinetic gene expression and growth assays were made culturing triple reporter plasmid pAAA containing bacteria (Figure 4A) using two different strains as well as two different carbon sources, as described in Flapjack (26).

Monoclonal colonies of *E. coli* strain TOP10 or MG1655z1 transformed with plasmid DNA were picked and cultured for 14–15 h overnight in M9 media with $50\,\mu g$/ml kanamycin, 0.2% w/v casaminoacids and 0.4% w/v glucose or 0.4% w/v glycerol. Overnight cultures were diluted 1000 times in 2-ml tubes. All the tubes were filled with $1996\,\mu l$ of fresh M9 media, $2\,\mu l$ of kanamycin and $2\,\mu l$ of the bacteria liquid culture obtaining a final volume of 2 ml. In each well of a 96-well plate were added $200\,\mu$ l, 4 wells with M9 media with the proper carbon source and kanamycin, 4 wells with non-transformed bacteria of the same strain and 10 wells of bacteria transformed with the appropriate plasmid to analyze from the previously prepared 2-ml tubes. Optical density and fluorescence in three channels (RFP, YFP and CFP) were measured approximately every 15 min for 24 h in a Synergy HTX plate reader with Gen5 software. Each assay was repeated on three different days, with 10 replicates on each day, and each 96 well plate contained experiments with the same carbon source and strain following the methods in Flapjack (26).

4.2 Computational Methods
Computational methods were performed using Flapjack (26), Python (49), Numpy (50), Scipy (51), Pandas (52), Matplotlib (53), Plotly (54), Jupyter (55), Google Colaboratory (56) and Matlab. The direct method was computed using the WellFARE package (29). The indirect method with anti-causal zero-phase digital filter was computed using the Matlab function 'butter' (27).

4.3 Web-based Software Implementation
Flapjack (26) (http://flapjack.rudge-lab.org) was extended to compute gene expression rate and growth rate profiles using the methods described above, using the WellFARE package (29) for the direct method. The indirect method is computed by Flapjack by filtering the measured signals (biomass and reporter levels) using a Savitzky–Golay filter (28) and then differentiating the resulting smooth polynomial interpolation.

Flapjack is a systems and synthetic biology data storage and analysis tool, built as a web app that provides a user-friendly web interface and a representational state transfer (REST) and web socket application programming interface (API). The system allows upload of kinetic gene expression data from a variety of sources and links it to metadata about experimental conditions. These data may then be queried and filtered and used to reconstruct gene expression and growth rates. Flapjack automatically subtracts background signal from both reporter and biomass measurements, taking the average of control samples (untransformed cells or media with no cells) at each time point.

4.4 Random Profile Generation
In order to generate a range of gene expression rate profiles, with minimal assumptions about their form, we generate lognormal random walks as,

$$\phi_t = \prod_{i=0}^{t} \xi_i \tag{10}$$

with $\log(\xi_i) \sim N(0,\sigma^2)$ and $\sigma^2 = 0.25$. The profiles ϕ_t were then smoothed using a second-order Savitzky–Golay filter (28) with window size 21 and normalized to $[0,1]$. To generate random growth rate profiles, we used the Gompertz equation,

$$\log\left(\frac{B}{B_0}\right) = A \exp\left(-\exp\left(\frac{\mu^* e}{A}(\lambda - t) + 1\right)\right) \tag{11}$$

with μ^* the maximal growth rate uniformly distributed on $[0.5,1]$ per hour, λ the lag phase length uniformly distributed in $[0,4]$ hours, $A = \log(B^*/B_0)$, where the maximal biomass $B^* = 1$ and minimum biomass $B_0 = 0.01$. The growth rate profile implied by this equation is given by,

$$\mu(t) = \mu^* \exp\left[\frac{\mu^* e \cdot (\lambda - t)}{A} - \exp\left(\frac{\mu^* e \cdot (\lambda - t)}{A} + 1\right) + 2\right]. \tag{12}$$

