

Research article

Open Access

## Master equation simulation analysis of immunostained Bicoid morphogen gradient

Yu Feng Wu<sup>\*1</sup>, Ekaterina Myasnikova<sup>2</sup> and John Reinitz<sup>1</sup>

Address: <sup>1</sup>Department of Applied Mathematics and Statistics, and Center for Developmental Genetics, Stony Brook University, Stony Brook, NY 11794-3600, USA and <sup>2</sup>Department of Computational Biology, Center of Advanced Studies, St.Petersburg State Polytechnic University, St.Petersburg, 195251, Russia

Email: Yu Feng Wu<sup>\*</sup> - yufeng1714@gmail.com; Ekaterina Myasnikova - myasnikova@spbcas.ru; John Reinitz - reinitz@odd.bio.sunysb.edu

<sup>\*</sup> Corresponding author

Published: 16 November 2007

Received: 26 July 2007

BMC Systems Biology 2007, 1:52 doi:10.1186/1752-0509-1-52

Accepted: 16 November 2007

This article is available from: <http://www.biomedcentral.com/1752-0509/1/52>

© 2007 Wu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The concentration gradient of Bicoid protein which determines the developmental pathways in early *Drosophila* embryo is the best characterized morphogen gradient at the molecular level. Because different developmental fates can be elicited by different concentrations of Bicoid, it is important to probe the limits of this specification by analyzing intrinsic fluctuations of the Bicoid gradient arising from small molecular number. Stochastic simulations can be applied to further the understanding of the dynamics of Bicoid morphogen gradient formation at the molecular number level, and determine the source of the nucleus-to-nucleus expression variation (noise) observed in the Bicoid gradient.

**Results:** We compared quantitative observations of Bicoid levels in immunostained *Drosophila* embryos with a spatially extended Master Equation model which represents diffusion, decay, and anterior synthesis. We show that the intrinsic noise of an autonomous reaction-diffusion gradient is Poisson distributed. We demonstrate how experimental noise can be identified in the logarithm domain from single embryo analysis, and then separated from intrinsic noise in the normalized variance domain of an ensemble statistical analysis. We show how measurement sensitivity affects our observations, and how small amounts of rescaling noise can perturb the noise strength (Fano factor) observed. We demonstrate that the biological noise level in data can serve as a physical constraint for restricting the model's parameter space, and for predicting the Bicoid molecular number and variation range. An estimate based on a low variance ensemble of embryos suggests that the steady-state Bicoid molecular number in a nucleus should be larger than 300 in the middle of the embryo, and hence the gradient should extend to the posterior end of the embryo, beyond the previously assumed background limit. We exhibit the predicted molecular number gradient together with measurement effects, and make a comparison between conditions of higher and lower variance respectively.

**Conclusion:** Quantitative comparison of Master Equation simulations with immunostained data enabled us to determine narrow ranges for key biophysical parameters, which for this system can be independently validated. Intrinsic noise is clearly detectable as well, although the staining process introduces certain limits in resolution.

## Background

Recently considerable attention has been given to the characterization and understanding of intrinsic molecular noise in biological systems [1-8]. Nearly all of these studies were performed using *in vivo* fluorescent reporters in single cell systems. In multicellular organisms, however, most quantitative gene expression data are obtained from fixed tissues. Examples of such data for the *Drosophila* segmentation system are contained in the FlyEx database, which provides spatiotemporal data on the expression of developmental segmentation genes [9]. These data on protein expression levels are at cellular resolution and were obtained by means of immunofluorescence histochemistry and confocal scanning microscopy. At a large spatial scale, expression levels in these embryos form expression domains characteristic for each gene, but smaller fluctuations in expression levels between adjacent nuclei appear random. In this paper, we investigate the question of whether these fluctuations are a consequence of intrinsic molecular noise or stem from some type of measurement uncertainty. These alternatives are, of course, not mutually exclusive.

A complicating factor in separating the above alternatives is that each one involves an unknown chemical mechanism. Intrinsic noise will have a major contribution from fluctuations in the rate of initiation of transcription, but the chemical mechanisms underlying this process in multicellular organisms are very poorly understood. Measurement uncertainty can stem from chemical causes such as fluctuations in the number of primary and secondary antibody molecules which bind to proteins in the fixed embryo. The chemistry of this process is also very poorly understood. If observations could be made on a process whose fluctuation properties could be reliably predicted by a numerical model, comparison of the predicted fluctuations with those observed will provide critical information for distinguishing whether observed nucleus-to-nucleus variations are a consequence of intrinsic biological noise or merely fluctuations arising from the staining procedure.

An excellent candidate for a process with predictable fluctuation properties is one that involves only diffusion and first order decay. There is good evidence that the formation of the protein gradient of the morphogen Bicoid (Bcd) takes place by means of these two processes. Bcd protein is distributed in an exponential profile along the anterior-posterior (A-P) axis with higher concentrations towards the anterior [10,11]. This gradient forms by translation of maternally deposited mRNA at the anterior pole of the embryo, and the synthesized protein spreads through the syncytial embryo by diffusion accompanied by decay [12]. The observed exponential profile corresponds to a solution of Fick's equation for a substance

undergoing first order decay and diffusing from a point source in one dimension, and hence it is reasonable to suppose that the reaction-diffusion mechanism leading to the formation of the gradient is reasonably well understood. Quantitative observations of this gradient in fixed tissue exhibit small fluctuations between neighboring nuclei, while the overall exponential profile ensures that such fluctuations can be monitored over a wide range of concentrations which extend to the lower limits of detectability.

The intrinsic fluctuations of the Bcd gradient will be well described by a stochastic Reaction-Diffusion Master Equation (RDME). Such equations typically do not have analytic solutions and are usually solved by running repeated stochastic simulations. A well known simulation algorithm due to Gillespie [13,14] performs an exact simulation of the Chemical Master Equation for a well mixed system. This method has been extended to spatially distributed systems by Elf and others [15,16]. These authors divide the spatially extended system into a series of subvolumes that are small enough to be regarded as well mixed. In each subvolume chemical reactions are simulated by Gillespie's algorithm, while diffusion between subvolumes is treated as a first order reaction.

In this study, we compare the results of such simulations to data in order to discover whether or not the data is sufficiently accurate as to exhibit the signature of a simple stochastic process. Stochastic processes underlying biological regulation can in general form complex patterns as a result of the reaction network, and for this reason a full consideration of 3 dimensional geometry is often necessary [15]. In the case considered here fluctuations occur passively in the course of diffusion and the statistical signature that we seek is independent of detailed geometry. For these reasons, we chose to model a 1 dimensional system.

