

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)**SciVerse ScienceDirect**

Procedia Computer Science 9 (2012) 373 – 382

**Procedia**  
Computer Science

International Conference on Computational Science, ICCS 2012

## Measuring gene expression noise in early *Drosophila* embryos: nucleus-to-nucleus variability

Nina E. Golyandina<sup>a</sup>, David M. Holloway<sup>b</sup>, Francisco J.P. Lopes<sup>c</sup>, Alexander V. Spirov<sup>d,e,\*</sup>, Ekaterina N. Spirova<sup>f</sup>, Konstantin D. Usevich<sup>a,\*\*</sup>,<sup>a</sup>Department of Mathematics and Mechanics, St.-Petersburg State University, St.-Petersburg, Russia<sup>b</sup>Mathematics Department, British Columbia Institute of Technology, Burnaby, B.C., Canada<sup>c</sup>Instituto de Biofísica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil<sup>d</sup>Computer Science and CEWIT, SUNY Stony Brook, Stony Brook, New York, USA<sup>e</sup>The Sechenov Institute of Evolutionary Physiology & Biochemistry, St.-Petersburg, Russia<sup>f</sup>AMS, SUNY Stony Brook, Stony Brook, New York, USA\*Corresponding author: [Alexander.Spirov@sunysb.edu](mailto:Alexander.Spirov@sunysb.edu)

\*\*Current address: School of Electronics and Computer Science, University of Southampton, Southampton, UK

The authors are listed in alphabetical order

---

### Abstract

In recent years the analysis of noise in gene expression has widely attracted the attention of experimentalists and theoreticians. Experimentally, the approaches based on in vivo fluorescent reporters in single cells appear to be straightforward and effective tools for bacteria and yeast. However, transferring these approaches to multicellular organisms presents many methodological problems. Here we describe our approach to measure between-nucleus variability (noise) in the primary morphogenetic gradient of Bicoid (Bcd) in the precellular blastoderm stage of fruit fly (*Drosophila*) embryos. The approach is based on the comparison of results for fixed immunostained embryos with observations of live embryos carrying fluorescent Bcd (Bcd-GFP). We measure the noise using two-dimensional Singular Spectrum Analysis (2D SSA). We have found that the nucleus-to-nucleus noise in Bcd intensity, both for live (Bcd-GFP) and for fixed immunostained embryos, tends to be signal-independent. In addition, the character of the noise is sensitive to the nuclear masking technique used to extract quantitative intensities. Further, the method of decomposing the raw quantitative expression data into a signal (expression surface) and residual noise affects the character of the residual noise. We find that careful masking of confocal images and use of appropriate computational tools to decompose raw expression data into trend and noise makes it possible to extract and study the biological noise of gene expression.

**Keywords:** fluorescence imaging; image segmentation; measuring gene expression; gene expression noise; noise filtration; noise analysis; Singular Spectrum Analysis (SSA); 2D SSA

---

### 1. Introduction

In recent years the analysis of noise in gene expression has widely attracted the attention of both experimentalists and theoreticians. Observations on live bacterial and yeast cells are a straightforward way to do this. The approach is based on genetically engineered fluorescent proteins and fluorescence imaging. Particularly, the dual-reporter method was firstly applied to bacteria [1,2], and a little later to yeast cells [3,4]. Noise is present in some of the

critical processes in living systems and robustness to noise should be one of the main design principles of the associated gene regulatory networks [5].

Noise in gene expression arises not only from the inherent randomness of biochemical processes such as transcription and translation (intrinsic noise), but also from fluctuations in cellular components that lead indirectly to variation in the expression of a particular gene (extrinsic noise) [e.g. 6]. Extrinsic noise arises from fluctuations in cellular components such as regulatory proteins and polymerases, and has a global effect [1]. Intrinsic noise arises from the stochastic nature of the biochemical process of gene expression and causes identical copies of a gene to express at different levels [1]. Early noise projects found that intrinsic noise in bacteria and yeast tended to be well-fit by Poisson birth and death processes [4,7].

Detailed consideration of the unicellular experiments raises some crucial questions as to what extent the measured noise is chiefly the molecular noise of gene transcription and translation. Even ignoring the observational noise (e.g. in fluorescence measurements), we still have to pay attention to biological noise sources such as active molecular transport, compartmentalization, and the mechanics of cell division etc. [e.g. 8]. We might expect that these noise sources are relatively low and controllable in bacterial and yeast populations, but this assumption becomes far less appropriate in the case of whole embryo observations (tissues of many thousands of cells).

