

RESEARCH

Open Access

# How reliable is the linear noise approximation of gene regulatory networks?

Philipp Thomas<sup>1,2</sup>, Hannes Matuschek<sup>3</sup>, Ramon Grima<sup>1\*</sup>

From IEEE International Conference on Bioinformatics and Biomedicine 2012  
Philadelphia, PA, USA. 4-7 October 2012

## Abstract

**Background:** The linear noise approximation (LNA) is commonly used to predict how noise is regulated and exploited at the cellular level. These predictions are exact for reaction networks composed exclusively of first order reactions or for networks involving bimolecular reactions and large numbers of molecules. It is however well known that gene regulation involves bimolecular interactions with molecule numbers as small as a single copy of a particular gene. It is therefore questionable how reliable are the LNA predictions for these systems.

**Results:** We implement in the software package intrinsic Noise Analyzer (iNA), a system size expansion based method which calculates the mean concentrations and the variances of the fluctuations to an order of accuracy higher than the LNA. We then use iNA to explore the parametric dependence of the Fano factors and of the coefficients of variation of the mRNA and protein fluctuations in models of genetic networks involving nonlinear protein degradation, post-transcriptional, post-translational and negative feedback regulation. We find that the LNA can significantly underestimate the amplitude and period of noise-induced oscillations in genetic oscillators. We also identify cases where the LNA predicts that noise levels can be optimized by tuning a bimolecular rate constant whereas our method shows that no such regulation is possible. All our results are confirmed by stochastic simulations.

**Conclusion:** The software iNA allows the investigation of parameter regimes where the LNA fares well and where it does not. We have shown that the parametric dependence of the coefficients of variation and Fano factors for common gene regulatory networks is better described by including terms of higher order than LNA in the system size expansion. This analysis is considerably faster than stochastic simulations due to the extensive ensemble averaging needed to obtain statistically meaningful results. Hence iNA is well suited for performing computationally efficient and quantitative studies of intrinsic noise in gene regulatory networks.

## Background

It is generally accepted that the relative size of molecular fluctuations scales as the inverse square root of the mean molecule numbers [1]. Since the key players of gene regulatory networks are present in amounts as small as one molecule it follows that gene expression is inherently noisy [2,3]. This molecular noise manifests itself in the copy number variations of transcripts and their proteins among genetically identical cells [4]. The main measures that have been used to quantify these

cell-to-cell variations both experimentally and through modeling are the coefficient of variation (CV) and the Fano factor [5-9].

Exact analytical results for these quantities have been derived only for very simple gene regulatory systems [10-12] and hence they are more commonly obtained by means of Monte Carlo simulations using the stochastic simulation algorithm (SSA) [13,14]. Despite being formally exact with the Chemical Master Equation (CME), in practice, this approach turns out to be computationally expensive mainly due to the considerable amount of sampling required to compute reliable statistical averages. The situation is exacerbated when networks

\* Correspondence: ramon.grima@ed.ac.uk

<sup>1</sup>School of Biological Sciences, University of Edinburgh, UK  
Full list of author information is available at the end of the article

are to be studied over a wide range of parameters. The main analytical tool to address this issue has since been the linear noise approximation (LNA) of the Chemical Master Equation (CME) [15-17] which allows one to approximate the dynamics of the latter by a set of linear stochastic differential equations from which all moments can be computed in closed form. In this approximation, the mean concentrations of the CME are approximated by the solution of the deterministic rate equations (REs) and the probability distribution of the fluctuations is approximated by a Gaussian. Thereby the LNA can give insight into the parametric dependence of the noise whenever the REs admit a unique steady state solution. However unlike the CME, this approximation is valid only in the limit of large molecule numbers and hence the accuracy of its predictions is questionable for intracellular biochemical reaction networks [18,19]. A handful of theoretical studies assess the accuracy of the REs and the LNA predictions by computing finite molecule number corrections to both approximations [20-22], a task which can be carried out analytically only for some simple systems. Hence, to-date, it is unclear how important these corrections are for many gene regulatory networks of interest.

We recently developed intrinsic Noise Analyzer (iNA) [23], the first software package enabling a fluctuation analysis for a broad class of biochemical networks of interest via the LNA and the Effective Mesoscopic Rate Equation (EMRE) approximations of the CME. The latter approximation gives accurate mean concentrations for systems characterized by intermediate to large molecule numbers and is hence more accurate than the conventional REs.

In this article we develop and efficiently implement in iNA, the Inverse Omega Square (IOS) method which gives the variances and covariances of fluctuations about the means calculated by the EMRE method. From these we can calculate the CVs and Fano factors of mRNA and protein fluctuations to an accuracy higher than possible with the LNA. Hence the software iNA provides a means of probing the validity of the LNA for any biochemical network under study. We use the EMRE and IOS methods to study the parametric dependence of the CV and Fano factors of mRNA and protein fluctuations in two examples of stochastic gene regulation involving nonlinear protein degradation, post-transcriptional, post-translational and negative feedback regulation. We show that these results agree with stochastic simulations but in many instances disagree with the LNA results. In particular the LNA predicts that the noise levels can be optimized by tuning a bimolecular rate constant whereas no such regulation is predicted by EMRE/IOS and simulations. It is also found that the LNA significantly underestimates the amplitude and period of noise-induced

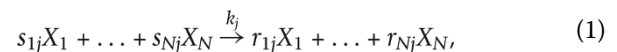
oscillations in genetic oscillators. Using detailed benchmarks we demonstrate that the present methodology is typically computationally more efficient than stochastic simulations using the SSA.

## Results

In this section we describe the results of the novel IOS method implemented in iNA. Its predictions are compared to the RE and LNA approximations of the CME and with exact stochastic simulations using the SSA for two examples of gene regulation. Finally we discuss its computational efficiency. The three methods (LNA, EMRE, IOS) are obtained from the system size expansion of the CME [15] which is applicable to monostable chemical systems. Technical details of the various approximation methods are provided in the section Methods.

### Investigating the parametric dependence of the size of molecular fluctuations

Biochemical reactions occur in random order and at random time. The stochastic description of biochemical reaction kinetics considers  $N$  distinct chemical species confined in a volume  $\Omega$  reacting via  $R$  chemical reactions of the form



where  $j$  varies from 1 to  $R$ . The mesoscopic state of the system is given by the vector of molecular populations  $\vec{n} = (n_1, n_2, \dots, n_N)$  and can be characterized by the probability  $P(\vec{n})$  to find the system in a particular configuration  $\vec{n}$ . The latter is however very difficult to obtain from analysis and hence intrinsic noise may be more easily characterized in terms of CVs and Fano factors which are defined in the following. We show how these quantities are calculated using the LNA and higher order approximations implemented in the software iNA.

The CV of the number of molecules of species  $X_i$  is defined by

$$C_{V,i} = \frac{\sqrt{\text{Var}(n_i)}}{E(n_i)}, \quad (2)$$

where  $\text{Var}(n_i)$  denotes the variance and  $E(n_i)$  the expectation value of the number of molecules of species  $X_i$ . The CV quantifies the relative spread about the mean or put in different terms, it measures the inverse signal-to-noise ratio. Because of the latter fact it is often referred to as the "size" of the noise. A different but commonly used noise measure is the Fano factor defined by

$$F_i = \frac{\text{Var}(n_i)}{E(n_i)}. \quad (3)$$

The Fano factor allows one to compare the spread of probability distributions relative to a Poissonian with the same mean. Notice that both quantifiers are determined solely by knowledge of the means and variances which can be obtained as a power series in the inverse compartment volume  $\Omega$  by van Kampen's system size expansion of the master equation [15]. This expansion is carried out at constant concentration and hence the large volume limit is equivalent to the limit of large molecule numbers. As shown in the section Methods, the expansion's leading order term ( $\Omega^0$ ) for the mean concentrations is given by the REs and the next to leading order term ( $\Omega^{-1}$ ) by the EMREs. Similarly the variances about these concentrations are given to leading order ( $\Omega^{-1}$ ) by the LNA and the next to leading order term ( $\Omega^{-2}$ ) by what we here refer to as the IOS approximation. It then follows that the CV has the expansion

$$C_{V,i} = C_{V,i}^{\text{LNA}}(\vec{k}, \Omega) \left( 1 + \Omega^{-1} c_i(\vec{k}) + O(\Omega^{-2}) \right), \quad (4)$$

and a similar expansion holds for the Fano factor

$$F_i = F_i^{\text{LNA}}(\vec{k}) \left( 1 + \Omega^{-1} f_i(\vec{k}) + O(\Omega^{-2}) \right). \quad (5)$$

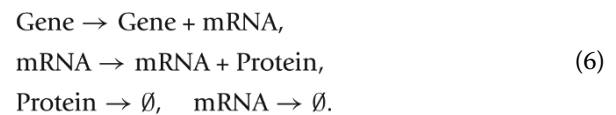
The above expressions fully characterize the parametric dependence of the size of the fluctuations in terms of the compartment volume  $\Omega$ , the set of reaction rate constants  $\vec{k} = (k_1, k_2, \dots, k_R)$  and the set of initial conditions whose explicit dependence has been omitted here. Note that the leading order contributions in the infinite volume limit,  $C_{V,i}^{\text{LNA}}(\Omega, \vec{k})$  and  $F_i^{\text{LNA}}(\vec{k})$ , are given by the LNA's result for the CV and Fano factor which can be shown to scale as  $\Omega^{-1/2}$  and  $\Omega^0$ , respectively. The factors  $c_i(\vec{k})$  and  $f_i(\vec{k})$  determine the relative corrections to the LNA result and can be obtained from the EMRE and IOS approximations as has been carried out explicitly in the Methods section. It can be argued that the size of these correction terms is proportional to the bimolecular reaction rate constants since the LNA is exact up to second moments for networks composed solely of unimolecular reactions since the propensities are linear functions of the concentrations (see section Methods). Summarizing, this analysis suggests novel correction terms to the CVs and Fano factors that are of order  $\Omega^{-3/2}$  and  $\Omega^{-1}$ , respectively, and hence of higher accuracy than the LNA.