4.5 Simulation of Kinetic Gene Expression and Biomass Data
Equations 2 and 3 were solved using the forward Euler integration scheme with time step $\Delta t = 2.4 \times 10^{-2}$ h for a period of 24 h. Noise and background were added according to the following equations with $B' = 0.1$ and $y' = 0.1$,

$$B_t = (B(t) + B')(1 + \epsilon_t) \tag{13}$$

$$y_t = (y(t) + y')(1 + \zeta_t), \tag{14}$$

where ϵ_t and ζ_t are uncorrelated white noise with variance σ^2, due to the measurement process. Simulated measurements were generated using LOICA (57), then uploaded to Flapjack (26) and analyzed using the API via Python (see Supporting information Figures S11 and S12 for an example of reporter and biomass raw data, respectively).

4.6 Choice of hyperparameters
Each of the methods tested in the main text is dependent on one or more hyperparameters. In the case of the direct method this is the so-called insignificant value ϵ_L (29), for the indirect method it is the Savitzky–Golay filter window size and for the inverse method they are Δ and λ. For reconstructions of simulated data these parameters were optimized by scanning a range of reasonable values and choosing the parameter that minimized the mean squared error (see Supporting information). For experimental data, the value of λ was chosen using the L-curve method (58), ϵ_L was taken from the original paper (29), the value of Δ was fixed at 1 h and the Savitzky–Golay window size fixed at 11. For the anti-causal zero-phase digital filter, a second-order Butterworth filter (27) with cut-off frequency 4/33 was used.

Supplementary Data

Supplementary data are available at *SYNBIO* online.

Data Availability Statement

The data are available at https://github.com/RudgeLab/Inverse_Characterization/tree/main/Data.

Author contributions statement

Conceptualization, T.J.R.; data curation, G.V., G.Y.F. and T.J.R.; formal analysis, G.V., C.V., C.C. and T.J.R; funding acquisition, F.F. and T.J.R.; investigation, G.V., C.V., M.M.S. and T.J.R.; methodology, G.V., C.V., M.M.S., F.F. and T.J.R.; project administration, T.J.R., G.V. and F.F.; resources, G.V., F.F. and T.J.R.; software, G.V., C.V., G.Y.F, C.C. and T.J.R.; supervision, T.J.R., F.F. and G.V.; validation, G.V.; visualization, C.V., G.V. and T.J.R.; writing—original draft, G.V., C.V. and T.J.R.; writing—review editing, G.V. C.V., G.Y.F., C.C., F.F. and T.J.R.

Conflict of interest statement. No potential conflict of interest was reported by the authors.

References

1. Zwietering,M.H., Jongenburger,I., Rombouts,F.M. and van't Riet,K. (1990) Modeling of the Bacterial Growth Curve. *Appl. Environ. Microbiol.*, **56**, 1875–1881. 10.1128/aem.56.6.1875-1881.1990.

2. Gefen,O., Fridman,O., Ronin,I. and Balaban,N.Q. (2014) Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 556–561. 10.1073/pnas.1314114111.

3. Klumpp,S., Zhang,Z. and Hwa,T. (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell*, **139**, 1366–1375. 10.1016/j.cell.2009.12.001.

4. Scott,M., Klumpp,S., Mateescu,E.M. and Hwa,T. (2014) Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.*, **10**, 747. 10.15252/msb.20145379.

5. Klumpp,S. and Hwa,T. (2014) Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Curr. Opin. Biotechnol.*, **28**, 96–102. 10.1016/j.copbio.2014.01.001.

6. Klumpp,S. and Hwa,T. (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20245–20250.

7. Berthoumieux,S., De Jong,H., Baptist,G., Pinel,C., Ranquet,C., Ropers,D. and Geiselmann,J. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.*, **9**, 634. 10.1038/msb.2012.70.