The ideas and approaches for transcription/translation noise in simple single cells has started to be transferred to higher multicellular organisms, primarily the early *Drosophila* embryo [9-12], with tempting parallels between fluorescence of separate nuclei in *Drosophila* blastoderm images and fluorescent reporters in separate cells in bacteria and yeast experiments. In both cases we can mask the sources of fluorescence (bacteria / yeast cells vs. nuclei in the blastoderm), measure their intensity, and estimate noise. The problem is to what extent the nucleus-to-nucleus variability corresponds to the intrinsic noise of gene expression: in bacteria and yeast, observed fluctuations are likely to be more directly due to intrinsic transcription/translation noise; in embryos, tissue-level effects such as molecular transport and compartmentalization begin to contribute more strongly to the observed fluctuations.

Serious observational limitations in confocal scanning of whole *Drosophila* embryos have been raised by several authors (e.g. Myasnikova et al. ([10,11], cf [12])). We find, by comparing fluorescence intensity noise in confocal scanning of live Bcd-GFP embryos (unpublished) with fixed immunostained embryos [13,14], that the nature of the noise is quite far from the simple poissonian birth and death processes applied to single cells. In particular, we have noticed that noise in the mRNA signal (FISH technique) can be practically independent of trend (Supplementary Text S3 & Supplementary Figure S12 [14]). These preliminary observations inspired us to apply the approach to protein data and to analyse the noise trends in detail. We believe the single-celled view of nuclear fluorescence directly resulting from point transcription and translation (using traditional chemical kinetic approaches) does not conform to the biological reality in embryos.

In complex multicellular tissues, nuclear fluorescence is affected by the complicated and relatively poorly studied processes of active (energy-dependent) transport of molecules from one compartment to another. The geometry and dynamics of the compartments and their relation to the molecular machinery of the active transport is still puzzling. Statistics for the abundance of a given transcription factor in a nucleus is not likely to be simply modeled by elementary polypeptide synthesis on ribosomes. We need to pay attention to the many coupled processes of stratification and compartmentalization of the cytoplasm, which is highly dynamic and cell-cycle dependent; for example, the well-known local and whole-embryo scale cytoplasmic movements, especially during mitosis [15,16]. In recent work, we have found that the ‘texture noise’, arising from the pre-cellular compartmentalization of the embryo surface, which is highly dynamic in time, is a major component of total biological noise, and can exceed gene transcription/translation noise [13].

In this communication, we focus on nucleus-to-nucleus variability in the Bicoid (Bcd) transcription factor. Our approach involves comparing confocal data for Bcd intensity between fixed immunostained and live embryos. We focus on several hour old embryos, at the 13<sup>th</sup> nuclear cleavage cycle (cc13). At this stage, the embryo is a precellular syncytial blastoderm, with  $\sim 2^{12}$  nuclei, mostly located just under the surface.

Since Bcd is a transcription factor, it is actively imported into nuclei, where it can act in gene regulation. Nuclei show much higher Bcd intensity than the surrounding cytoplasm. Looking closely at the fluorescence, it can be seen that Bcd tends to form granules inside nuclei. This is seen in live embryos (Fig 1A) and in fixed immunostained embryos (Fig 1B). As noted in [17] “GFP–Bcd appears to form punctate spots within anterior nuclei, possibly reflecting binding of GFP–Bcd to target genes, or localization of GFP–Bcd to a subnuclear compartment.” And in [17] “Confocal microscopy revealed that GFP–Bcd binds to many bands in the polytene chromosomes of larval salivary glands.” Similarities and differences between the Bcd patterns in live and fixed embryos can form a basis

for extracting real biological noise from the confocal images.