Complementary to the LNA and IOS analysis the noise coefficients can be obtained by stochastic simulations using the SSA. Although this method is formally exact we note that noise estimators may be strongly biased for small to intermediate sample sizes [24]. The large amount of ensemble averaging required makes it computationally expensive to obtain these estimates from stochastic

simulations. A commonly used method to accelerate the statistical averaging procedure involves replacing the ensemble average by a time average which is allowed under steady state conditions and the assumption of ergodicity of sample paths [25]. In particular, using the SSA one time-averages over a sufficiently long time series to estimate the noise coefficients. The present version of iNA facilitates this procedure by computing the stationary moments via a time-averaged SSA. A convenient on-the-fly dialog allows one to remove transients from simulated trajectories and to monitor the convergence of mean, variance, CV or Fano factor statistics. In Figure 1 we present a screenshot of this dialog.

### Applications

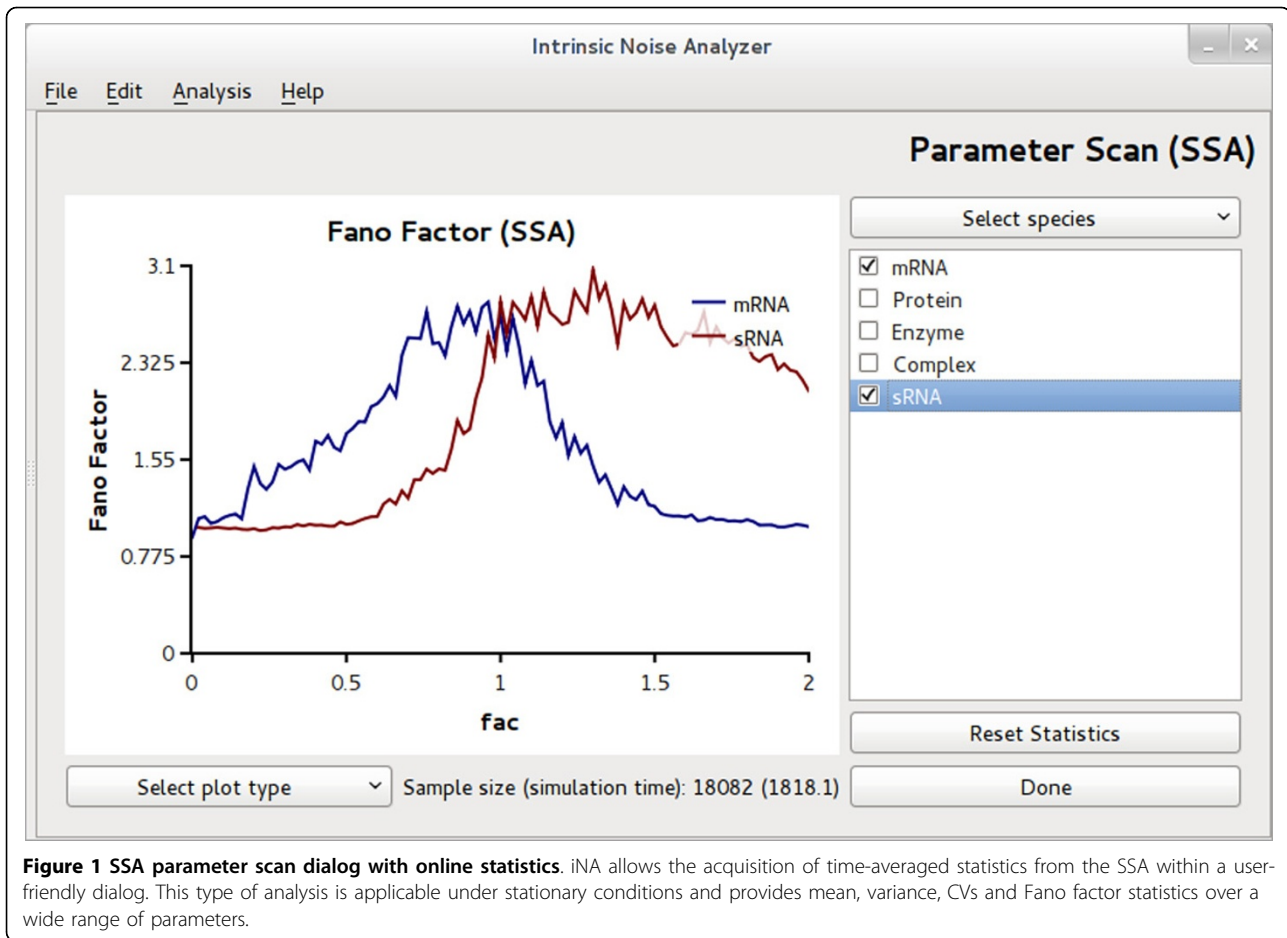
While gene expression is a complex process the most commonly used model is naturally the simplest. The model describes the transcription of mRNA and the translation of proteins from mRNA and the subsequent degradation by the effective first order reactions:



The model has been used to quantify variability in the proteome of *E. coli* [3,5,10], yeast [26] and mammalian cells [27] as well as having being subject to a number of theoretical studies [6,11]. The LNA's predictions of the first two moments of this model are exact since it is composed of only unimolecular reactions.

Given the complexity of the intracellular biochemistry, it is clear that this simple model cannot fully account for regulation which occurs at transcriptional, post-transcriptional, translational and post-translational stages. These processes typically involve bimolecular reactions with regulatory molecules such as transcription factors, functional RNAs or enzymes. While it is obvious that the CVs and Fano factors of more realistic models will differ from those predicted by the "standard" linear model (6), it is however not immediately clear whether these noise measures are qualitatively different than those obtained from the LNA.

In this section we demonstrate the use of the software iNA, which makes use of the approximation methods described in the previous section, to predict the noise characteristics of two gene regulatory networks involving post-transcriptional regulation by non-coding RNA and negative autoregulation via post-translational modification. Specifically we focus on how well these characteristics are described by the LNA both quantitatively and qualitatively and point out the LNA's limitations using correction terms of the IOS analysis and stochastic simulations provided by iNA.

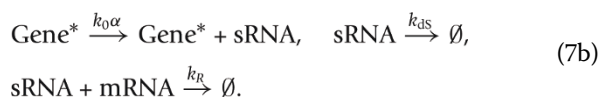
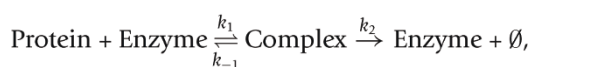
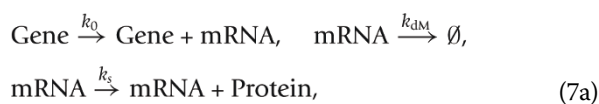


**Figure 1 SSA parameter scan dialog with online statistics.** iNA allows the acquisition of time-averaged statistics from the SSA within a user-friendly dialog. This type of analysis is applicable under stationary conditions and provides mean, variance, CVs and Fano factor statistics over a wide range of parameters.

### sRNA mediated post-transcriptional regulation

A large number of functional RNAs called small RNAs (sRNAs) have been found in bacteria which are not actively translated. This non-coding form of RNA is believed to coordinate pathways in response to external stimuli such as stress [28,29]. To investigate the robustness of critical pathways it is therefore important to understand the impact of intrinsic noise on their regulation.

Here we extend the mechanism of gene expression with nonlinear degradation studied in [30] to include the regulation of gene expression by sRNA in response to stress. A generic model of non-catalytic sRNA-mRNA interaction is



Note that reactions (7a) are as considered in Ref. [30]. In (7b) we describe the transcription and degradation of sRNA with respective rates  $k_{0\alpha}$  and  $k_{dS}$ . The parameter  $\alpha$  is given by the ratio of sRNA to mRNA transcription and can be used to describe the coordination of the stress response due to tight regulation of sRNA transcription. When sRNA is expressed it binds with its mRNA target at a rate  $k_R$  and quickly degrades thereafter. Similar models have been studied in Refs. [31,32].

Understanding how pathways are regulated in the presence of noise requires to study their response over a wide range of parameter values. Such a task is typically computationally expensive when carried out by stochastic simulations. We used iNA to investigate the impact of stress on our gene regulatory network both using the system size expansion and the SSA method. To our knowledge the effect of sRNA regulation on protein noise with a nonlinear degradation mechanism has not been studied before. Using parameter set (i) in Table 1 with iNA, we obtained the mean concentrations and standard deviations estimated from the REs and the LNA, respectively; these are shown in Figure 2 (a). Similarly the EMRE concentrations and IOS standard

