molecular
systems
biology

Dynamics of protein noise can distinguish between alternate sources of gene-expression variability

Abhyudai Singh, Brandon S. Razooky, Roy D. Dar, Leor S. Weinberger

Corresponding author: Leor S. Weinberger, The Gladstone Institute

Review timeline:

Submission date:	24 January 2012
Editorial Decision:	13 March 2012
Revision received:	08 June 2012
Editorial Decision:	06 July 2012
Revision received:	27 July 2012
Accepted:	30 July 2012

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision   13 March 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from two of the three referees who agreed to evaluate your manuscript, and have decided to render a decision now to avoid further delay.

As you will see from the reports below, the referees generally found the method proposed in this work potentially interesting. Both independently indicated, though, that additional experimental work would be needed to rigorously validate the ability of this method to distinguish transcription and translational noise, before it would be appropriate for publication in Molecular Systems Biology. Each reviewer makes specific recommendations for additional experiments that could potentially help address their concerns.

If you feel you can satisfactorily deal with the points listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favorable.

When preparing your revised work, the raw intensity data for Fig. 3, requested by Reviewer #1, can be provided as "figure source data". Please see our Instructions of Authors for more details on preparation and formatting of figure source data (<http://www.nature.com/msb/authors/index.html#a3.4.3>).

PLEASE NOTE As part of the EMBO Publications transparent editorial process initiative (see

http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology now publishes online a Review Process File with each accepted manuscript. Please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this file, which will be available to the scientific community. Authors may opt out of the transparent process at any stage prior to publication (contact us at msb@embo.org). More information about this initiative is available in our Instructions to Authors.

Sincerely,
Editor - Molecular Systems Biology
msb@embo.org

Referee reports:

Reviewer #1 (Remarks to the Author):

Starting with a minimal mathematical model of gene expression, the author derived a clever noise argument to distinguish between two (extreme) regimes of intrinsic noise in gene expression, which can be connected smoothly in the analysis by a single parameter (the burst size). While the steady state in protein expression would not allow distinguishing the two cases, the noise behavior (normalized CV) after a perturbation (transcription block) shows specific signatures of both regimes. The authors conclude from single cell analysis of four clones that the cells must be in the regime where the burst size is high, but they do not specify how high.

Overall this is an interesting paper that would benefit from some additions as outlined below.

Comments:

1.What the authors derive (Eq 3) is that the CV is a the minimum possible for a constitutive gene (that transcribes one transcript at time) times a positive number B_e that they call the effective burst size. I have two comments here: I am not sure that the splitting in the transcriptional and translational bursting is very useful here and it seems to some extend arbitrary. In the end the point is to make an estimate of B_e under the assumption of the model used. I do not see why B_e is called the 'effective' burst size as the formula for it is not very telling. Perhaps some more explanations would be useful (what distributions give $B_e = 1$?, etc.)

2. Fig. 3. Show the estimates for B_e from the data, even if that's only possible when assuming that ActD blocks transcription fully. This parameter is important as it allows comparison with other studies that have estimated burst sizes. It should be possible to find ActD conditions such that transcription rate goes almost to zero. I found the argument with the delta's artificial. In the two curves shown (Fig. 3), use the same values for delta1 and delta2, ideally 0 for delta1.

3. Analysis for Fig. 3. Show the raw intensity data as well. Provide the raw data as supplemental material.

4. To be fully convincing, the authors should show conditions in which the noise signatures go towards translational bursting. According to the model, this should be the case for reduced B_e (which the authors claim is achieved under partial transcription block). Can the authors run few additional experiments in which they use several different low doses of ActD to gradually reduce transcription before increasing the dose to block transcription? Then the different CV(t) curves should be staggered. This would ensure that higher CV can indeed be measured under the particular set up used.

Minor:
- Fig. 1: I find the parallel with the switching model misleading as the model used with the variable B does not explicitly model the switching.
- Supplement: Eq 1a, should there be some alphas?

- Fig. 3 label colors also with value of B_e instead of only % transcriptional bursting (cf. point 1 above)
- Supplement: Is it always the case that the moment equations are closed for the second moments or are the authors truncating some higher moments? Does this depend on the structure of the master equation?
- It may also be useful to graph all four clones on the same graph (Fig 3.) to compare values.