8. Jishage,M., Iwata,A., Ueda,S. and Ishihama,A. (1996) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.*, **178**, 5447–5451. 10.1128/jb.178.18.5447-5451.1996.

9. Scott,M., Gunderson,C.W., Mateescu,E.M., Zhang,Z. and Hwa,T. (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science*, **330**, 1099–1102. 10.1126/science.1192588.

10. McLaughlin,J.A., Beal,J., Misirli,G., Grunberg,R., Bartley,B.A., Scott-Brown,J., Vaidyanathan,P., Fontanarrosa,P., Oberortner,E. and Wipat,A. *et al.* (2020) The synthetic biology open language (SBOL) version 3: Simplified data exchange for bioengineering. *Front. Bioeng. Biotechnol.*, **8**. 10.3389/fbioe.2020.01009.

11. Carr,S.B., Beal,J. and Densmore,D.M. (2017) Reducing DNA context dependence in bacterial promoters. *PloS One*, **12**, e0176013. 10.1371/journal.pone.0176013.

12. Yeung,E., Dy,A.J., Martin,K.B., Ng,A.H., Del Vecchio,D., Beck,J.L., Collins,J.J. and Murray,R.M. (2017) Biophysical constraints arising from compositional context in synthetic gene networks. *Cell Syst.*, **5**, 11–24. 10.1016/j.cels.2017.06.001.

13. Vecchio,D.D., Ninfa,A.J. and Sontag,E.D. (2008) Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.*, **161**. 10.1038/msb4100204.

14. Vecchio,D.D. (2015) Modularity, context-dependence, and insulation in engineered biological circuits. *Trends Biotechnol.*, **33**, 111–119. 10.1016/j.tibtech.2014.11.009.

15. Beal,J., Haddock-Angelli,T., Baldwin,G., Gershater,M., Dwijayanti,A., Storch,M., De Mora,K., Lizarazo,M., Rettberg,R. and with the iGEM Interlab Study Contributors (2018) Quantification of bacterial fluorescence using independent calibrants. *PLoS One*, **13**, e0199432. 10.1371/journal.pone.0199432.

16. Lichten,C.A., White,R., Clark,I.B.N. and Swain,P.S. (2014) Unmixing of fluorescence spectra to resolve quantitative time-series measurements of gene expression in plate readers. *BMC Biotechnol.*, **14**, 1–11. 10.1186/1472-6750-14-11.

17. Kalir,S., McClure,J., Pabbaraju,K., Southward,C., Ronen,M., Leibler,S., Surette,M.G. and Alon,U. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science*, **292**, 2080–2083. 10.1126/science.1058758.

18. Beal,J., Farny,N.G., Haddock-Angelli,T., Selvarajah,V., Baldwin,G.S., Buckley-Taylor,R., Gershater,M., Kiga,D., Marken,J. and Sanchania,V., et al. (2020) Robust estimation of bacterial cell count from optical density. *Commun. Biol.*, **3**, 1–29. 10.1038/s42003-020-01127-5.

19. Nuñez,I., Matute,T., Herrera,R., Keymer,J., Marzullo,T., Rudge,T. and Federici,F. (2017) Low cost and open source multi-fluorescence imaging system for teaching and research in biology and bioengineering. *PLoS One*, **12**, 1–21. 10.1371/journal.pone.0187163.

20. Rudge,T.J., Brown,J.R., Federici,F., Dalchau,N., Phillips,A., Ajioka,J.W. and Haseloff,J. (2016) Characterization of intrinsic properties of promoters. *ACS Synth. Biol.*, **5**, 89–98. 10.1021/acssynbio.5b00116.

21. Nielsen,A.A.K., Der,B.S., Shin,J., Vaidyanathan,P., Paralanov,V., Strychalski,E.A., Ross,D., Densmore,D. and Voigt,C.A. (2016) Genetic circuit design automation. *Science*, **352**, aac7341-aac7341. 10.1126/science.aac7341.