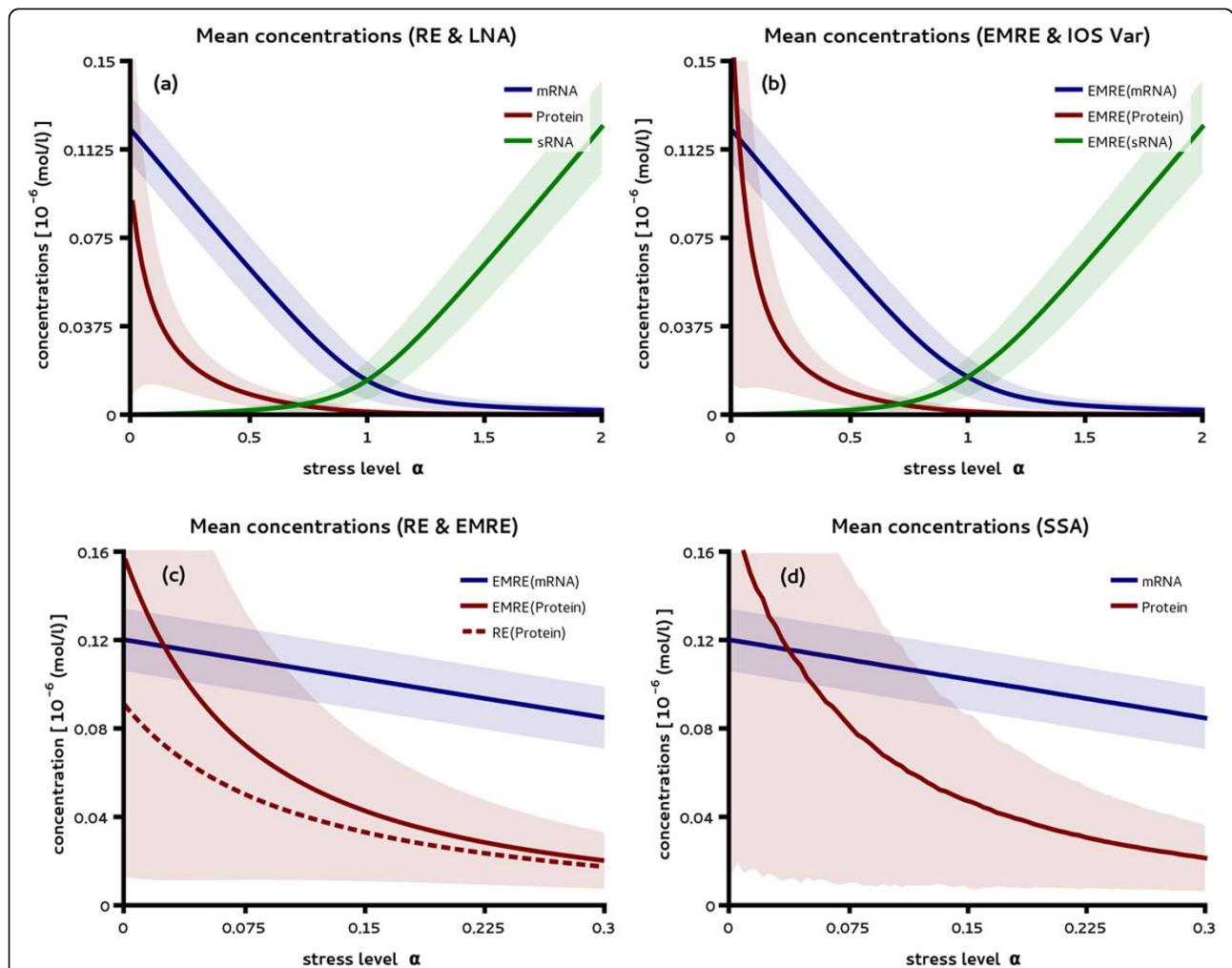


**Table 1 Gene expression model with sRNA regulation**

parameter	set(i)	set(ii)
$k_0[G]$	$0.024\text{min}^{-1}\mu\text{M}$	$0.0024\text{min}^{-1}\mu\text{M}$
$k_{dM}$	$0.2\text{min}^{-1}$	$0.2\text{min}^{-1}$
$k_s$	$1.5\text{min}^{-1}$	$1.5\text{min}^{-1}$
$k_{-1}, k_2$	$2\text{min}^{-1}$	$2\text{min}^{-1}$
$k_1$	$400(\mu\text{M min})^{-1}$	$4000(\mu\text{M min})^{-1}$
$k_{dS}$	$0.2\text{min}^{-1}$	$0.2\text{min}^{-1}$
$k_R$	$100(\mu\text{M min})^{-1}$	$1000(\mu\text{M min})^{-1}$

Kinetic parameters used for the gene expression model with sRNA regulation, scheme (7), as discussed in the main text. The volume is fixed to  $10^{15}$  with total enzyme concentration of  $0.1\mu\text{M}$  for parameter set (i) and  $0.01\mu\text{M}$  for parameter set (ii) corresponding to 60 and 6 enzyme molecules, respectively. Note that  $[G]$  denotes the constant gene concentration.

deviations for the same parameters are shown in (b). As stress levels are increased the characteristic threshold response is observed, i.e., as the expression level of sRNA rises it down-regulates the target mRNA concentrations [29]. Note that for  $\alpha < 1$  the protein is expressed while it is turned off for  $\alpha > 1$  while at the crossover point ( $\alpha = 1$ ) the levels of sRNA and mRNA are equal. Depending on the relative abundance of sRNAs and mRNAs in unstressed cells protein expression may be activated or silenced. A comparison of Figure 2 (a) and (b) shows that the predictions of the RE/LNA and of the EMRE/IOS agree well over large ranges of  $\alpha$ . The two differ for small  $\alpha$ , i.e., for small stress, (see Figure 2 (c)) where REs predict the mRNA levels to be higher than the protein ones while the EMRE method predicts



**Figure 2 Threshold responses of mRNA and protein concentrations under sRNA regulation.** We compare the mean concentrations predicted by the LNA method as a function of the parameter  $\alpha$  shown in (a) with those predicted by the more accurate EMREs shown in (b). The average expression levels predicted by both methods are in good agreement except for small  $\alpha$  (c) which is verified by stochastic simulations using the SSA (d).

the opposite. This phenomenon has been called discreteness-induced inversion [19] and has been discussed for the unregulated case in [30]. The effect is validated by stochastic simulations in Figure 2 (d). It is interesting that this inversion disappears for minute concentrations of sRNA corresponding to less than a single transcript and hence shows that the REs are surprisingly accurate over a wide range of stress levels.

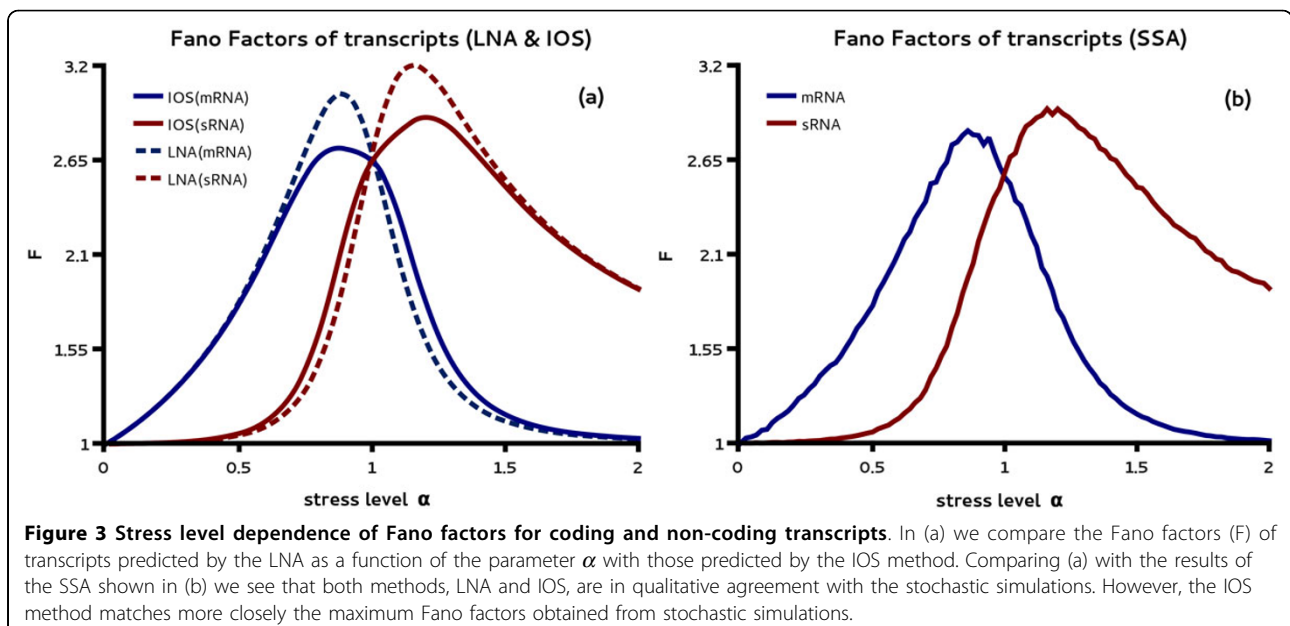
Next we study the dependence of the CVs and Fano factors of coding and non-coding transcripts on the stress level. The mean and variances shown in Figure 2 can be used to compute both of these quantities. Here the Fano factor is of particular interest since in the absence of sRNA control, the molecule number of mRNA is exactly Poissonian distributed with Fano factor one [11]. Hence the Fano factor can be used to study the impact of regulation on the noise. In Figure 3 (a) it is shown that the Fano factor of transcripts is increased shortly before and after the crossover point for mRNA and sRNA, respectively. This highlights an increase of the mRNA noise levels by sRNA regulation in comparison to the situation where the same average mRNA concentration is obtained by regulating its transcription rate. Interestingly we find that the LNA prediction for the Fano factors of transcripts near their peak values are larger than those from the IOS. This over-estimation by the LNA is confirmed by stochastic simulations shown in Figure 3 (b). Note that while the Fano factors of the transcripts have a peak for intermediate stress levels the dependence of the associated CVs is monotonous (Additional file 1).

In order to investigate the impact of stress on protein noise levels we analyzed the CVs of protein noise as predicted by the LNA and the IOS theory. The result is shown in Figure 4 (a). We find a minimum of the noise coefficient for intermediate levels of stress in the activated regime. Comparing the result to stochastic simulation also shown in Figure 4 (a) we see that both approximations yield the same qualitative result but the IOS coefficient is slightly more accurate in predicting the position and value of the minimum. Note that here the protein level corresponds to a copy number of about 60 molecules.