Reviewer #3 (Remarks to the Author):

Authors propose a new method for quantifying the source of intrinsic noise in protein fluctuations that aims at distinguishing between ``translational'' noise due to low copy numbers of mRNAs and ``transcriptional noise'' due to bursting in mRNA synthesis. I put these terms in quotes sine I am not completely satisfied with this nomenclature. In fact, I believe that their ``translational'' noise is also transcriptional in nature, just with presumably different statistical properties due to burstiness in mRNA synthesis. Indeed, authors in their model do not take into account translational fluctuations at all, and use deterministic mass-action equations, therefore the proposed method is not really quantifying translational noise.

It is an interesting study, since the direct measurements of mRNA dynamics are difficult. The proposed method generally makes sense in that indeed, if one assumes knowledge of the degradation rates for both proteins and mRNAs, one can in principle deduce the ``burstiness'' of the mRNA synthesis based on the temporal evolution of the relative noise strength (as characterized by the coefficient of variation). However, the applicability of the method is limited by a number of significant assumptions, some of which are mentioned in the manuscript (like the necessity to have comparable degradation rates of mRNAs and proteins), but sime are not. Among the latter the most important I believe are the implicit assumptions that degradation reactions of both mRNAs and proteins (i) are first-order, (ii) their rates are known, and (iii) the rates do not change after a transcription-blocking drug has been administered. In many practically important cases, these assumptions may not be satisfied, and so they have to be clearly stated and carefully evaluated.

Another assumption used in the formulating of the method is that the protein synthesis is completely deterministic. However, if the degradation rate of the protein is comparable to that of the mRNA (as for the destabilized GFP used in the experimental test), their abundances may be comparable as well, and then stochasticity of translation may also play an important role. It would be instructive to see how this additional source of noise affects the results.

One would definitely want to see an independent experimental validation of the method by using direct measurements of mRNA statistics. Such methods do exist (they are cited in the manuscript), therefore it would be of obvious interest to perform such a control assay.

In conclusion, I find the proposed method is new, interesting, and deserving publication in some form. However, the manuscript needs a significant revision before the decision about publication can be made.

1st Revision - authors' response   08 June 2012

Thank you for obtaining detailed and constructive reviews of the paper entitled "*Dynamics of protein noise can be used to distinguish transcriptional bursting from translational bursting*". In response to the reviewer's comments, we have performed new experiments and made significant revisions to the manuscript. Most notably, we present new single-molecule mRNA FISH data to directly measure mRNA population statistics across cell populations. The mRNA FISH data reveals transcriptional bursting from the HIV LTR promoter and provides an independent experimental verification of the proposed method. We have also addressed all other reviewer comments including the estimation of transcriptional burst size from data and incorporating noise at the translational

level into the model. Responses to specific reviewer's comments are detailed below. We request that the revised manuscript be considered for publication in *Molecular Systems Biology*.

Reviewer #1 (Remarks to the Author):

Starting with a minimal mathematical model of gene expression, the author derived a clever noise argument to distinguish between two (extreme) regimes of intrinsic noise in gene expression, which can be connected smoothly in the analysis by a single parameter (the burst size). While the steady state in protein expression would not allow distinguishing the two cases, the noise behavior (normalized CV) after a perturbation (transcription block) shows specific signatures of both regimes. The authors conclude from single cell analysis of four clones that the cells must be in the regime where the burst size is high, but they do not specify how high.

Overall this is an interesting paper that would benefit from some additions as outlined below.

Comments:

1.What the authors derive (Eq 3) is that the CV is a the minimum possible for a constitutive gene (that transcribes one transcript at time) times a positive number B_e that they call the effective burst size. I have two comments here: I am not sure that the splitting in the transcriptional and translational bursting is very useful here and it seems to some extend arbitrary. In the end the point is to make an estimate of B_e under the assumption of the model used. I do not see why B_e is called the 'effective' burst size as the formula for it is not very telling. Perhaps some more explanations would be useful (what distributions give B_e =1?, etc.)