## Results and Discussion

In the following, we first characterize the statistical properties of random variations in expression level between adjacent nuclei of individual embryos and compare them to the results of stochastic simulations. At the level of single embryos, we do not find a clear signature of stochastic processes in the data, but the need to separate spatially changing mean expression values from their variation limits the amount of statistical information that can be obtained from individual embryos. To address this problem, we consider ensembles of embryos, both over the whole dataset and over a restricted subset with low embryo-to-embryo variation.

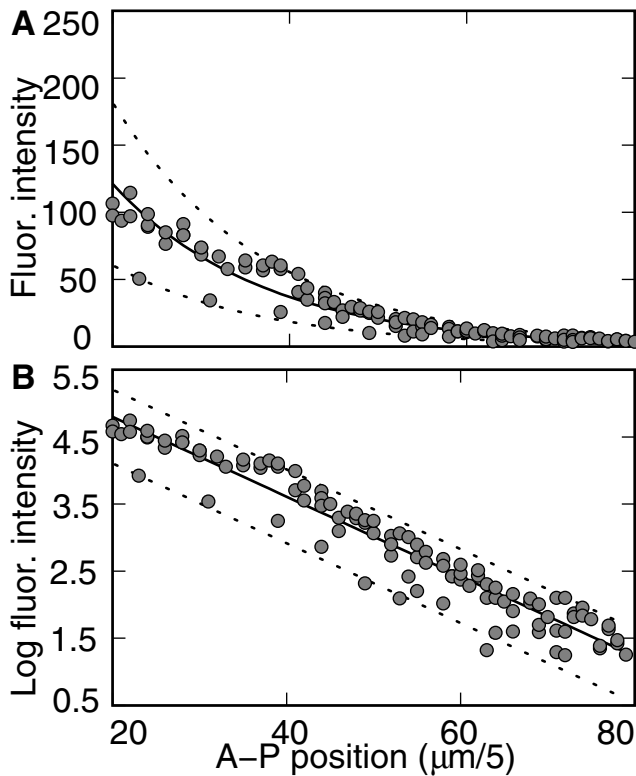
In order to interpret these data, we introduce a stochastic model of the immunostaining procedure. We denote all

random variables in this article with an upper hat but write a specific value without the hat. Thus, for example,  $\hat{n}_j$  denotes a random variable for the number of Bcd molecules in subvolume  $j$ , while  $n_j$  denotes a particular value of this variable.

**Single embryo analysis in the logarithmic domain**

We find that the Bcd profiles of individual embryos observed by immunostaining (Fig. 1A) are exponential, as previously reported [17-19]. This profile strongly supports the model of an effective Fickian diffusion which gives an exponential decay solution at steady state [12,20]. Each embryo contains a collection of observations of expression (with variation)  $I_j$ , and the observed properties of the exponential gradient will depend on the data through a function  $F$  embodying the least squares fitting procedure such that

$$F [I_j] = a \exp(-j/\lambda).$$



**Figure 1**  
**Bicoid profile from FlyEx embryo ms18 after background removal.** (A) The spatial index  $j$  ( $\mu\text{m}/5$ ) is the index of  $5 \mu\text{m}$  bins. The fluorescence intensity  $I_j$  (circles) were fit to an exponential  $F[I_j] = a \exp(-j/\lambda)$  (solid line) with two scaling index lines  $1.5F[I_j]$  and  $0.5F[I_j]$  (dashed). The fluorescence intensity were then converted into log scale in panel (B).

The residuals of the exponential vary from embryo to embryo in our data. In logarithmic coordinates the size of the residuals is independent of  $j$  for the anterior portion of the embryo, and in certain embryos (such as ms18) the size of the residuals is completely independent of  $j$  (Fig. 1B). For these embryos, the residuals are well described by

$$\ln(I_j) = \ln(a) - j/\lambda + \hat{W},$$

where  $\hat{W}$  is an unknown random variable independent of  $j$ , so that the variance of  $\ln(I_j)$  is equal to the variance of  $\hat{W}$ .

**Intrinsic noise is insufficient**

In order to understand whether the observed characteristic variation is a consequence of intrinsic fluctuations in Bcd molecular number, we performed stochastic simulations of an RDME which describes the time evolution of the Bcd gradient and compared them to data. We imagine the observed intensity  $I_j$  to be a particular value of the random variable  $\hat{I}_j$ . We further suppose that the random variable  $\hat{I}_j$  is determined by a direct linear rescaling of the Bcd molecular number such that

$$\hat{I}_j = m\hat{n}_j,$$

where the factor  $m$  represents the proportionality between one Bcd molecule and the corresponding fluorescence intensity, which includes the combined effects of tissue fixation, first and second antibody binding, fluorescence excitation and image processing.

As we do not know the exact *in vivo* system parameters and molecular number within each nucleus, we performed a complete inspection of the behavior of the variance of  $m\hat{n}_j$  in the four dimensional parameter space  $f(m, J, D, \omega)$ , where  $J$  is the synthesis rate of Bcd in the anterior compartment,  $\omega$  is its decay rate, and  $D$  is the diffusion coefficient. It was always true that the residuals in logarithmic coordinate increased towards the posterior of the embryo, or in other words at lower levels of Bcd. This behavior strongly contrasts with the position-independent residuals seen in Figure 1 and described in equation (2).

**Measurement rescaling noise dominates**

The simulation results suggest that intrinsic noise cannot explain the pattern of variance seen in Figure 1. Another possibility is the measurement process itself. In order to

analyze this process, we consider a simple model of the measurement of fluorescence intensity, where

$$\hat{I}_j = \hat{\alpha} \hat{n}_j + \hat{\beta}.$$

Here  $\hat{\alpha}$  is a spatially uniform random variable which replaces  $m$  in equation (3), and  $\hat{\beta}$  is a spatially uniform random variable which represents nonspecific background staining. This picture allows us to consider noise that arises from both intrinsic and measurement related sources.

The simplest way to understand the consequences of (4) is to imagine the consequences if only one of  $\hat{\alpha}$ ,  $\hat{n}_j$ , and  $\hat{\beta}$  are allowed to have finite variance while the other two are constrained to deterministic (zero variance) behavior. We have already pointed out that allowing finite variance for  $\hat{n}_j$  with  $\hat{\alpha}$  and  $\hat{\beta}$  deterministic leads to an increase of variance towards the posterior. The same is true if all noise comes from  $\hat{\beta}$ , that is if noise from background staining dominates. This will also lead to more noise in the logarithmic domain towards the posterior as  $\hat{\beta}$  provides a larger proportion of the total detected signal. Let us next consider the case where all noise comes from  $\hat{\alpha}$ .