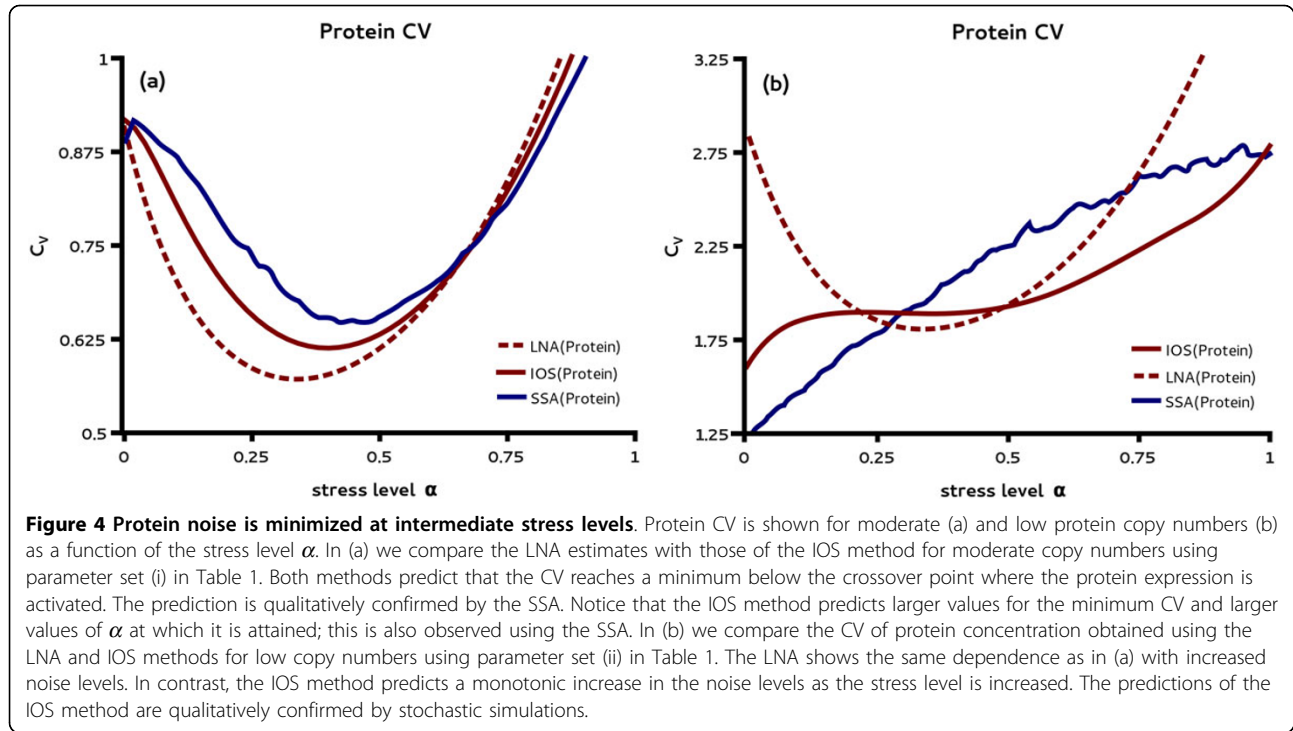
Genome-wide studies in *E. coli* revealed that some proteins can be expressed in much lower copy numbers than 60 [3]. We next make use of parameter set (ii) in Table 1 to probe the validity of the LNA under low copy number conditions. The results for the CV are shown in Figure 4 (b). In the absence of sRNA control the protein levels correspond to approximately 6 protein molecules. We observe that for such low copy numbers the LNA is in severe qualitative disagreement with the IOS. In particular, the LNA predicts that there exists a stress level for which the size of the noise is minimized while, in contrast, the IOS predicts the noise level to increase monotonically with stress. The latter is also reproduced by simulations using the SSA in Figure 4 (b) which hence signals the breakdown of the LNA under low copy number conditions.

#### Gene expression with negative autoregulation

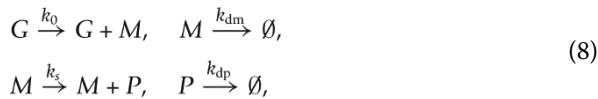
Autoregulation represents a common mechanism by which cells regulate their expression levels. Specifically



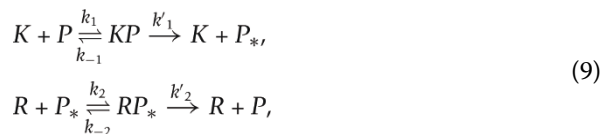
**Figure 3 Stress level dependence of Fano factors for coding and non-coding transcripts.** In (a) we compare the Fano factors (F) of transcripts predicted by the LNA as a function of the parameter  $\alpha$  with those predicted by the IOS method. Comparing (a) with the results of the SSA shown in (b) we see that both methods, LNA and IOS, are in qualitative agreement with the stochastic simulations. However, the IOS method matches more closely the maximum Fano factors obtained from stochastic simulations.



in *E. coli* about 35% - 43% of transcription factors are autoregulated [33,34] while in *S. cerevisiae* these account for about 10% [35]. We consider the common model of transcription and translation



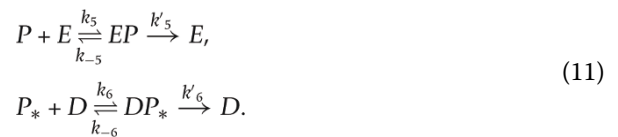
where  $G$ ,  $M$  and  $P$  refer to gene, mRNA and protein species, respectively. The first order degradation reactions here may also account for dilution due to cell growth. Next we add a negative feedback loop via a reversible phosphorylation mechanism and consequent transcriptional repression. Protein phosphorylation is common to post-translational regulation mechanisms [36-39] as for instance the negative feedback loop of the *Drosophila* circadian rhythm [40]. In fact, the majority of phosphorylated proteins in yeast are transcription factors [41]. The modification is modeled explicitly via a protein kinase  $K$  and a phosphatase  $R$  as follows



where  $P_*$  denotes the phosphorylated form of protein which can bind to the DNA and thereby inhibits its own expression.



Note that unlike the previous case of post-transcriptional regulation, here the promoter can be in one of three states  $G$ ,  $GP_*$  and  $GP_*^2$  depending on the number of bound protein molecules. The degradation as in the preceding example is assumed to occur via two proteases  $E$  and  $D$



A similar model has been analyzed using the LNA and the EMREs implemented in a previous version of iNA in Ref. [23]. With the present version of iNA the more accurate IOS analysis is available and is used here to investigate the reliability of the LNA estimates of the CVs. This presents a major benchmark for the LNA since the analysis includes fluctuations of a single promoter.

We start by exploring the dependence of mRNA and protein CVs by varying the transcription rate  $k_0$  of the promoter. Using the LNA for the parameter set given in Table 2 we calculated the CV as a function of the

**Table 2 Genetic expression model with negative autoregulation**

parameter	Value	parameter	value
$k_0$	$50\bar{\Omega}^{-1} (nM)$	$k_5$	$50h^{-1}$
$k_{dM}$	$5h^{-1}$	$k_{dP}$	$0.5h^{-1}$
$k_1$	$0.5 \cdot (nMh)^{-1}$	$k_1$	$1h^{-1}$
$k_2$	$0.5 \cdot (nMh)^{-1}$	$k_2$	$k_1$
$k_3$	$0.5\bar{\Omega}^{-1}$	$k_3$	$450h^{-1}$
$k_4$	$50 \cdot k_3$	$k_4$	$k_3$
$k_5$	$0.25 \cdot (nMh)^{-1}$	$k_5$	$0.5h^{-1}$
$k_6$	$5 \cdot (nMh)^{-1}$	$k_6$	$5h^{-1}$
$k'_6$	$10h^{-1}$	$k'_5$	$0.5h^{-1}$
$\bar{\Omega}$	$450 \cdot (nM)^{-1}$		

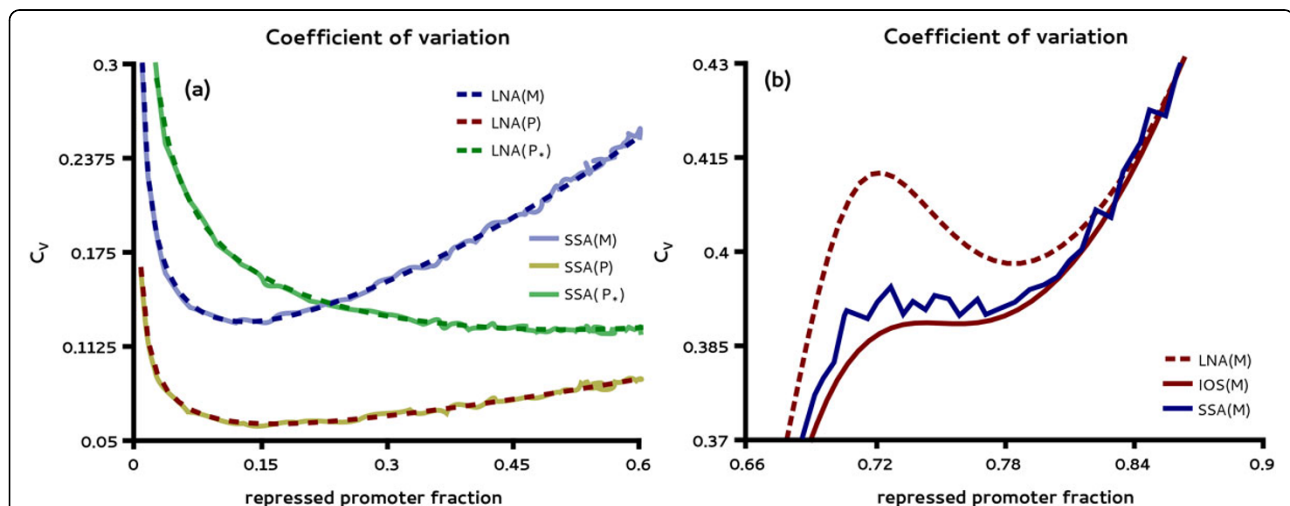
Kinetic parameters for the model of autoregulated gene model for a volume of  $7.5 \times 10^{-13}l$ . Note that here the parameter  $\bar{\Omega}$  is the inverse concentration of a single molecule. The total enzyme concentration of  $E, K, R$  is given by  $0.5nM$  while the concentration of  $D$  is  $5nM$ .

average fraction of repressed promoter states. The latter is given by the sum of the average occupation of the protein bound promoter states  $GP^*$  and  $GP^2_*$  and hence represents a measure of the feedback strength. We observe that both mRNA and the unphosphorylated protein CVs are minimized for small values of the feedback strength as shown in Figure 5 (a). These predictions are in excellent agreement with the results of the SSA. Next we investigate the parametric dependence of the fluctuations for larger values of the feedback strength varied through the rate constant  $k_3$  of the protein-DNA association rather than the transcription rate.