22. Kelly,J.R., Rubin,A.J., Davis,J.H., Ajo-Franklin,C.M., Cumbers,J., Czar,M.J., de Mora,K., Glieberman,A.L., Monie,D.D. and Endy,D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.*, **4**, 4. 10.1186/1754-1611-3-4.

23. Ronen,M., Rosenberg,R., Shraiman,B.I.I. and Alon,U. (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 10555–10560. 10.1073/pnas.152046799.

24. Aïchaoui,L., Jules,M., Le Chat,L., Aymerich,S., Fromion,V. and Goelzer,A. (2012) BasyLiCA: a tool for automatic processing of a Bacterial Live Cell Array. *Bioinformatics*, **28**, 2705–2706. 10.1093/bioinformatics/bts422.

25. De Jong,H., Ranquet,C., Ropers,D., Pinel,C. and Geiselmann,J. (2010) Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst. Biol.*, **4**, 1–17. 10.1186/1752-0509-4-55.

26. Feliú,G.Y., Gómez,B.E., Berrocal,V.C., Silva,M.M., Nuñez,I.N., Matute,T.F., Medina,A.A., Vidal,G., Céspedes,C.V., Dahlin,J., Federici,F. and Rudge,T.J. (2020) Flapjack: Data management and analysis for genetic circuit characterization. *ACS Synth. Biol.*, **10**, 183–191. 10.1021/acssynbio.0c00554.

27. Butterworth,S. *et al.* (1930) On the theory of filter amplifiers. *Wirel. Netw.*, **7**, 536–541.

28. Savitzky,A. and Golay,M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, **36**, 1627–1639. 10.1021/ac60214a047.

29. Zulkower,V., Page,M., Ropers,D., Geiselmann,J. and de Jong,H. (2015) Robust reconstruction of gene expression profiles from reporter gene data using linear inversion. *Bioinformatics*, **31**, 71–79. 10.1093/bioinformatics/btv246.

30. Beck,J.V. and Woodbury,K.A. (1998) Inverse problems and parameter estimation: integration of measurements and analysis. *Meas. Sci. Technol.*, **839**, 839–847. 10.1088/0957-0233/9/6/001.

31. Gardner,T.S., Cantor,C.R. and Collins,J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, **403**, 339–342. 10.1038/35002131.

32. Elowitz,M.B. and Leibler,S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338. 10.1038/35002125.

33. Naylor,L.H. (1999) Reporter gene technology: the future looks bright. *Biochem. Pharmacol.*, **58**, 749–757. 10.1016/S0006-2952(99)00096-9.

34. Andersen,J.B., Sternberg,C., Poulsen,L.K., Bjørn,S.P., Givskov,M. and Molin,S. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.*, **64**, 2240–2246. 10.1128/AEM.64.6.2240-2246.1998.

35. Branch,M.A., Coleman,T.F. and Li.,Y. (1999) A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM J. Sci. Comput.*, **21**, 1–23. 10.1137/S1064827595289108.

36. IGEM. (2006) Registry of Standard Biological Parts.

37. Gompertz,B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. R. Soc.*, **115**, 513–583. 10.1098/rstl.1825.0026.

38. Silva,M.A.M., Matute,T., Nuñez,I., Valdes,A., Ruiz,C.A., Peña,G.A.V., Federici,F. and Rudge,T.J. (2019) Phase space characterization for gene circuit design. *BioRxiv*, 590299. 10.1101/590299.

39. Durfee,T., Nelson,R., Baldwin,S., Plunkett III,G., Burland,V., Mau,B., Petrosino,J.F., Qin,X., Muzny,D.M. and Ayele,M. *et al.* (2008) The complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse. *J. Bacteriol.*, **2597**, 2597–2606. 10.1128/JB.01695-07.

40. Toni,T., Welch,D., Strelkowa,N., Ipsen,A. and Stumpf,M.P.H. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface*, **6**, 187–202. 10.1098/rsif.2008.0172.