In order to understand the role of  $\hat{\alpha}$ , note that the spatial pattern of variance observed in Figure 1 can be captured by an exponential function multiplied by a normal random variable. This suggests that the simplest picture for (4) is given by assuming that there is no background noise  $\hat{\beta}$  and no intrinsic noise for Bcd so that  $\hat{n}_j = \langle \hat{n}_j \rangle$ . Moreover, it suggests that the rescaling noise  $\hat{\alpha}$  from measurement uncertainty is uniform across the embryo (independent of  $j$ ) and is normally distributed with

$$\hat{\alpha} = m(1 + \sigma_\alpha \hat{N}(0,1)),$$

where  $\hat{N}(0,1)$  is a normal independent random variable with mean 0 and variance 1. Then we can model  $\hat{I}_j$  by

$$\hat{I}_j = \hat{\alpha} \langle \hat{n}_j \rangle.$$

In steady state,  $\langle \hat{n}_j \rangle = a/m \exp(-j/\lambda)$ . Taking logarithms allows us to write

$$\ln(I_j) = \ln(a) - j/\lambda + \ln(1 + \sigma_\alpha \hat{N}(0,1)).$$

Comparison with equation (2) indicates that  $\widehat{W} = \ln(1 + \sigma_\alpha \hat{N}(0,1))$ .

The reason we do not see intrinsic noise in this individual embryo is most likely low measurement sensitivity, that is to say a low molecule-to-fluorescence mean rescaling ratio  $m$ . When  $m$  is small enough, it can mask the variance of  $\hat{n}_j$  because the variance of the rescaled gradient is given by

$$\text{var}(m \hat{n}_j) = m^2 \text{var}(\hat{n}_j).$$

As  $m^2 \rightarrow 0$  with  $\text{var}(\hat{n}_j)$  bounded in a reasonable way throughout the embryo, we will get  $\text{var}(m \hat{n}_j) \rightarrow 0$ . Hence the rescaled gradient  $m \hat{n}_j$  can be treated as deterministic by letting

$$m \hat{n}_j = \langle m \hat{n}_j \rangle = m \langle \hat{n}_j \rangle.$$

In summary, we suspect the nucleus-to-nucleus variation observed in our data comes chiefly from the experimental rescaling noise  $\hat{\alpha}$ , which is normally distributed. If Bcd intrinsic noise is to be observed, then the fluorescence noise intensity should be a function of the mean intensity in logarithm, instead of a constant as observed in embryo ms18. Nevertheless, the necessity of considering data in spatially resolved bins limits the amount of information that can be obtained from a single embryo. More information can be obtained by pooling data from many embryos, and we discuss this point in the next section.

#### Statistical analysis of an ensemble of embryos

Statistical analysis of many embryos is required in order to take our analysis further. This analysis will show how physical constraints on the model can be inferred from the ensemble dataset, and independent random variables separated. We consider a set of embryos indexed by  $i$  with expression levels  $I_{ij}$ , and then pool data from corresponding bins to obtain the ensemble dataset  $\bigcup_i I_{ij}$ . Since this

dataset includes embryo-to-embryo variability, i.e. the variation of system parameters and experimental conditions over different embryos, the variance of the ensemble profile will be an upper bound for the average variance within each embryo. This allows us to identify physical

constraints for system parameters and to determine if the model behaves properly within the permitted range of parameters.

We define independent global random variables  $\hat{I}_j = \hat{\alpha}\hat{n}_j + \hat{\beta}$  as described in the last section with normal distributed measurement uncertainty  $\hat{\alpha} = m(1 + \sigma_\alpha \hat{N}(0,1))$ . We also now assume that background noise is normally distributed with  $\hat{\beta} = \sigma_\beta \hat{N}(0,1)$ . We assume such global random variable represent the average variability for each embryo. The statistics of simulated global random variables were then collected from 2000 stochastic simulation runs and the Bcd molecular number random variable  $\hat{n}_j$  were sampled after reaching steady state. In this section we explicitly consider the effects of different choices for the molecule-to-fluorescence rescaling ratio  $m$  by comparing simulations to data at differing values of this parameter. Because  $m$  is not an explicit input to the model, this comparison is effected by varying the synthesis rate  $J$ , which varies the molecular number, and comparing the behavior of the model to fluorescence data which is on a fixed but arbitrary scale.

We seek optimal values of  $m$  such that the simulated global random variable  $\hat{I}_j$  is constrained by the variation observed in the immunostained ensemble data  $\bigcup_i I_{ij}$  by the condition

$$\langle \hat{I}_j \rangle = \langle \bigcup_i I_{ij} \rangle$$

$$\text{var}(\hat{I}_j) \leq \text{var}(\bigcup_i I_{ij}).$$

Because comparison with mean and variance of the ensemble data is not straightforward, we approximate the above conditions by comparing their exponential fit,  $F$ , and the normalized variance  $\eta^2$ , where  $\eta^2 = \text{var}(I)/\langle I \rangle^2$ , and  $I = \hat{I}_j$  or  $I_{ij}$  for simulation and data respectively. Then (7a) and (7b) become

$$\langle \hat{I}_j \rangle = F[\bigcup_i I_{ij}]$$

$$\eta^2(I_j) \leq \eta^2(\bigcup_i I_{ij}).$$

In (8a),  $\langle \hat{I}_j \rangle = ma' \exp(-j/\lambda')$  from simulation and  $F[\bigcup_i I_{ij}] = a \exp(-j/\lambda)$  from the ensemble data. Because  $\lambda'$  is only determined by  $D$  and  $\omega$ , we first select combinations of  $D$  and  $\omega$  such that  $\lambda' = \lambda$ . Selecting a synthesis rate  $J$  determines  $a'$ , and also  $m$ , because  $m = a/a'$ . Finally, biologically reasonable values of  $J$  are determined by the constraint (8b).