The reviewer brings up two excellent points. First, we now understand that referring to B_e as the "effective" burst size was not helpful. Since B_e is also the steady-state Fano factor for the mRNA count (see Eq. 8 in Appendix A), we have now renamed B_e using the notation used for Fano factor in previous studies, η_m (see Kaern et al. Nature Reviews Genetics 2005) and refer to it as the "mRNA Fano factor" in place of the "effective burst size" throughout the text.

Second, our main motivation for splitting the protein noise was to quantify how much of the variability in protein levels could be attributed to low-copy mRNA fluctuations (translational bursting) vs. promoter-state fluctuations (transcriptional bursting). Previous studies have established that the contribution of translational bursting to protein variability is:

$$\frac{k_p}{(\gamma_m + \gamma_p)\overline{\langle p \rangle}}.$$

(see for example: Singh et al. Biophysical Journal 2010). Hence, we defined the noise contribution from transcriptional bursting as the total noise minus the contribution from translational bursting. Although splitting protein noise is not essential to the results presented in the paper, we feel it is useful in attributing stochastic variability to transcriptional vs. translational bursting.

2. Fig. 3. Show the estimates for B_e from the data, even if that's only possible when assuming that ActD blocks transcription fully. This parameter is important as it allows comparison with other studies that have estimated burst sizes. It should be possible to find ActD conditions such that transcription rate goes almost to zero. I found the argument with the delta's artificial. In the two curves shown (Fig. 3), use the same values for delta1 and delta2, ideally 0 for delta1.

As instructed by the reviewer, we have now estimated η_m (previously called B_e) from data keeping the same values for delta1 and delta2 for both the curves in Fig 3 (details in Appendix C). Our analysis shows that η_m is at least 10, i.e., the contribution of transcriptional bursting to GFP expression variability is at least 90%. The value of η_m obtained form our method is consistent with direct measurements of Fano factor from mRNA FISH data ($\eta_m \approx 80$; see Appendix G).

Unfortunately, for all non-cytotoxic concentrations of Actinomycin D (ActD) that we used,

transcription rate does not go to zero. Even at the highest concentration of ActD (10 μg/ml), mean GFP levels asymptotically approach 10% of their maximum value (See Fig. S1 in Appendix C).

An important limitation of the method, which we now discuss in the paper (see third paragraph on page 8), is that it may only allow an order of magnitude estimate of η_m. This is because the change in protein expression variability after a transcription block becomes insensitive to η_m, when η_m is large. For example, model predictions for η_m equal to one and ten are well separated in Fig. 2, but predictions for η_m equal to ten and eighty are close to each other. Thus, given the errors in the data it may not be possible to get precise estimates of η_m using this method, especially when η_m is large (i.e., high levels of transcriptional bursting form the promoter).

3. Analysis for Fig. 3. Show the raw intensity data as well. Provide the raw data as supplemental material.

As suggested, raw intensity data has been uploaded with the paper.

4. To be fully convincing, the authors should show conditions in which the noise signatures go towards translational bursting. According to the model, this should be the case for reduced B_e (which the authors claim is achieved under partial transcription block). Can the authors run few additional experiments in which they use several different low doses of ActD to gradually reduce transcription before increasing the dose to block transcription? Then the different CV (t) curves should be staggered. This would ensure that higher CV can indeed be measured under the particular set up used.

This is an excellent suggestion. We have now analytically quantified the change in η_m (previously called B_e) needed to create a significant shift in the CV(t) curve. mRNA FISH data has revealed high levels of transcriptional bursting form the HIV-1 LTR with $\eta_m \approx 80$ (Appendix G). Fig 2 shows that η_m will have to change from a value of 80 to 4 (20-fold change) for a significant shift in CV(t) towards the translational bursting line.

Our analysis has shown that at the highest concentration of Actinomycin D used, the reduction in η_m is only 2-fold (Appendix H). Given the insensitivity of CV(t) to η_m, a change from 80 to 40 would be indistinguishable in Fig. 2. Note that at lower doses of ActD (as proposed above), the change in η_m would be much smaller than 2-fold and will not be enough to shift the CV(t) curves.