In Figure 5 (b) we show that the LNA predicts that the mRNA CV has a non-monotonic dependence on the average fraction of repressed promoter states. In contrast to the LNA, the IOS analysis predicts the CV to be a monotonically increasing function of the feedback strength showing no maximum. We have verified this dependence by stochastic simulations also shown in Figure 5 (b). In contrast to the mRNA CV, the same analysis carried out for the CVs of proteins yields qualitative agreement between LNA and SSA (Additional file 2).

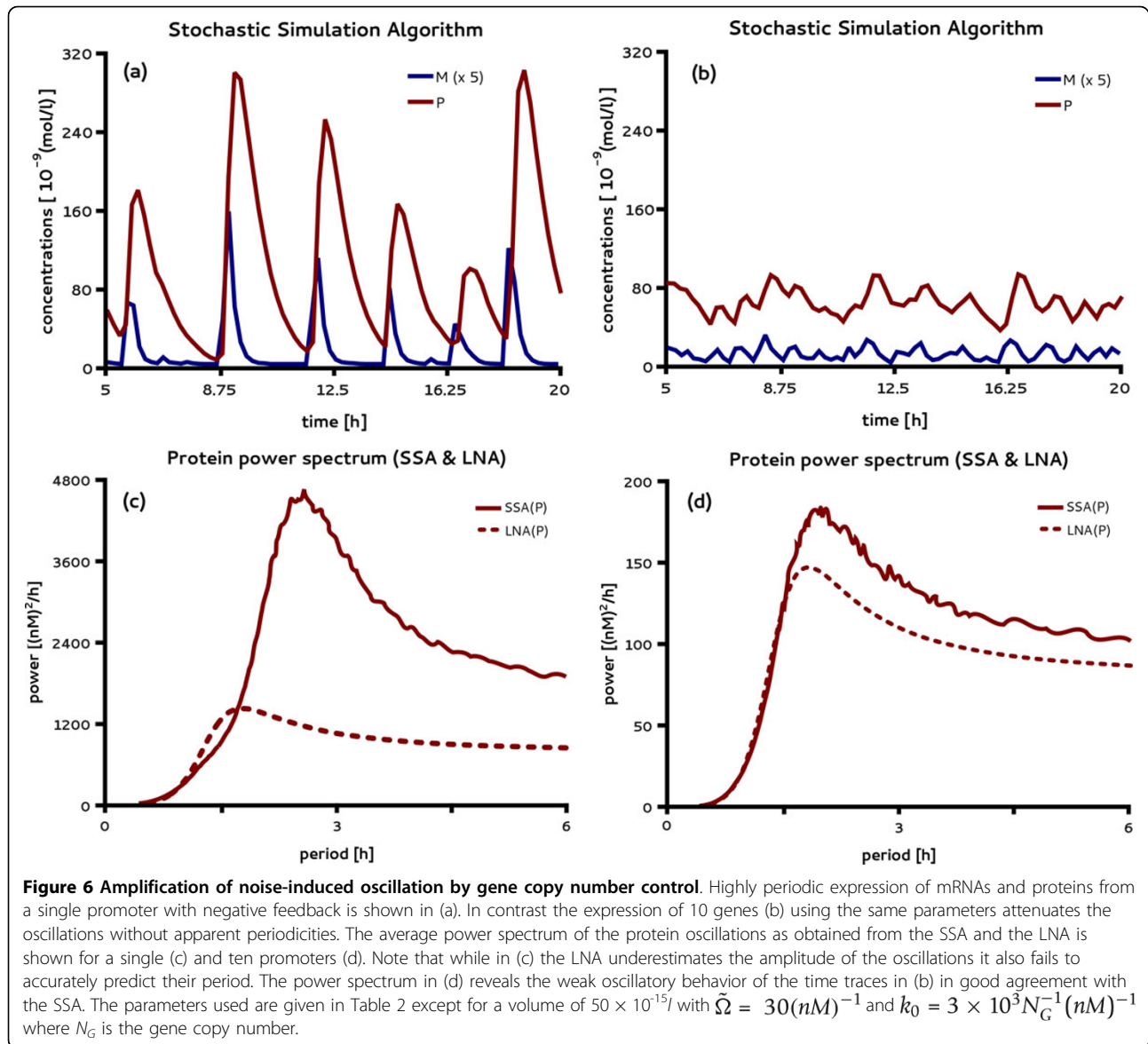
This result suggests that the corrections to the LNA are susceptible to the fluctuations in the promoter states. In order to test this hypothesis we investigated the oscillatory dynamics (that is often associated with the presence of a negative feedback loop) as a function of the gene copy number. In rapidly growing *E. coli*, for instance, the copy number of chromosomal genes located near the origin of DNA replication can be increased by 4-fold over genes located near the terminus [42]. Moreover genes located on plasmids can be present in higher copy numbers than those integrated in the genome. For synthetic circuits the plasmid copy number can also be controlled experimentally [43,44].

Variation of the copy number typically yields elevated protein concentrations if dosages are not compensated. For simplicity, here we scaled the transcription rate  $k_0$  by the number of genes which yields the same steady state expression levels for the deterministic REs independent of the gene copy number. Figure 6 (a) shows the oscillatory protein expression of a single gene



**Figure 5 CV of mRNA and protein fluctuations in gene autoregulation.** In (a) CVs of mRNA and protein levels are shown as function of the average repressed promoter fraction by variation of the transcription rate  $k_0$ . The result obtained using the LNA indicates that the CVs of mRNA and unphosphorylated protein are minimized when the repressed promoter fraction is small (10 - 15%) which is in good agreement with stochastic simulations. In (b) we compare the mRNA CV obtained using the LNA with the IOS as function of the average occupation of repressed promoters varied through the protein-DNA association rate  $k_3$  which is a measure of feedback strength. While the LNA predicts the CV to have a maximum for moderately repressed promoters the IOS analysis shows a monotonically increasing dependence on the feedback strength with significantly smaller noise levels. The latter is confirmed by stochastic simulations.





obtained from a SSA realization in a reduced volume of  $5 \times 10^{-14}l$  which roughly corresponds to that of yeast [45]. In Figure 6 (b) the same is shown for the expression of 10 genes. Notice that in contrast to the case of a single gene, the trajectories lack apparent periodicities. Hence we conclude that these oscillations are induced by limited gene copies. The signature of a noise-induced oscillation is a peak in the power spectrum of a system for which the deterministic REs show no sustained oscillations. In Additional file 3 we have verified that this is indeed the case. We next obtained the average power spectrum of the noise-induced oscillations from a large number of SSA realizations and compared it to the power spectrum that can be calculated from the LNA [46]. Figure 6 (c) and (d) show the power spectra of

protein expression of a single gene and of 10 genes, respectively. We note that in both cases the LNA qualitatively captures the presence of noise-induced oscillations since the LNA spectra exhibit a peak at a non-zero frequency. However for the case shown in Figure 6 (c) both the oscillation amplitude (the square root of the peak power) and period are underestimated by about 50% percent using the LNA. In actuality, the dampening of single cell oscillations has been observed in synthetic circuits with varying plasmid copy number with similar shifts in their periods [47].

### Implementation

iNA is a GUI-based software which at heart is based on the SBML description of stoichiometric reaction

networks. With the current release we introduce model manipulation capabilities which are tailored to fit the needs of stochastic modeling, as well as a just-in-time compilation engine that increases the overall execution speed of the analysis.

### Model editor capabilities

The software is based on the wide-spread SBML file format [48]. Although in common use, SBML has the shortcoming that it is barely human readable. The present version of iNA supports the compatible format SBML shorthand (SBML-sh) which represents the essential SBML model structure in an easy to read and write description language [49]. Therefore SBML-sh complements the existing SBML functionality by allowing import and export of both formats together with an online SBML-sh editor, see Additional file 4 (c).

Apart from this iNA's GUI also incorporates basic model editing capabilities. Additional file 4 (a) shows the list of reactions which allows to add or edit reactions within a dialog shown in Additional file 4 (b). Within this dialog the propensity of the reaction is either constructed automatically from the statistical formulation of the law of mass action [23,50,51] or to be specified by the user.

### Performance

The system size expansion of the CME yields a high dimensional system of coupled ODEs of order  $N^2$  equations for the LNA and  $N^3$  equations for the IOS analysis. Parameter scans as well as numerical integration of large systems are particularly challenging because of the large number of function evaluations needed to obtain accurate results. iNA's initial release addressed this issue by providing a bytecode interpreter for efficient expression evaluation [23]. The present version improves on this using a just-in-time (JIT) compiler provided by the LLVM infrastructure [30,52]. This technique provides the means of platform specific code generation for the system size expansion ODEs at runtime mimicking the performance of statically compiled code.