We seek to establish which terms of equation (4) dominate the observed variance in different parts of the embryo. To do this, we will graphically compare the total observed variance with a set of simulations in which variance arises from different subsets of the random variables in (4). We denote these restricted models by placing brackets around the subset of random variables which contribute to the variance. Thus the full model in (4) can be denoted by  $\hat{I}_j = [\widehat{\alpha n \beta}]$ . A model with no variance contributed by background is denoted by  $[\widehat{\alpha n}] = \alpha n_j$ , while models in which all variance is contributed only by rescaling, background, or molecular number are denoted respectively by  $[\alpha] = \alpha \langle n_j \rangle$ ,  $[\beta] = m \langle n_j \rangle + \beta$ , and  $[n] = m n_j$ . Using this notation, a comparison of the definition of  $\eta^2$  with (4) shows that we can write

$$\eta^2(I_j) = \eta^2([\widehat{\alpha n}]) + \eta^2([\beta])$$

$$= \eta^2([n]) + \sigma_\alpha^2 \text{var}(n_j N(0,1)) / \langle n_j \rangle^2 + \eta^2([\beta]).$$

The middle term of the above equation simplifies even further when the molecular number is large so that  $\hat{n}_j = \langle \hat{n}_j \rangle$  and hence  $\sigma_\alpha^2 \text{var}(n_j N(0,1)) = \sigma_\alpha^2$ . If the contribution of background noise  $\hat{\beta}$  is also small then it is the case that

$$\eta^2(I_j) = \sigma_\alpha^2 = \eta^2([\alpha]).$$

In summary, when Bcd molecular number is large in the anterior region of the embryo, we expect to see a constant level of normalized variance in our fluorescence data, contributed solely by rescaling noise  $\hat{\alpha}$ . Thus,  $\hat{\alpha}$  can be identified independently from  $m$  and other random variables in our data.

### Physical constraints from a high-variance ensemble of embryos

We first show results of the above statistical comparison using all 89 FlyEx cycle 13 embryos. This ensemble contains 9400 nuclei, with about 150 nuclei per bin. In Figure 2A we see that the normalized variance of the ensemble data  $\eta^2(\bigcup_i I_{ij})$  asymptotically approaches the simulation curve  $\eta^2([\alpha]) = \sigma_\alpha^2$  in the anterior region of the embryo.  $\sigma_\alpha$  values that are too high will place the black model curve above the data points on the left, and this constrains  $\sigma_\alpha$  to be less than 0.2.

With regard to the molecular parameters of diffusion, adopting the value of  $D = 17.2$  ( $\mu\text{m}^2/\text{s}$ ) given in the literature [20] together with our observed value of  $\lambda = 80.65$  ( $\mu\text{m}$ ) for this ensemble implies that  $\omega = 0.0027$  ( $\text{s}^{-1}$ ) in order to satisfy (8a). These values, for any  $J$ , will cause the mean molecular number gradient,  $\langle \hat{n}_j \rangle$ , to be in steady state after about 4000 seconds (cycle 8). A lower bound of  $J > 30$  (molecules/s) and an upper bound of  $m < 0.7$  is imposed by the experimentally observed magnitude of  $\eta^2$ . Finally, we estimate the constraint for background noise  $\hat{\beta}$  by satisfying (8b) in the posterior end of the embryo. Violation of the inequality (8b) would cause the black model curve to be above the data on the right hand side of Figure 2A, and hence we have an upper bound  $\sigma_\beta < 1.7$ .

In summary, analysis of our high-variance ensemble of embryos dataset implies that the Bcd synthesis rate is higher than 30 (molecules/s). Figure 2B shows a scatterplot of the molecular number associated with this synthesis rate. The panel indicates that most nuclei in the anterior fifth of the embryo contain more than 200 molecules of Bcd after reaching steady state, and that these molecular numbers can fluctuate over a range of more than 80 molecules in this region (Fig. 2C). This panel shows that these molecular fluctuations, even at the largest level of normalized variance compatible with data, still do not account for the observed variance in experimental observations. The additional variance comes chiefly from rescaling noise  $\hat{\alpha}$ . Background noise  $\hat{\beta}$  only has a significant effect at the posterior end of the embryo, and indeed dominates the normalized variance in that region (Fig. 2A). Towards the posterior,  $\eta^2([\beta])$  rises faster than  $\eta^2(n_j)$ , and its sharp rise may in certain cases serve as a

marker to distinguish regimes dominated by molecular noise from those dominated by background noise.

### Physical constraints from a low-variance ensemble of embryos

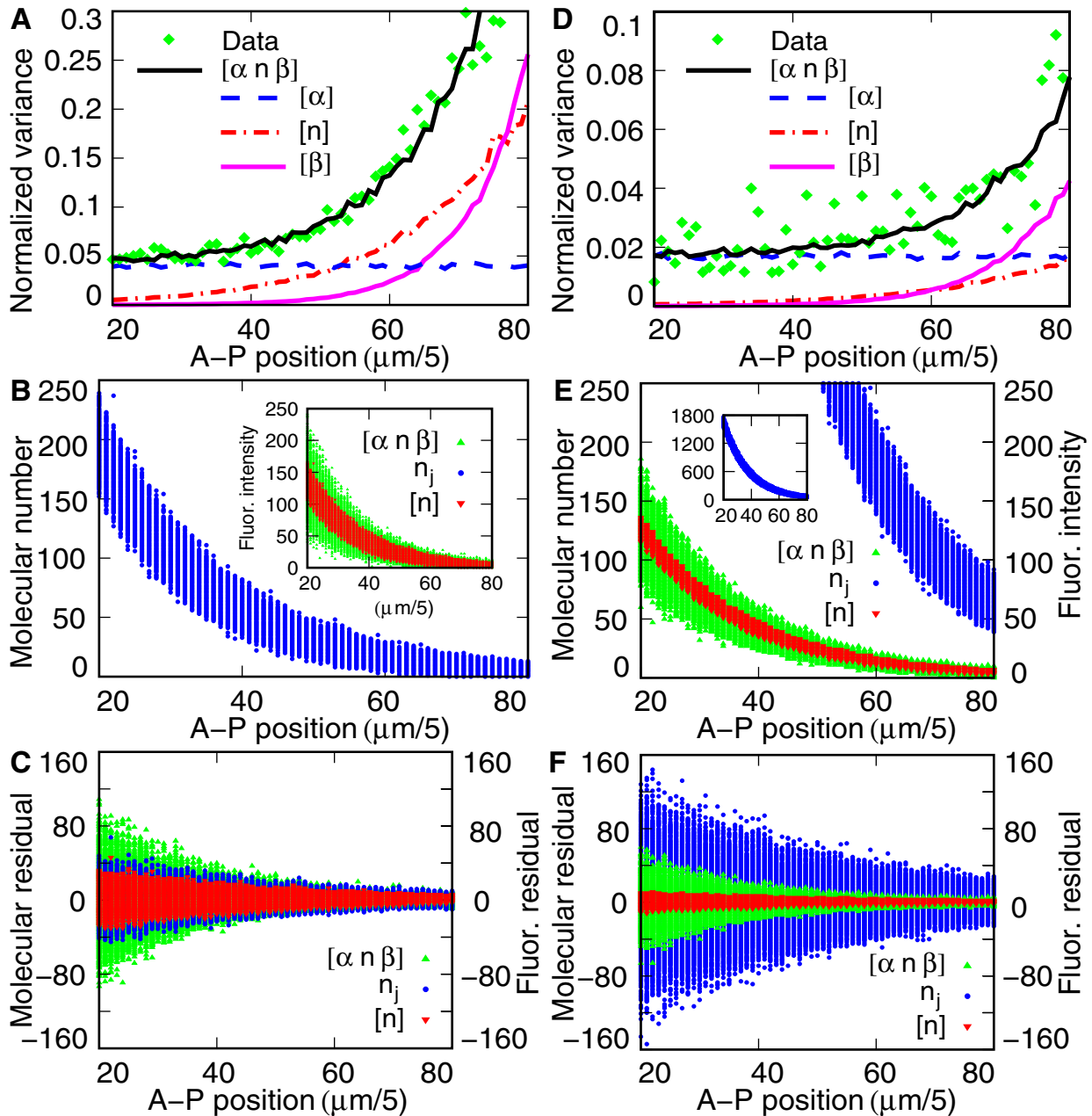
In the ensemble of embryos discussed above, a portion of the variance observed is likely to stem from embryo to embryo variation in staining conditions and inherent biological parameters. In order to find a better upper limit for observed molecular fluctuations, it is desirable to analyze a set of embryos whose properties are as uniform as possible. In order to generate such a set, we considered the value of  $\lambda_i$  in the exponential fit  $F[I_{ij}] = a_i \exp(-j/\lambda_i)$  to each embryo. We then grouped embryos according to common values of  $1/\lambda_i$ , taken to two decimal places. In the largest such group, which contained 17 members, we rescaled the data to a common amplitude by letting