Finally, measurements of GFP expression variability after blocking transcription with various doses of ActD show similar curves for CV(t) (Appendix I).

In summary, while we feel this was a clever experiment suggested by the reviewer and we attempted to get it to work, analytical predictions and experimental data show that it will not be possible to shift CV(t) curves towards translation bursting by using different concentrations of the transcription-blocking drug.

Minor:
- Fig. 1: I find the parallel with the switching model misleading as the model used with the variable B does not explicitly model the switching.

  As correctly pointed out by the reviewer, we do not explicitly model promoter switching. However, the stochastic model proposed in the manuscript is equivalent to the two-state promoter model in physiological relevant parameter regimes. For example, if the promoter spends most of the time in the 'ON' state then it corresponds to B=1 with probability one in the model. Alternatively, if the promoter spends most of the time in the 'OFF' state then it corresponds to B being a geometrically distributed random variable. We have modified the text in the paper emphasizing this connection between the proposed model and the two-state promoter model (see last paragraph on page 4).

- Supplement: Eq 1a, should there be some alphas?

  We thank the review for pointing this typo. Eq 1a has been modified in the supplemental.

- Fig. 3 label colors also with value of B_e instead of only % transcriptional bursting (cf. point 1 above)

  Values for η_m (previously called B_e) are provided in Figures 2 & 3.

- Supplement: Is it always the case that the moment equations are closed for the second moments or are the authors truncating some higher
  moments? Does this depend on the structure of the master equation?

  Since the probabilities defining mRNA time evolution (Eq.1 in the supplemental) and protein dynamics (Eq 2 in the paper) are linear with respect to m(t) and x(t), the moment dynamics are always closed. Thus, truncation of higher order moments is not needed to solve for the moment dynamics in this case.

  Nonlinearities in the systems (which can arise with feedback regulation) can make time derivative of lower order moments dependent on higher order moments. In such cases, moment-closure methods are typically needed to solve for the statistical moments.

- It may also be useful to graph all four clones on the same graph (Fig 3.) to compare values.

  A figure showing the response of all the clones on the same graph is now provided in the supplemental (see Fig. S7).

Reviewer #3 (Remarks to the Author):

Authors propose a new method for quantifying the source of intrinsic noise in protein fluctuations that aims at distinguishing between ``translational'' noise due to low copy numbers of mRNAs and ``transcriptional noise'' due to bursting in mRNA synthesis. I put these terms in quotes sine I am not completely satisfied with this nomenclature. In fact, I believe that their ``translational'' noise is also transcriptional in nature, just with presumably different statistical properties due to burstiness in mRNA synthesis. Indeed, authors in their model do not take into account translational fluctuations at all, and use deterministic mass-action equations, therefore the proposed method is not really quantifying translational noise.

We completely agree. Our analysis does not consider noise at the translational level and we have now attempted to clarify this in the main text. As discussed below, due to the high copy numbers of proteins, any noise originating at the translational level (through probabilistic protein synthesis and decay) is negligible compared to noise at the transcriptional level. Our method specifically discriminates between two different sources of transcriptional noise: (i) fluctuations in promoter transcriptional states (which we refer to as "transcriptional busting" noise) and (ii) low-copy fluctuations in mRNA counts (which we refer to as "translational bursting" noise). Importantly, "translational bursting noise" is stochasticity in the gene expression process that is completely originating at the transcriptional level. We realize that this nomenclature could have caused confusion and we now clarify it by discussing this early on in the manuscript (see end of page 3).

It is an interesting study, since the direct measurements of mRNA dynamics are difficult. The proposed method generally makes sense in that indeed, if one assumes knowledge of the degradation rates for both proteins and mRNAs, one can in principle deduce the ``burstiness'' of the mRNA synthesis based on the temporal evolution of the relative noise strength (as characterized by the coefficient of variation). However, the applicability of the method is limited by a number of significant assumptions, some of which are mentioned in the manuscript (like the necessity to have comparable degradation rates of mRNAs and proteins), but sime are not. Among the latter the most important I believe are the implicit assumptions that degradation reactions of both mRNAs and proteins (i) are first-order, (ii) their rates are known, and (iii) the rates do not change after a transcription-blocking drug has been administered. In many practically important cases, these

assumptions may not be satisfied, and so they have to be clearly stated and carefully evaluated.