In order to access the performance of the present implementation we consider the model of negative auto-regulation studied in the section Applications which involves 14 species and 20 reactions. After conservation analysis the model reduces to only 9 species which yields a total number of 273 simultaneous equations for the IOS method and it is hence well suited for direct benchmarking purposes. This is particularly challenging in terms of ODE integration since the mean couples to higher statistical moments and causes the full system of 273 coupled equations to exhibit damped oscillations (see Additional file 3). The results of the benchmarks are summarized in Table 3 highlighting the performance

**Table 3 Performance of iNA's time course analysis**

method	IOS, LSODA	IOS, Rosenbrock	SSA, single (ens.)
BCI	18.2s (18.5s)	59.5s (59.8s)	0.04s (0.5h)
JIT	0.9s (3.6s)	13.0s (35.8s)	0.03s (0.4h)

Execution times of iNA's timecourse analysis for IOS and SSA for the gene oscillatory network reproducing Additional file 3 (a) and (b), respectively. The table compares the IOS method using the ODE integrator LSODA and Rosenbrock in combination with a bytecode interpreter or iNA's JIT feature. The value in the brackets denotes the overall time including bytecode assembly or JIT compilation. The SSA value denotes the average time for a single run while the values in the brackets denote the extrapolated value for an ensemble of 50, 000 independent realizations needed to generate Additional file 3 (b). Benchmarks were performed on Ubuntu (64bit), Dell OptiPlex 9010 SF, Core i7-3770 @ 3.4GHz using a single core.

of the present version of iNA. The improvements of iNA's system size expansion using the LSODA algorithm [53] over the previous Rosenbrock method reduce the execution time by up to a factor of 10 using the JIT compiler. In comparison the overall execution of the SSA requires about half an hour and hence is computationally extremely expensive because of the considerable number of trajectories to be averaged in order to obtain accurate statistics.

The analysis using the system size expansion is particularly advantageous when it is performed under steady state conditions because it can be readily obtained for large sets of parameters as we have shown in the section Applications. In this case the problem reduces to finding the solution of the 9 nonlinear deterministic REs and solving the remaining 264 linear equations (obtained from the system size expansion) from which the noise statistics are obtained. In Table 4 we summarize the detailed computation times for the REs, LNA and IOS analysis that have been employed to calculate the protein CVs (see Additional file 2 and Figure 5 (b)) showing the protein CVs. All analyses were performed in less than a second albeit the LNA is typically much quicker than the more accurate IOS method. For comparison we also show the average execution time per sample of the SSA using a finite sampling rate. We remark that

**Table 4 Performance of iNA's steady state parameter scan**

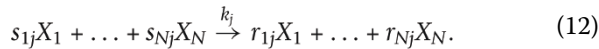
method	REs	LNA	IOS	SSA per sample (overall)
GiNaC	0.61s	1.48s	39.93s	–
BCI	0.19s	0.19s	0.39s	11ms (16h)
JIT	0.17s	0.17s	0.28s	7ms (10h)

Execution times of iNA's parameter scan for the gene autoregulation network reproducing Additional file 2 (a) and (b), respectively. The table compares the REs, LNA, IOS and SSA method using the bytecode interpreter or iNA's JIT feature. For comparison we also show the execution time using the computer algebra system GiNaC which is limited by its arbitrary precision arithmetic. The SSA value denotes the average time per sample per parameter while the values in the brackets denote the extrapolated value for a sample size of 50, 000 and 100 sets of parameter values as has been used in Additional file 2. Benchmarks were performed on Ubuntu (64bit), Dell OptiPlex 9010 SF, Core i7-3770 @ 3.4GHz using a single core.

the performance of the system size expansion methods could be increased by about 30% for the IOS and 40% for the SSA using iNA's integrated JIT compilation feature. We emphasize that the computation of the IOS analysis requires the same time as computing only 40 SSA samples. In particular to reproduce Additional file 2 and Figure 5 (b) a thousand-fold of this sample size was needed. The advantage of the IOS method as a complementary method to traditional means of stochastic simulation is therefore readily obvious.

## Methods

We consider a reaction network confined in a volume  $\Omega$  composed of  $N$  distinct chemical species reacting via  $R$  chemical reactions of the form



Here  $j$  is the reaction index running from 1 to  $R$ ,  $X_i$  denotes chemical species  $i$ ,  $k_j$  is the reaction rate constant of the  $j^{th}$  reaction and  $s_{ij}$  and  $r_{ij}$  are the stoichiometric coefficients. Let  $n_i$  be the number of molecules of the  $i^{th}$  species. Under well-mixed conditions, the time evolution of the mesoscopic state  $\vec{n} = (n_1, \dots, n_N)^T$  can be obtained either by Monte Carlo simulations using the SSA [14] or directly by determining the probability  $P(\vec{n}, t)$  of finding the system in a particular mesoscopic state using the CME

$$\frac{\partial P(\vec{n}, t)}{\partial t} = \Omega \sum_{j=1}^R (\hat{f}_j(\vec{n} - \vec{\mu}_j, \Omega) P(\vec{n} - \vec{\mu}_j, t) - \hat{f}_j(\vec{n}, \Omega) P(\vec{n}, t)), \quad (13)$$

where  $\hat{f}_j(\vec{n}, \Omega)$  is the probability per unit time and unit volume for the  $j^{th}$  reaction to occur [14] and  $(\vec{\mu}_j)_i = r_{ij} - s_{ij}$  is the stoichiometry of species  $i$  in the  $j^{th}$  reaction. Note that both approaches are equivalent [51].

## Rate equations and the Linear Noise Approximation

The CME determines the probability of observing any combination of molecule numbers at any point in time. Hence for closed systems the state space grows exponentially with the number of species while for open systems it is generally infinite. It is this complexity which prevents one to obtain exact analytical solutions of the CME except in particular cases [11,12]. The most common approximation method is the LNA which has been derived by van Kampen through the system size expansion of the Master equation. In brief, the method separates the instantaneous concentration vector into a macroscopic part,  $[\vec{X}]$ , and the fluctuations  $\vec{\epsilon}$  about it:

$$\frac{\vec{n}}{\Omega} = [\vec{X}] + \Omega^{-1/2} \vec{\epsilon}. \quad (14)$$

The macroscopic part is obtained as the solution of the conventional REs

$$\frac{\partial}{\partial t} [\vec{X}] = \sum_j^R \vec{\mu}_j f_j([\vec{X}]). \quad (15)$$

The implicit assumption made by ansatz (14) is that in the infinite volume limit the instantaneous concentrations equal the solution of the REs. It can then be shown that the macroscopic rate function of the  $j^{th}$  reaction is obtained from the relation  $f_j([\vec{X}]) = \lim_{\Omega \rightarrow \infty} \hat{f}_j([\vec{X}], \Omega)$  [30]. Since the limit is taken at constant concentration, this implies the large molecule number limit as well.

The method now proceeds by using the ansatz (14) together with Eq. (13) in order to obtain an equation for  $\vec{\epsilon}$ . The result is an expansion of the CME in powers of the inverse square root of the volume which can be truncated. The first term in the expansion ( $\Omega^{1/2}$ ) yields the REs. The next term ( $\Omega^0$ ) is given by a linear Fokker-Planck equation called the linear noise approximation of CME. This approximation is in wide-spread use mainly because of the simplicity of the result: the LNA solution is given by a Gaussian distribution describing the fluctuations around the macroscopic concentrations predicted by the REs. Specifically, to the order of approximation this implies that the macroscopic equations determine the average concentrations. The covariance of fluctuations  $\sigma_{ij} = \langle \epsilon_i \epsilon_j \rangle$  then satisfies the following matrix equation [16]

$$\mathbb{J} \underline{\sigma} + \underline{\sigma} \mathbb{J}^T + \underline{D} = 0, \quad (16)$$

where  $\mathbb{J} = \sum_{j=1}^R \vec{\mu}_j \nabla_{[\vec{X}]}^T f_j([\vec{X}])$  is the Jacobian of the macroscopic equations (15) and  $\underline{D} = \sum_{j=1}^R \vec{\mu}_j \vec{\mu}_j^T f_j([\vec{X}])$  is the diffusion matrix. Using Eq. (14) we can write expressions for the CVs and Fano factor of the  $i^{th}$  species using the definitions (2) and (3) in the main text

$$C_{V,i}^{\text{LNA}} = \Omega^{-1/2} \frac{\sqrt{\sigma_{ii}}}{[X_i]}, F_i^{\text{LNA}} = \frac{\sigma_{ii}}{[X_i]}, \quad (17)$$

which are of order  $\Omega^{-1/2}$  and  $\Omega^0$ , respectively.

It is well known that the means and variances of concentrations predicted by the LNA are exact only for reaction networks involving at most unimolecular reactions. For bimolecular reactions the LNA can be inaccurate if some species are present only in low molecule numbers. This is because for unimolecular reactions the hierarchy of moment equations obtained from the CME is closed, i.e., the  $n^{th}$  moment depends only the  $(n - 1)^{th}$  moment and all lower order moments [54]. The equations for the first moment are given by the REs since

the propensities are linear functions of the concentrations. The equations for the second moments are a system of linear equations for the variances and covariances which depends parametrically on the solution of the REs and are equivalent to the LNA result. For bimolecular reactions this is not the case since the hierarchy of moment equations obtained from the CME is not closed, i.e., the equations for the means involve the covariances and similarly the equations for the covariances involve higher moments. A systematic approximation of these equations can be achieved using the system size expansion which yields the REs and the LNA in the limit of large volumes [22]. The latter represent a closed system of equations for the first two moments.