$$I'_{ij} = \frac{\langle a_i \rangle}{a_i} I_{ij}.$$

These 17 processed embryos constitute an ensemble  $\bigcup_i I'_{ij}$  for statistical analysis as described in the previous section. A trade-off of this treatment is the loss of statistical sample size, with only around 30 nuclei in each bin.

Figure 2D shows that this ensemble of 17 embryos has lower normalized variance compared to the 89 embryos ensemble in Figure 2A. The fluctuation of normalized variance is also higher because of smaller sample size. Note that rescaling noise is dominant over a larger portion of the embryo than is the case for the full 89 embryo ensemble. We estimate an upper bound for rescaling noise  $\sigma_\alpha$  to be 0.13. At this point it is possible to determine the smallest  $J$  compatible with variance as was done for the high variance ensemble. We do not do so, however, because in this ensemble the contribution of  $\hat{\beta}$  is too small to be separated from  $\hat{n}_j$ .

We compared this data to simulations performed using the same parameter values reported in the previous section, except that  $\lambda$  (for the whole ensemble) had a value of 90.91 ( $\mu\text{m}$ ), leading to  $\omega$  taking on a value of 0.00215 ( $\text{s}^{-1}$ ). These values cause the deterministic system to relax to steady state after 4000 seconds, as was the case with the high variance ensemble. In the high variance ensemble (Fig. 2A), the choice of  $\sigma_\beta$  was dictated by the necessity of matching the observed variance along the entire A-P axis. In the case of the low variance ensemble this is not



**Figure 2**

**Physical constraints of Bicoid gradient from the ensemble of embryos data.** (A) The normalized variance. In the key, "Data" denotes the members of  $\eta^2(\bigcup_i I_{ij})$  from the high-variance ensemble of 89 embryos. Lines denote simulation results as shown in the key.  $[\alpha n \beta]$  denotes the full model  $\eta^2(I_j) = \eta^2([\widehat{\alpha n \beta}])$ ,  $[\alpha]$  denotes  $\eta^2([\alpha])$ ,  $[n]$  denotes  $\eta^2([n])$ , and  $[\beta]$

denotes  $\eta^2([\beta])$ . The parameters used in simulation were  $\sigma_\alpha = 0.2$ ,  $\sigma_\beta = 1.7$ ,  $J = 30$  (molecules/s),  $m = 0.7$ ,  $D = 17.2$  ( $\mu\text{m}^2/\text{s}$ ) and  $\omega = 0.0027$  ( $\text{s}^{-1}$ ). (B) is a scatterplot of Bcd molecular number  $\hat{n}_j$ , while the inset shows simulated fluorescence intensity for models and parameters used in panel A. (C) shows the residuals (deviations from mean) of panel B. (D), (E) and (F) show the same information as panel A, B, and C respectively, but from the low-variance ensemble of 17 embryos, with parameters used in simulation  $\sigma_\alpha = 0.13$ ,  $\sigma_\beta = 1.0$ ,  $J = 200$  (molecules/s),  $m = 0.07$ ,  $D = 17.2$  ( $\mu\text{m}^2/\text{s}$ ) and  $\omega = 0.00215$  ( $\text{s}^{-1}$ ). Note that the axes in panel A and D are scaled differently, and the absolute molecule number is shown in the inset to panel E.

required, but measurements from nonexpressing nuclei indicate that  $\sigma_\beta$  is equal to about 1. It is important to have at least a rough estimate for  $\sigma_\beta$  so that fluctuations from this source are not spuriously assigned to fluctuations in molecular number. Finally, these constraints require that the lower bound of synthesis rate  $J$  be 200 (molecules/s) in order that that  $\eta^2(I_j) = \eta^2(\bigcup_i I'_{ij})$ . The corresponding upper bound of molecule-to-fluorescence rescaling ratio is  $m = 0.07$ .

The lower limit of  $J$  imposed by this ensemble of low variance embryos in turn implies that there must be more than 300 Bcd molecules per subvolume in the middle of the embryo ( $j = 50$ ) after reaching steady state (Fig. 2E). Moreover, it implies that the Bcd molecular gradient does not drop to 0 as the fluorescence intensity reaches the presumed background level, but remains at a level of at least 50 molecules per subvolume at  $j = 80$ , and 36 molecules per subvolume at the posterior pole ( $j = 100$ ). Note that the variance of fluorescence measurements is similar between the two embryo ensembles (compare the green areas in Fig. 2C and Fig. 2F), but that the portion of that variance assigned to fluctuations in molecular number is smaller for the ensemble of 17 embryos (compare the red areas in Fig. 2C and Fig. 2F). The tighter constraints from the smaller ensemble make the lower limits on molecular number higher (compare Fig. 2B and Fig. 2E), and the lower limit on the variance of molecular number higher (compare the blue areas in Fig. 2C and Fig. 2F), although the higher limit of the normalized variance becomes smaller (compare the data points in Fig. 2A and Fig. 2D). The Bcd molecular number will thus vary by more than 100 molecules in the middle of the embryo. The 13% rescaling noise  $\hat{\alpha}$  is still the main source of the characteristic variation observed in the anterior region of our fluorescence intensity data.