We have modified the text in the paper (see page 7 end of second paragraph) explicitly stating these assumptions. The reviewer makes a very good point regarding our assumptions on the degradation reactions. As the reviewer requests, we now directly measure the rate of GFP degradation, show that it is first order, and show the degradation rate of GFP remains unchanged in the presence of the transcription-blocking drug (see Fig. S4 in supplemental).

Another assumption used in the formulating of the method is that the protein synthesis is completely deterministic. However, if the degradation rate of the protein is comparable to that of the mRNA (as for the destabilized GFP used in the experimental test), their abundances may be comparable as well, and then stochasticity of translation may also play an important role. It would be instructive to see how this additional source of noise affects the results.

This is an excellent point. To investigate how noise at the translational levels affects our results, we have considered an expanded model in Appendix E. In this model, both mRNA/protein dynamics are modeled through stochastic jumps. Analysis of the model reveals that as long as the mean protein count is atleast 50 molecules, noise at the translation level does not change the results presented in Fig 2 (see Supplemental Fig. S3). The clones considered in this paper have mean GFP population counts in the order of 10,000 (Singh et al. Biophysical Journal 2010) and hence modeling protein synthesis and decay using mass-action equations is a valid approximation in this system. We thank the reviewer for this suggestion.

One would definitely want to see an independent experimental validation of the method by using direct measurements of mRNA statistics. Such methods do exist (they are cited in the manuscript), therefore it would be of obvious interest to perform such a control assay.

As requested by the reviewer, we now perform single-molecule mRNA FISH analysis to directly measure mRNA population statistics across cell populations (details in Appendix G). The mRNA FISH data verifies extensive transcriptional bursting from the HIV LTR promoter and provides an independent experimental verification of the proposed method.

In conclusion, I find the proposed method is new, interesting, and deserving publication in some form. However, the manuscript needs a significant revision before the decision about publication can be made.

2nd Editorial Decision   06 July 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who agreed to evaluate your revised study. As you will see, the referees agreed that the revisions made to this work had addressed their main concerns, and they are mostly supportive. They raise however a series of minor concerns and conceptual issues, which we would ask you to carefully address in a final revision of the present work.

The editor feels that these issues, while not entirely trivial, should be addressable with some textual clarifications and some minor additional analysis. Given the overall supportive comments from the reviewers, the editor feels that this work will, in principle, be acceptable for publication after these last issues are addressed.

Reviewer #3 has some remaining conceptual concerns with the appropriateness of the "translational bursting" terminology. I can see the reviewer's point here, but I do feel that this issue is somewhat semantic and the other reviewers did not seem to share this concern. I think that you have done a reasonable job of clarifying your meaning in the introduction of this work, and, as such, I would prefer to leave it to your discretion to decide whether or not to further change the terminology used in this manuscript.

When preparing your revised manuscript, please also address the following format and content issues:

1. Please provide high-resolution versions of the manuscript figures, ideally in the EPS format (PDF or TIFF are also be acceptable).

2. Please add a Table of Contents to the beginning of the Supplementary Information pdf, listing the page numbers for each Appendix section and supplementary figure.

Thank you for submitting this paper to Molecular Systems Biology.

Sincerely,
Editor - Molecular Systems Biology
msb@embo.org

Referee reports:

Reviewer #1 (Remarks to the Author):

I am mostly satisfied with the revision of the manuscript. I have two remaining points:

Point 2. Since I was mislead in thinking that B_e was the effective burst size, my point was about reporting the values of <B>, i.e. the mean burst size or how many mRNA copies are made per burst, since this was estimated previously in mammalian cells. Can the authors estimate this quantity?

mRNA fish: I strongly suggest to show some images of mRNA fish together with the histogram in Fig. S5.

Reviewer #3 (Remarks to the Author):

In the revised version, authors significantly improved the presentation both by clarifying previously confusing statements and adding new data as requested by the reviewers. However, a couple of important issues still remain.