41. Chan,T.E., Stumpf,M.P.H. and Babtie,A.C. (2017) Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.*, **251**, 251–.e3. 10.1016/j.cels.2017.08.014.

42. Bansal,M., Gatta,G.D. and Bernardo,D.D. (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, **22**, 815–822. 10.1093/bioinformatics/btl003.

43. Gaal,T., Ross,W., Estrem,S.T., Nguyen,L.H., Burgess,R.R. and Gourse,R.L. (2001) Promoter recognition and discrimination by eσs RNA polymerase. *Mol. Microbiol.*, **42**, 939–954. 10.1046/j.1365-2958.2001.02703.x.

44. Mishra,D., Rivera,P.M., Lin,A., Vecchio,D.D. and Weiss,R. (2014) A load driver device for engineering modularity in biological networks. *Nat. Biotechnol.*, **32**, 1268. 10.1038/nbt.3044.

45. Lou,C., Stanton,B., Chen,Y.-J., Munsky,B. and Voigt,C.A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.*, **30**, 1137–1142. 10.1038/nbt.2401.

46. Storch,M., Casini,A., Mackrow,B., Fleming,T., Trewhitt,H., Ellis,T. and Baldwin,G.S. (2015) Basic: a new biopart assembly standard for idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth. Biol.*, **4**, 781–787. 10.1021/sb500356d.

47. Tas,H., Grozinger,L., Stoof,R., de Lorenzo,V. and Goñi-Moreno,A. (2021) Contextual dependencies expand the re-usability of genetic inverters. *Nat. Commun.*, **12**, 1–9. 10.1038/s41467-020-20656-5.

48. Shin,J., Zhang,S., Der,B.S., Nielsen,A.A.K. and Voigt,C.A. (2020) Programming *Escherichia coli* to function as a digital display. *Mol. Syst. Biol.*, **16**, e9401. 10.15252/msb.20199401.

49. Van Rossum,G. and Drake,F.L. *Python 3 Reference Manual.* CreateSpace, Scotts Valley, CA, pp. 2009.

50. Harris,C.R., Jarrod Millman,K., van der Walt,S.J., Gommers,R., Virtanen,P., Cournapeau,D., Wieser,E., Taylor,J., Berg,S., Smith,N.J. and Kern,R. et al. (2020) Array programming with NumPy. *Nature*, **585**, 357–362. 10.1038/s41586-020-2649-2.

51. Virtanen,P., Gommers,R., Oliphant,T.E., Haberland,M., Reddy,T., Cournapeau,D., Burovski,E., Peterson,P., Weckesser,W. and Bright,J. et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272. 10.1038/s41592-019-0686-2.

52. McKinney,W. (2010) Data structures for statistical computing in Python. In: *Proceedings of the 9th Python in Science Conference*, pp. 56–61. 10.5281/zenodo.3509134.

53. Hunter,J.D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95. 10.1109/MCSE.2007.55.

54. Plotly Technologies Inc. (2015) Collaborative data science.

55. Kluyver,T., Ragan-Kelley,B., Pérez,F., Granger,B.E., Bussonnier,M., Frederic,J., Kelley,K., Hamrick,J.B., Grout,J. and Corlay,S. *et al.* (2016)

Jupyter Notebooks-a Publishing Format for Reproducible Computational workflows. Vol. 2016. https://ebooks.iospress.nl/publication/42900.

56. Bisong,E. (2019) Google colaboratory. Apress, Berkeley, CA In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Springer, pp. 59–64. 10.1007/978-1-4842-4470-8_7.

57. Vidal,G., Vidal-Céspedes,C. and Rudge,T.J. (2022) LOICA: Integrating models with data for genetic network design automation. *ACS Synth. Biol.*, **11**, 1984–1990. 10.1021/acssynbio.1c00603.

58. Hansen,P.C. and O'Leary,D.P. (1993) The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, **14**, 1487–1503. 10.1137/0914086.