### Noise strength

In most applications the most important measure of fluctuation is the normalized variance  $\eta^2$  [21,22]. A different quantity, known as the Fano factor or noise strength [23-25], has been used by some authors as a marker to distinguish different stochastic mechanisms. The Fano factor  $\nu$  is given by  $\nu(I_j) = \text{var}(I_j)/\langle I_j \rangle$ , where  $\nu = 1$  in a Poisson process. All stochastic simulations of Bcd intrinsic noise  $\hat{n}_j$  give  $\nu = 1$ , as shown in Figure 3A. By contrast, the full statistical model for either of the two ensembles of

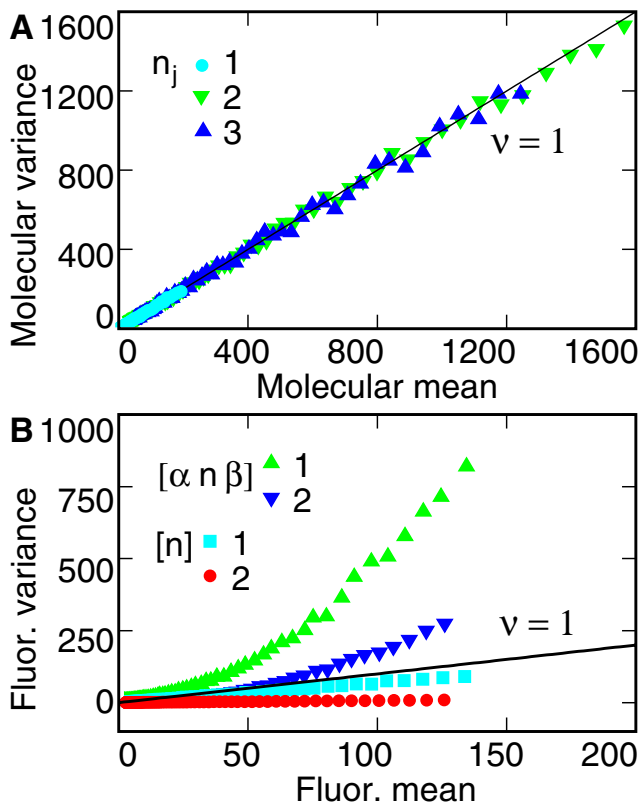
embryos examined (green and blue data in Fig. 3B) is obviously non-Poisson, not only because  $\nu \neq 1$  but also because in the data,  $\nu$  is a function of the mean. This happens because at larger values of molecular number, the variance of  $\hat{\alpha}$  has a dominant role, even when its value is small. Even if we model our data without rescaling noise using the  $[n]$  random variable alone, uncertainty in the value of the rescaling constant  $m$  itself leads to ambiguity in the observed value of the Fano factor (red and cyan data in Fig. 3B). This is a natural consequence of the dimensions of the Fano factor.

### Conclusion

We have compared the nucleus to nucleus variation in expression levels of the exponentially distributed Bcd gradient observed in fixed tissue in a steady state with a stochastic model of the diffusion equation. The model is well supported, in the sense that there is a well-supported physical model for the spatial dependence of mean concentrations of Bcd [12,20] on the scale of the embryo. The first major result of our analysis is to note that in many individual embryos the nucleus to nucleus variation in the log of concentration is independent of spatial position. This pattern of variation, which amounts to multiplicative noise in concentration space, is completely incompatible with the stochastic behavior of the diffusion equation. Simulations of the diffusion equation over an exhaustively large region of parameter space without exception give rise to solutions in which nucleus to nucleus variation of the *bcd* gradient is a function of position in the embryo, whether this variation is measured directly in Bcd levels or in their logarithms.

The data which we compare the model to is in the form of fluorescence levels, not concentrations. Although there is now good evidence that the specific batch of serum used to obtain this data has a mean response to Bcd [26] which is linear, there is no quantitative information about the variance of this sensitivity. Previous work on intrinsic molecular noise in yeast and bacteria utilized GFP [27,28] *in vivo*, a situation where fluorescence is detected without molecular amplification. In the data reported here, and in most studies with fixed tissue, the signal from bound primary antibodies is amplified by incubation with secondary antibodies conjugated to a fluorescent dye. It is easily imaginable that this amplification process itself is subject to molecular fluctuations. These fluctuations would then give rise to rescaling noise in the proportionality between fluorescence levels and primary antigen molecular number from nucleus to nucleus. We have shown that such variation can explain the multiplicative noise observed.



**Figure 3**

**Mean to variance correlation.** (A) The noise strength (Fano factor)  $\nu$  of simulated Bicoid molecular number gradient is defined as molecular variance  $\text{var}(n_j)$  divided by molecular mean  $\langle \hat{n}_j \rangle$ . The key indicates that parameters were obtained from (1) the high-variance ensemble of 89 embryos, (2) the low-variance ensemble of 17 embryos, and (3) the extreme condition of high diffusion rate  $D = 7890$  ( $\mu\text{m}^2/\text{s}$ ), decay rate  $\omega = 1.0$  ( $\text{s}^{-1}$ ) and synthesis rate  $J = 70000$  (molecules/s). (B) As shown in the key, noise strength of simulated fluorescence intensity  $\hat{I}_j = [\alpha n \beta]$  and rescaled gradient  $[n]$  were obtained using parameters from (1) the high-variance ensemble of 89 embryos and (2) the low-variance ensemble of 17 embryos.

When data from fixed embryos are pooled, better statistics are obtained. We analyzed data pooled from the entire dataset ( $N = 89$ ) as well as a smaller pool of data from a set of embryos selected to have nearly identical mean Bcd profiles ( $N = 17$ ). In order to analyze these data, we constructed an explicit statistical model. The model considers three sources of variance: intrinsic noise, rescaling noise, and background noise. By means of this statistical model it is possible to separate, at least roughly, the contributions of different sources of fluctuation. We have dis-

cussed mechanisms for the first two of these; the third, background noise, represents small fluctuations in the quantity of nonspecific molecules (background) from nucleus to nucleus. Because mean background has been previously removed from this data [29], the background noise has a mean of zero. We have confirmed that the background removal method does not affect our results, and the mean background intensity before removal is independent from the Bcd molecule-to-fluorescence rescaling ratio (data not shown). We also found from non-expressing areas of our data that the background noise (standard deviation) has about 54% positive correlation to the mean background intensity.