First, I found their response to the question of the other referee on the connection between their model and the stochastic two-state promoter switching model not convincing. I do not see why their model is equivalent to the promoter switching model in any parameter regime. If the authors insist on that point, they have to show it explicitly, otherwise they should remove this point. In fact, it is not essential to their presentation anyway.

Second, and more important, I am still is not convinced that they can use the term "translational bursting" in their context. I still believe that this terminology is misleading, and the small comment they now placed in the bottom of p. 3 did not remedy that. What their method in fact measures is the magnitude of transcriptional bursting, as quantified by the Fano factor of the mRNA distribution. Their measurements answer the question whether the transcriptional bursting is significant (large Fano factor) or not (Fano factor is close to one), it really has nothing to do with translational bursting.

I believe they should correct the presentation in these two respects (which would also entail changing the title accordingly), and then the paper could probably be published

Below we answer the reviewer's remaining concerns.   As requested, we have included a representative image of the mRNA FISH in Fig. S4.   We have also decided to change the terminology in the manuscript to avoid confusion and as such the manuscript has been re-titled: "Dynamics of Protein Noise can Distinguish Between Alternate Sources of Gene-Expression Variability".

Below we provide specific responses to each reviewer comment.

Please do not hesitate to contact us with any questions.

Reviewer #1 (Remarks to the Author):

I am mostly satisfied with the revision of the manuscript. I have two remaining points:

Point 2. Since I was mislead in thinking that B_e was the effective burst size, my point was about reporting the values of , i.e. the mean burst size or how many mRNA copies are made per burst, since this was estimated previously in mammalian cells. Can the authors estimate this quantity?

Yes, the mean transcriptional burst size $\langle B \rangle$ can be estimated for the HIV LTR. We have modified the text explaining the relationship between the Fano factor η_m and $\langle B \rangle$ (see below Eq. 5 in the paper). Analysis shows that in a two-state promoter model, where the ON state is unstable and the promoter spends most of the time in the OFF state, $\eta_m = 1 + \langle B \rangle$ (see Appendix B). Thus, estimates of η_m are representative of the mean burst size for many promoters including HIV LTR, which have been shown to mostly reside in the OFF state (Singh et al, 2010; Raj et al, 2006; Suter et al, 2011).

mRNA fish: I strongly suggest to show some images of mRNA fish together with the histogram in Fig. S5.

Figure has been modified to include images of mRNA FISH (see Fig. S4).

Reviewer #3 (Remarks to the Author):

In the revised version, authors significantly improved the presentation both by clarifying previously confusing statements and adding new data as requested by the reviewers. However, a couple of important issues still remain.

First, I found their response to the question of the other referee on the connection between their model and the stochastic two-state promoter switching model not convincing. I do not see why their model is equivalent to the promoter switching model in any parameter regime. If the authors insist on that point, they have to show it explicitly, otherwise they should remove this point. In fact, it is not essential to their presentation anyway.

Since the equivalence between the two-state promoter model and the model presented in the manuscript is not essential to the paper, we have decided to remove this point, as suggested by the reviewer.

Second, and more important, I am still is not convinced that they can use the term "translational bursting" in their context. I still believe that this terminology is misleading, and the small comment they now placed in the bottom of p. 3 did not remedy that. What their method in fact measures is the magnitude of transcriptional bursting, as quantified by the Fano factor of the mRNA distribution.

Their measurements answer the question whether the transcriptional bursting is significant (large Fano factor) or not (Fano factor is close to one), it really has nothing to do with translational bursting.

We realize that the term "translational bursting" was confusing and misleading. To address the reviewer's concern we now refer to protein expression variability arising from stochastic synthesis and decay of individual transcripts as "mRNA birth/death fluctuations" and not "translational bursting noise". We refer to protein noise arising from stochastic promoter transition as "promoter fluctuations". The entire paper has now been modified to discriminate between "mRNA birth/death fluctuations vs. promoter fluctuations" rather than "transcriptional vs. translational bursting".

I believe they should correct the presentation in these two respects (which would also entail changing the title accordingly), and then the paper could probably be published