### Finite molecule number corrections

Finite molecule number corrections to the LNA can be obtained by considering higher order terms in the system size expansion. The latter implies that the moments  $\langle \varepsilon_k \varepsilon_l \dots \varepsilon_m \rangle$  have an expansion of the form [21,55]:

$$\langle \varepsilon_k \varepsilon_l \dots \varepsilon_m \rangle = \sum_{j=0}^{\infty} [\varepsilon_k \varepsilon_l \dots \varepsilon_m]_j \Omega^{-j/2}. \quad (18)$$

For each term in the above expression an closed form equation can be derived [21,22]. In particular to leading order, the mean concentrations and the covariances are given by  $[\varepsilon_i]_0 = 0$  and  $[\varepsilon_i \varepsilon_j]_0 = \sigma_{ij}$  which is the LNA result. Note that for deterministic initial conditions  $[\varepsilon_i \varepsilon_j]_1 = 0$  as shown in [21] and hence the next to leading order corrections to these quantities are given by  $[\varepsilon_i]_1$  and  $[\varepsilon_i \varepsilon_j]_2$  which are generally non-zero if bimolecular reactions are considered. In order to relate the above moments back to the moments of the concentration variables we use Eqs. (14) and (18) to find expressions for the mean concentrations and covariance of fluctuations

$$\left\langle \frac{n_i}{\Omega} \right\rangle = [X_i] + \Omega^{-1} [\varepsilon_i]_1 + O(\Omega^{-2}), \quad (19a)$$

$$\begin{aligned} \Sigma_{ij} &= \left\langle \left( \frac{n_i}{\Omega} - \left\langle \frac{n_i}{\Omega} \right\rangle \right) \left( \frac{n_j}{\Omega} - \left\langle \frac{n_j}{\Omega} \right\rangle \right) \right\rangle \\ &= \Omega^{-1} [\varepsilon_i \varepsilon_j]_0 + \Omega^{-2} ([\varepsilon_i \varepsilon_j]_2 - [\varepsilon_i]_1 [\varepsilon_j]_1) + O(\Omega^{-3}) \end{aligned} \quad (19b)$$

Again, the leading order ( $\Omega^0$ ) contribution to the mean concentrations, Eq. (19a), is given by the macroscopic REs while the leading order contribution given by the  $\Omega^{-1}$  term in Eq. (19b) corresponds to the LNA estimate for the variance and the covariance. Including terms to order  $\Omega^{-1}$  in Eq. (19a) gives the EMRE estimate of the mean concentrations which corrects the solution of the REs [20]. Finally, considering also the  $\Omega^{-2}$  term in Eq. (19b) gives the IOS (Inverse Omega Squared) estimate of the variance and the covariance. From the form

of this higher order contribution it is clear that the variance estimate is centered around the EMRE concentrations and is of higher accuracy than the LNA method.

It now follows using definitions (2) and (3) in the main text that the CV and Fano factor have the following expansions in powers of the inverse volume

$$\begin{aligned} C_{V,i} &= \frac{\Omega \Sigma_{ii}^{1/2}}{\langle n_i \rangle} = C_{V,i}^{\text{LNA}} \left( 1 + \Omega^{-1} \left( \frac{[\varepsilon_i \varepsilon_i]_2 - [\varepsilon_i]_1^2}{\sigma_{ii}} - \frac{1}{2} \frac{[\varepsilon_i]_1}{[X_i]} \right) + O(\Omega^{-2}) \right), \\ F_i &= \frac{\Omega^2 \Sigma_{ii}}{\langle n_i \rangle} = F_i^{\text{LNA}} \left( 1 + \Omega^{-1} \left( \frac{[\varepsilon_i \varepsilon_i]_2 - [\varepsilon_i]_1^2}{\sigma_{ii}} - \frac{[\varepsilon_i]_1}{[X_i]} \right) + O(\Omega^{-2}) \right). \end{aligned} \quad (20)$$

Again, the leading order contributions are determined by the LNA result, Eq. (17). Note that the factors multiplying  $\Omega^{-1}$  in Eq. (20) yield the relative corrections to the LNA measures and are denoted by  $c_i$  and  $f_i$  in the main text. Note also that each of these factors contains a contribution stemming from a change in the variance of concentration fluctuations and another one reflecting the change in the mean of the concentrations. The equations determining the coefficients  $[\varepsilon_i]_1$  and  $[\varepsilon_i \varepsilon_j]_2$  needed to compute Eqs. (20) have been derived in Refs. [21] and [22]. As shown therein, the quantities depend only on  $[\vec{X}]$  and the vector of reaction rate constants  $\vec{k} = (k_1, k_2, \dots, k_R)$ . Hence the CVs and Fano factors depend parametrically on the reaction rate constants through the solution of the REs together with their initial conditions.

In Additional file 5 we have verified the correctness of iNA's implementation of the IOS for the example of a simple enzyme catalyzed reaction against an analytical result obtained in Ref. [22], Eq. (74) therein. We remark that using the IOS it is also possible to deduce the mean concentrations accurately to order  $\Omega^{-2}$  [21] which is superior to the EMRE and hence can be used as an error estimate of the method. However the variance about these concentrations is of  $\Omega^{-3}$  as can be seen from Eq. (19b) and hence requires to consider higher orders in the system size expansion.

### Discussion

In this article we have analyzed the parametric dependence of intrinsic noise in gene regulatory networks by means of average concentrations and variances as well as noise measures such as the CV and the Fano factor. The leading order contributions to the average concentrations and variances of the fluctuations as obtained from the system size expansion are given by the deterministic REs and the LNA respectively. The next to leading order contribution are given by novel terms which we have referred to as the EMRE and the IOS approximations respectively. The relative size of these corrections to the LNA are proportional to the inverse compartment volume and to the size of the bimolecular reaction rate constants. Hence, as we have demonstrated, these higher



order terms can be significant for networks involving low copy number of molecules and nonlinear reaction kinetics as is common in gene regulation.

In the case of sRNA regulation we have found that for highly expressed proteins the size of the noise can attain a minimum at intermediate stress levels. This result is in line with the LNA's prediction and may perhaps be advantageous as a mechanism of noise minimization in gene expression. The LNA result predicts that such optimization is indeed possible even for low protein expression levels yet the EMRE/IOS analysis and stochastic simulations show that this is not the case, i.e., the CV in the protein fluctuations increases monotonically with the stress levels.

For the case of gene autoregulation we observed that the LNA reliably describes the CV of mRNA and protein fluctuations when the transcription rate, a first order rate constant, is varied but gives very different results from simulations and the EMRE/IOS analysis when the protein-DNA association constant is varied (a bimolecular rate constant measuring the strength of the feedback loop). In particular in the latter case the LNA predicts a maximum in the CV is achieved as one increases the strength of the negative feedback loop whereas the EMRE/IOS analysis shows that the CV increases monotonically. We also found that the LNA can give considerably misleading results for the amplitude and period of noise-induced oscillations in the expression of a single gene while it becomes increasingly more accurate as the gene copy number is increased.

Hence in summary we have shown by means of these examples that the LNA's predictions regarding the regulation of noise by genetic regulatory networks can be quite different than those obtained by stochastic simulations using the SSA. In contrast the results from the EMRE/IOS methods agree well with those obtained from the SSA for the examples studied here. This is surprising since transcriptional feedback involves transitions between internal states of a gene represented by only one or two copies in a cell. We note that the methods presented can become inaccurate when the noise contribution of the feedback loop dominates. However, our methods enjoy the advantage that they can be computed in a fraction of the time needed to calculate the SSA. Hence the EMRE/IOS analysis tools implemented in iNA 0.4 present a quick means to accurately study the stochastic properties of biochemical reaction networks of intermediate or large size involving many bimolecular reactions.

## Availability

**Project name:** intrinsic Noise Analyzer

**Version:** 0.4.2

**Project home page:** <http://code.google.com/p/intrinsic-noise-analyzer>

**Operating systems:** platform independent, binaries available for Mac OSX, Linux and Windows

**Programming language:** C++

**License:** GNU GPL v2

## Additional material

### Additional file 1: Coefficients of variation of coding and non-coding transcripts as a function of stress levels.

CV of coding and non-coding transcripts in sRNA regulated gene expression as a function of the stress level  $\alpha$ . In (a) we see that the mRNA CV increases while the CV of sRNA decreases as the stress level increases. Notice that the IOS theory "linearizes" the LNA predictions around the crossover point. The predictions are well confirmed by stochastic simulations shown in (b).

### Additional file 2: Coefficients of variation of proteins in autoregulated gene expression as a function of feedback strength.

Protein CV of the autoregulated gene expression model is shown as a function of average fraction of repressed promoter states which are a measure of the feedback strength. Unlike the CV of mRNAs, under low copy number conditions the CV of unphosphorylated (a) and phosphorylated proteins (b) predicted by the LNA is in qualitative agreement with the IOS analysis. Notice that the IOS results more closely match those predicted by the SSA.

### Additional file 3: Amplification of damped oscillations in a single gene negative feedback loop.

Transient oscillations in the average mRNA and protein levels from negative feedback with a single gene copy number per cell. (a) compares the mean concentrations of the REs and EMREs. The latter predicts an amplification of the damped oscillations which is not captured by the REs. The result is in good agreement with the SSA shown in (b). The parameters used are given in Table 2 except for a volume of  $50 \times 10^{-15}l$ ,  $\bar{\Omega} = 30(nM)^{-1}$  and  $k_0 = 3 \times 10^3(nM)^{-1}$ .

### Additional file 4: Improved model editing capabilities in iNA 0.4.

iNA's GUI is equipped with a user friendly model editor. In (a) shows the list of reaction definition along with their propensities. The reaction editor shown in (b) facilitates the creation and editing of reactions through their chemical equation. The propensities are generated automatically using the law of mass action or can be specified explicitly. (c) In addition to the standard SBML format, iNA also supports the convenient SBML shorthand format which allows the specification of the essential model structure using a simple markup language.