The results of this analysis constrained the physical parameters of the stochastic model considerably, with sharper constraints provided by the smaller dataset. The data require that the synthesis rate,  $J$ , of Bcd from its pool of anteriorly deposited mRNA be greater than 200 (molecules/s). We chose the subvolumes of the model to have the same volume as a nucleus, and hence the constraint on  $J$  also implies that there are a mean of at least 300 molecules of Bcd per nucleus at the midpoint of the embryo, and mean levels of at least 36 molecules of Bcd per nucleus at the posterior pole. In terms of concentration, our results show that Bcd concentration at the midpoint of the embryo is greater than 4 nM. Recently, a direct *in vivo* measurement of Bcd concentration was performed and yielded a mid embryo Bcd concentration of 8 nM [30], fully compatible with our results.

Although we are able to extract a clear signature of intrinsic molecular noise from the data by means of the statistical model, we also showed that at least one quantity diagnostic for different stochastic mechanisms, the Fano factor, cannot be read out from the data. Although the Fano factor  $\nu = 1$  in the simulations, the full statistical model gives rise to a Fano factor which is a function of the mean, and even if all noise is restricted to be intrinsic, the observed Fano factor depends on the scale for conversion from fluorescence to molecular number.

We believe that our results in general demonstrate that fixed material processed with secondary fluorophores is not well suited to studies of molecular fluctuations. This arises from three issues, which may be separable. Fixation obviously prevents repeated observations on the same cell. While that is clearly a limitation, it need not affect an investigation of a molecule whose mean values are in steady state. The other two issues concern amplification. GFP is intrinsically fluorescent, but antibodies must bind to antigen, a process that is in itself subject to intrinsic molecular fluctuations. In the present study, the level of molecular fluctuation is doubtless increased by the need to bind secondary antibodies conjugated to fluorophore

to the already bound primary antibody. It is thus possible that better data from fixed tissue could be obtained by conjugating dye directly to the primary antibody. This may prove to be a useful measure in situations where constructing a GFP fusion that can functionally substitute for the native gene is difficult or impossible.

More generally, we suggest that it is important to know how precision and robustness of developmental control is achieved at the molecular number level throughout development. Indeed, there is little doubt that stochastic processes are important later in development. Adult *Drosophila* normally have four scutellar bristles. In mutants of the *scute* gene, the number of bristles varies between one and three [31], strongly suggesting a stochastic process. On a more theoretical level, it has been suggested that fluctuations can augment the operating capabilities of biological regulatory networks [32]. Our results indicate that *in vivo* monitoring of gene expression will be required to obtain high quality data on stochastic gene expression phenomena in eucaryotes. The central technical problem that must be solved to conduct such studies is the complete replacement of the endogenous gene with a fluorescently tagged functional version [30].

## Methods

### Stochastic simulations

The Bcd gradient was modeled in one dimension with 101 homogeneous cubic subvolumes indexed by  $j$  from 0 to 100. Each subvolume has sides of length  $l = 5 \mu\text{m}$  and volume  $\Delta = l^3$ . These dimensions were chosen in light of those of actual *Drosophila* embryos, which are  $500 \mu\text{m}$  long. The subvolume dimensions are very close to those of blastoderm nuclei. In the first subvolume ( $j = 0$ ), corresponding to the anterior pole of the embryo, we assume a zero-order synthesis reaction of Bcd molecules with constant rate  $J$  (molecules/s), representing the translation of a maternally deposited and localized stationary mRNA pool after egg deposition. The  $j$ th subvolume contains  $n_j$  molecules of Bcd, which are the state variables of the model. We take initial conditions to be  $n_j = 0 \forall j$ . Diffusion of Bcd is modeled as a first-order elementary reaction for the exchange of molecules between neighboring subvolumes with rate constant  $k = D/l^2$  seconds<sup>-1</sup>, where  $D$  is the effective Fickian diffusion constant. Dispersed degradation (decay) of Bcd is also modeled as a first-order reaction in all subvolumes with rate constant  $\omega$  (seconds<sup>-1</sup>).

Thus, for subvolumes  $j = 0$  to  $j = 100$ , the RDME is given by

$$\begin{aligned} \frac{dP(\{n_j\}, t)}{dt} = & (\mathbb{E}_0^{-1} - 1)JP(\{n_j\}, t) \\ & + \sum_{j=0}^{100} [(\mathbb{E}_j^{+1} - 1)\omega n_j P(\{n_j\}, t)] \\ & + \sum_{j=0}^{100} \sum_{k \in \{j \pm 1\}} [(\mathbb{E}_j^{+1} \mathbb{E}_k^{-1} - 1) \left(\frac{D}{l^2}\right) n_j P(\{n_j\}, t)], \end{aligned}$$

where  $P(\{n_j\}, t)$  is the joint probability of state vector  $\{n_j\} = [n_0, \dots, n_j, \dots, n_{100}]$ . The state operator,  $\mathbb{E}$ , is defined so that  $\mathbb{E}_j^{\pm 1} f(\dots, n_j, \dots) = f(\dots, n_j \pm 1, \dots)$ . Monte Carlo simulations of the behavior of this equation were obtained using the publicly available software MesoRD 0.2.0 [33]. MesoRD can automatically generate a stochastic or deterministic model from its input, and we make use of this feature in the work presented here. In the deterministic limit, mesoRD calculates the mean trajectory  $\langle \hat{n}_j \rangle(t)$ . By converting the initial values of the state variables into concentrations, mesoRD can then integrate the classical Reaction-Diffusion Rate Equation (RDRE) given by

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} - \omega C,$$

where  $C$  is the concentration of Bcd. Boundary conditions are given by  $\partial_x C|_{x=0} = -J/\Delta$  and  $\partial_x C|_{x=500} = 0$ . The steady-state solution can be well approximated by  $C(x) \approx a \exp(-x/\lambda)$ , where  $a = (J/\Delta)\lambda$  and  $\lambda = \sqrt{(D/\omega)}$ . MesoRD solves the deterministic equation by fixing the mesh size at the number of subvolumes chosen for the stochastic system. We used the built-in Euler solver with a stepsize of 0.001 second after verifying that these settings yielded stable and accurate solutions. Further analysis of simulations in comparison with quantitative immunostained data were performed in MATLAB.