**Additional file 5: Verification of iNA's implementation.** We have verified the soundness of our implementation by comparison with the analytical result using the IOS derived in Ref. [22]. The graph shows the ratio of the IOS and LNA variance (given by the contributions up to orders  $\Omega^{-1}$  and  $\Omega^{-2}$  of Eq. (19b), respectively) obtained from the system size expansion (SSE) against the fraction  $\delta$  of free enzyme per total enzyme concentration at steady state. This is also compared to the SSA where the ratio of SSA and LNA variance has been used.

## Competing interests

The authors declare that there are no competing interests.

## Authors' contributions

RG, HM and PT conceived the study, HM and PT implemented the software, PT performed the analysis, HM performed the benchmarks, PT and RG wrote the paper.

## Acknowledgements

Based on "Computation of biochemical pathway fluctuations beyond the linear noise approximation using iNA", by P. Thomas, H. Matuschek and R. Grima which appeared in *Bioinformatics and Biomedicine (BIBM)*, 2012 IEEE

International Conference on. ©2012 IEEE [30]. RG gratefully acknowledges support by SULSA (Scottish Universities Life Science Alliance).

#### Declarations

The publication costs for this article were funded by the corresponding author.

This article has been published as part of *BMC Genomics* Volume 14 Supplement S4, 2013: Selected articles from the IEEE International Conference on Bioinformatics and Biomedicine 2012: Genomics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/14/S4>.

#### Authors' details

<sup>1</sup>School of Biological Sciences, University of Edinburgh, UK. <sup>2</sup>School of Mathematics, University of Edinburgh, UK. <sup>3</sup>Institute of Physics and Astronomy, University of Potsdam, Germany.

Published: 1 October 2013

#### References

1. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N: **Noise in protein expression scales with natural protein abundance.** *Nat Genet* 2006, **38**:636-643.
2. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell.** *Sci Signal* 2002, **297**:1183.
3. Taniguchi Y, Choi P, Li G, Chen H, Babu M, Hearn J, Emili A, Xie X: **Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.** *Science* 2010, **329**:533-538.
4. Kaern M, Elston TC, Blake WJ, Collins JJ: **Stochasticity in gene expression: from theories to phenotypes.** *Nat Rev Genet* 2005, **6**:451-464.
5. Ozbudak E, Thattai M, Kurtser I, Grossman A, van Oudenaarden A: **Regulation of noise in the expression of a single gene.** *Nat Genet* 2002, **31**:69-73.
6. Paulsson J: **Models of stochastic gene expression.** *Phys Life Rev* 2005, **2**:157-175.
7. Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl U, Kerner MJ, Frishman D: **Protein abundance profiling of the Escherichia coli cytosol.** *BMC Genomics* 2008, **9**:102.
8. Hensel Z, Feng H, Han B, Hatem C, Wang J, Xiao J: **Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis.** *Nat Struct Mol Biol* 2012, **19**:797-802.
9. Voliotis M, Bowsher CG: **The magnitude and colour of noise in genetic negative feedback systems.** *Nucleic Acids Res* 2012, **40**:7084-7095.
10. Thattai M, Van Oudenaarden A: **Intrinsic noise in gene regulatory networks.** *Proc Natl Acad Sci* 2001, **98**:8614-8619.
11. Shahrezaei V, Swain P: **Analytical distributions for stochastic gene expression.** *Proc Natl Acad Sci* 2008, **105**:17256.
12. Grima R, Schmidt T, Newman T: **Steady-state fluctuations of a genetic feedback loop: An exact solution.** *J Chem Phys* 2012, **137**:035104.
13. McAdams H, Arkin A: **Stochastic mechanisms in gene expression.** *Proc Natl Acad Sci* 1997, **94**:814-819.
14. Gillespie D: **Stochastic simulation of chemical kinetics.** *Annu Rev Phys Chem* 2007, **58**:35-55.
15. van Kampen N: *Stochastic processes in physics and chemistry.* 3 edition. North-Holland; 2007.
16. Elf J, Ehrenberg M: **Fast evaluation of fluctuations in biochemical networks with the linear noise approximation.** *Genome Res* 2003, **13**:2475-2484.
17. Paulsson J: **Summing up the noise in gene networks.** *Nature* 2004, **427**:415-418.
18. Hayot F, Jayaprakash C: **The linear noise approximation for molecular fluctuations within cells.** *Phys Biol* 2004, **1**:205.
19. Ramaswamy R, González-Segredo N, Sbalzarini I, Grima R: **Discreteness-induced concentration inversion in mesoscopic chemical systems.** *Nat Commun* 2012, **3**:779.
20. Grima R: **An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions.** *J Chem Phys* 2010, **133**:035101.
21. Grima R, Thomas P, Straube A: **How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations?** *J Chem Phys* 2011, **135**:084103.
22. Grima R: **A study of the accuracy of moment-closure approximations for stochastic chemical kinetics.** *J Chem Phys* 2012, **136**:154105.
23. Thomas P, Matuschek H, Grima R: **Intrinsic Noise Analyzer: A Software Package for the Exploration of Stochastic Biochemical Kinetics Using the System Size Expansion.** *PLoS ONE* 2012, **7**:e38518.
24. Bao Y: **Notes and problems: Finite-sample moments of the coefficient of variation.** *Economet Theor* 2009, **25**:291-297.
25. Hamilton JD: *Time series analysis, Volume 2* Cambridge University Press; 1994.
26. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS: **Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise.** *Nature* 2006, **441**:840-846.
27. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M: **Global quantification of mammalian gene expression control.** *Nature* 2011, **473**:337-342.
28. Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H: **Regulation of gene expression by small non-coding RNAs: a quantitative view.** *Mol Syst Biol* 2007, **3**:138.
29. Levine E, Hwa T: **Small RNAs establish gene expression thresholds.** *Curr Opin Microbiol* 2008, **11**:574-579.
30. Thomas P, Matuschek H, Grima R: **Computation of biochemical pathway fluctuations beyond the linear noise approximation using iNA.** *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on: 4-7 October 2012* 2012, **1**-5.
31. Levine E, Zhang Z, Kuhlman T, Hwa T: **Quantitative characteristics of gene regulation by small RNA.** *PLoS Biol* 2007, **5**:e229.
32. Mehta P, Goyal S, Wingreen N: **A quantitative comparison of sRNA-based and protein-based gene regulation.** *Mol Syst Biol* 2008, **4**:221.
33. Pérez-Rueda E, Collado-Vides J: **The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12.** *Nucleic Acids Res* 2000, **28**:1838-1847.
34. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of Escherichia coli.** *Nat Genet* 2002, **31**:64-68.
35. Lee TI, Rinaldi NJ, Robert F, Odum DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al: **Transcriptional regulatory networks in Saccharomyces cerevisiae.** *Sci Signal* 2002, **298**:799.
36. Hunter T, Karin M, et al: **The regulation of transcription by phosphorylation.** *Cell* 1992, **70**:375.
37. Macek B, Mijakovic I, Olsen JV, Gnäd F, Kumar C, Jensen PR, Mann M: **The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis.** *Mol Cell Proteomics* 2007, **6**:697-707.
38. Macek B, Gnäd F, Soufi B, Kumar C, Olsen JV, Mijakovic I, Mann M: **Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation.** *Mol Cell Proteomics* 2008, **7**:299-307.
39. Wu YB, Dai J, Yang XL, Li SJ, Zhao SL, Sheng QH, Tang JS, Zheng GY, Li YX, Wu JR, et al: **Concurrent quantification of proteome and phosphoproteome to reveal system-wide association of protein phosphorylation and gene expression.** *Mol Cell Proteomics* 2009, **8**:2809-2826.
40. Hardin PE: **The circadian timekeeping system of Drosophila.** *Curr Biol* 2005, **15**:R714-R722.
41. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, Jona G, Breikreutz A, Sopko R, et al: **Global analysis of protein phosphorylation in yeast.** *Nature* 2005, **438**:679-684.
42. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ: **Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli.** *Cell* 2003, **113**:597-607.
43. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators.** *Nature* 2000, **403**:335-338.
44. Becskei A, Serrano L: **Engineering stability in gene networks by autoregulation.** *Nature* 2000, **405**:590-593.
45. Jorgensen P, Nishikawa JL, Breikreutz BJ, Tyers M: **Systematic identification of pathways that couple cell growth and division in yeast.** *Sci Signal* 2002, **297**:395.
46. McKane A, Nagy J, Newman T, Stefanini M: **Amplified biochemical oscillations in cellular systems.** *J Stat Phys* 2007, **128**:165-191.
47. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M: **A tunable synthetic mammalian oscillator.** *Nature* 2009, **457**:309-312.

48. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, *et al*: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.** *Bioinformatics* 2003, **19**:524-531.
49. Gillespie C, Wilkinson D, Proctor C, Shanley D, Boys R, Kirkwood T: **Tools for the SBML Community.** *Bioinformatics* 2006, **22**:628-629.
50. Bartholomay AF: **A stochastic approach to statistical kinetics with application to enzyme kinetics.** *Biochemistry* 1962, **1**:223-230.
51. Gillespie DT: **A rigorous derivation of the chemical master equation.** *Phys A* 1992, **188**:404-425.
52. Lattner C, Adve V: **LLVM: A compilation framework for lifelong program analysis & transformation.** *Code Generation and Optimization, 2004. CGO 2004. International Symposium on, IEEE 2004*, 75-86.
53. Petzold L: **Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations.** *SIAM J Sci Stat Comput* 1983, **4**:136-148.
54. Sotiropoulos V, Kaznessis YN: **Analytical derivation of moment equations in stochastic chemical kinetics.** *Chem Eng Sci* 2011, **66**:268-277.
55. Grima R: **Construction and accuracy of partial differential equation approximations to the chemical master equation.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2011, **84**:056109.

doi:10.1186/1471-2164-14-S4-S5

**Cite this article as:** Thomas *et al.*: How reliable is the linear noise approximation of gene regulatory networks? *BMC Genomics* 2013 **14** (Suppl 4):S5.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