### Quantitative data

We used Bcd protein expression data from cleavage cycle 13 [34] that were downloaded from the FlyEx database [9,35]. In FlyEx, confocal scans have been processed into tables containing average fluorescence levels in each nucleus [36]; these fluorescence levels are linearly proportional to Bcd concentration [26], and hence to molecular number. Data were taken from the central 10% strip along the A-P axis with their D-V coordinate suppressed, and normalized to remove the non-specific background [29]. The gradient is in a steady state at cycle 13 [19]. For quantitative analysis and comparison with the model, we

pooled data into 5  $\mu\text{m}$  1D intervals to compare with the 5  $\mu\text{m}$  subvolumes of the stochastic model. For certain purposes we considered a 17 member subset of embryos among which the inverse of the spatial exponential coefficient  $\lambda$  varied by less than 1%. This subset contained embryos ab15, cb2, ac1, hz8, ac6, iz4, cb19, ac2, cb31, iz15, ac16, ad18, ad31, hz30, be1, hz33 and as22 from FlyEx.

### Authors' contributions

All authors participated in the design of the study, and wrote the manuscript together. YF conducted the analysis of data and simulations, and made the figures. All authors read and approved the final manuscript.

### Acknowledgements

We thank Alexander Spirov and Manu for valuable discussions. We thank David Fange for the help with MesoRD installation. This work was supported by NIH award ROI RR07801 and by CRDF GAP Award RUB-1578.

### References

- 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.** *Nature Genetics* 2006, **38**:636-643.
- Blake WJ, Kaern M, Cantor CR, Collins JJ: **Noise in eukaryotic gene expression.** *Nature* 2003, **422**:633-637.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell.** *Science* 2002, **297**:1183-1186.
- Newman JRS, 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.
- Pedraza JM, van Oudenaarden A: **Noise propagation in gene networks.** *Science* 2005, **307**:1965-1969.
- Raj A, Peskin CS, Tranchinal D, Vargas DY, Tyagi S: **Stochastic mRNA synthesis in mammalian cells.** *PLoS Biology* 2006, **4**:0001-0013.
- Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Gene regulation at the single-cell level.** *Science* 2005, **307**:1962-1965.
- Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J: **Origins of extrinsic variability in eukaryotic gene expression.** *Nature* 2006, **439**:861-864.
- Poustelnikova E, Pisarev A, Blagov M, Samsonova M, Reinitz J: **A database for management of gene expression data in situ.** *Bioinformatics* 2004, **20**:2212-2221.
- Driever W, Nüsslein-Volhard C: **A gradient of Bicoid protein in *Drosophila* embryos.** *Cell* 1988, **54**:83-93.
- Struhl G, Struhl K, Macdonald PM: **The gradient morphogen Bicoid is a concentration-dependent transcriptional activator.** *Cell* 1989, **57**:1259-1273.
- Houchmandzadeh B, Wieschaus E, Leibler S: **Precise domain specification in the developing *Drosophila* embryo.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2005, **72**(6 Pt 1):061920. [Art. No. 061920 Part 1]
- Gillespie DT: **A general method for numerically simulating the stochastic time evolution of coupled chemical reactions.** *The Journal of Computational Physics* 1976, **22**:403-434.
- Gillespie DT: **Exact stochastic simulation of coupled chemical reactions.** *The Journal of Physical Chemistry* 1977, **81**:2340-2361.
- Elf J, Ehrenberg M: **Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases.** *Syst Biol* 2004, **2**:230-236.
- Fange D, Elf J: **Noise-induced Min phenotypes in *E. coli*.** *PLoS Computational Biology* 2006, **2**:0637-0648.
- Driever W, Nüsslein-Volhard C: **The Bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner.** *Cell* 1988, **54**:95-104.
- Houchmandzadeh B, Wieschaus E, Leibler S: **Establishment of developmental precision and proportions in the early *Drosophila* embryo.** *Nature* 2002, **415**:798-802.
- Surkova S, Samsonova M, Myasnikova E, Kosman D, Kozlov K, Samsonova A, Vanario-Alonso CE, Reinitz J: **Characterization of the *Drosophila* segment determination morphome.** *Dev Biol* 2007 in press.
- Gregor T, Bialek W, Steveninck RR, Tank DW, Wieschaus EF: **Diffusion and scaling during early embryonic pattern formation.** *PNAS* 2005, **102**:18403-18407.
- Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, Brent R: **Regulated cell-to-cell variation in a cell-fate decision system.** *Nature* 2005, **437**:699-706.
- Paulsson J: **Summing up the noise in gene networks.** *Nature* 2004, **427**:415-418.
- Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A: **Regulation of noise in the expression of a single gene.** *Nature Genetics* 2002, **31**:69-73.
- Paulsson J: **Models of stochastic gene expression.** *Physics of Life Reviews* 2005, **2**:157-175.
- Raser JM, O'Shea EK: **Control of stochasticity in eukaryotic gene expression.** *Science* 2004, **304**:1811-1814.
- Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW: **Stability and nuclear dynamics of the Bicoid morphogen gradient.** *Cell* 2007, **130**:141-152.
- Austin DW, Allen MS, McCollum JM, Dar RD, Wilgus JR, Sayler GS, Samatova NF, Cox CD, Simpson ML: **Gene network shaping of inherent noise spectra.** *Nature* 2006, **439**:608-611.
- Becskei A, Kaufmann BB, van Oudenaarden A: **Contributions of low molecule number and chromosomal positioning to stochastic gene expression.** *Nature Genetics* 2005, **37**:937-944.
- Myasnikova E, Samsonova M, Kosman D, Reinitz J: **Removal of background signal from *in situ* data on the expression of segmentation genes in *Drosophila*.** *Development, Genes and Evolution* 2005, **215**:320-326.
- Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW: **Probing the limits to positional information.** *Cell* 2007, **130**:153-164.
- Rendel JM: **Canalization of the scute phenotype of *Drosophila*.** *Evolution* 1959, **13**:425-439.
- Paulsson J, Berg OG, Ehrenberg M: **Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation.** *PNAS* 2000, **97**:7148-7153.
- Hattne J, Fange D, Elf J: **Stochastic reaction-diffusion simulation with MesoRD.** *Bioinformatics* 2005, **21**:2923-2924.
- Foe VE, Alberts BM: **Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis.** *J Cell Sci* 1983, **61**:31-70.
- FlyEx Database [<http://flyex.ams.sunysb.edu/flyex/>]
- Janssens H, Kosman D, Vanario-Alonso CE, Jaeger J, Samsonova M, Reinitz J: **A high-throughput method for quantifying gene expression data from early *Drosophila* embryos.** *Development, Genes and Evolution* 2005, **215**:374-381.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