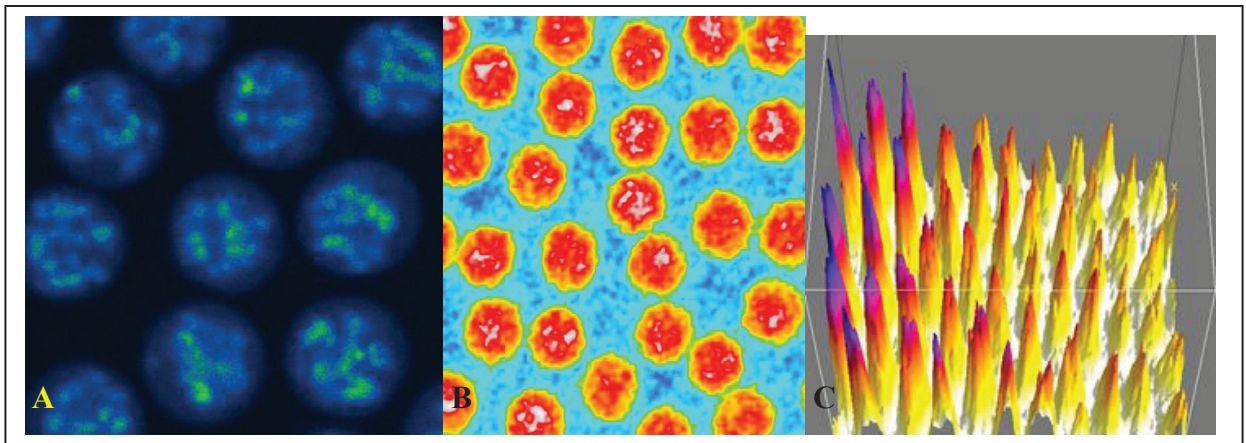


Figure 1. Nucleus-to-nucleus variability: Bcd signal localization in live (A) and fixed (B) nuclei. A) Bcd-GFP signal in live nuclei is mainly concentrated in compact dense particles and only weak signal is seen in the rest of the nucleoplasm. Probably what we are observing is euchromatin with hundreds of Bcd-GFP molecules bound to it. B) Nuclei in a fixed embryo immunostained for Bcd. Gray image has been colorized by ImageJ. Brighter granules are clearly seen and comparable with live observations (A). C) 3D representation of nuclear images like (A) and (B) by the Surface Plot 3D plugin for ImageJ software, making nuclei look like peaks.

At this stage of development each nucleus can contain  $\sim 2500$  Nucleus Pore Complexes, NPC [18]. This density of NCPs can pump a few hundred Bcd molecules into a nucleus in a few minutes ([19] and our observations). This keeps the Bcd nuclear concentration about four times higher than in the surrounding cytoplasm [20]. Specifically, maximum concentrations near the anterior pole are estimated to be  $\sim 5420$  Bcd molecules per nucleus (cc14), while concentrations near the midpoint of the embryo are  $\sim 690$  molecules per nucleus [19]. There are estimated to be hundreds to thousands of binding sites in each nucleus, probably mainly occupied at higher Bcd concentrations [21,22]. The molecular machinery of Bcd nuclear sequestration is complicated and still poorly understood [22].

The speed of Bcd nuclear import keeps nuclear concentrations in dynamic equilibrium with the surrounding cytoplasm. At the relatively low numbers of Bcd molecules and NPCs, it is expected that natural fluctuations in the nuclear import/export process could contribute significantly to observed variability – such a central process as the re-population of nuclear Bcd after each nuclear cleavage is likely to strongly affect Bcd's 'external noise' contribution to gene expression variability.

In this communication, we process and analyze the same embryo images from the FlyEx database used in [9] and compare these to our own observations on live Bcd-GFP embryos. We use original methods to process embryos and decompose spatially 2D fluorescence data into signal and noise, both for fixed and live embryos. We find that the observed nucleus-to-nucleus variability tends to be signal-level independent, both in fixed and live embryos. We conclude by discussing how our approach allows us to characterize biological noise, separated from observational or experimental contributions to noise.

## 2. Methods and Approaches

### 2.1. Confocal scanning of live embryos

We use a transgenic fly line in which GFP (Green Fluorescent Protein) has been fused onto the Bcd protein (gift of Dr. Wieschaus, [20]). With these flies, we can follow developmental dynamics in live embryos. FPs have been instrumental to studying noise in single-celled organisms, and it is hoped that using FP-tagged proteins can eliminate many sources of experimental noise in whole embryo imaging. However, confocal scanning of FPs has its own serious limitations. FPs are very sensitive to bleaching and live embryos are very sensitive to photo-damage. This makes photon-detection noise substantially higher than in fixed immunostained embryos, making it hard to

follow long developmental periods. Yolk autofluorescence also interferes with GFP. We now have a fair degree of experience with Bcd-GFP, however, and have developed a series of confocal protocols to minimize these difficulties. For instance, published studies with Bcd-GFP have been performed at low speed (one frame per 1-3 minutes). We found that scanning the mid-sagittal plane with a very small pinhole (0.5 airy) and low resolution (512\*512 or even 256\*256 pixels) minimizes bleaching. This allows us both to achieve 1 frame/sec. time resolution, allowing us to observe such processes as Bcd redistribution during mitosis.

## 2.2 Processing of confocal images

Confocal images of fixed immunostained embryos from the FlyEx database and scanning results of live Bcd-GFP embryos were processed using the same data processing protocol (summarized in Fig 2):

- *(Segmentation)* The regions of interest (ROI) of nuclei are detected in a confocal image and the average expression of each nuclei is computed.
- *(Regularization)* The irregularly sampled expression data is cropped and resampled to the image (data mapped onto a regular grid).
- *(Smoothing)* The image of expression is smoothed by non-parametric methods. The values of smoothed expression in nuclei centers and the nucleus-to-nucleus variability are determined.

In this paper we use circular ROI for segmentation, with both automatic and manual detection of nuclei centers. For this purpose special ImageJ scripts were designed. Images from FlyEx were also segmented by Khoros / VisiQuest software by using publicly available workspaces and recommended settings (<http://urchin.spcas.ru/flyex/>) to compare with our segmentation results.

In the smoothing step the data is decomposed by a non-parametric 2D-extension of Singular Spectrum Analysis (2D-SSA [23]), which we have previously applied to Bicoid pattern analysis. For the regularization step MATLAB scripts were developed, and for smoothing the 2D-SSA software was employed (as described in [14]).

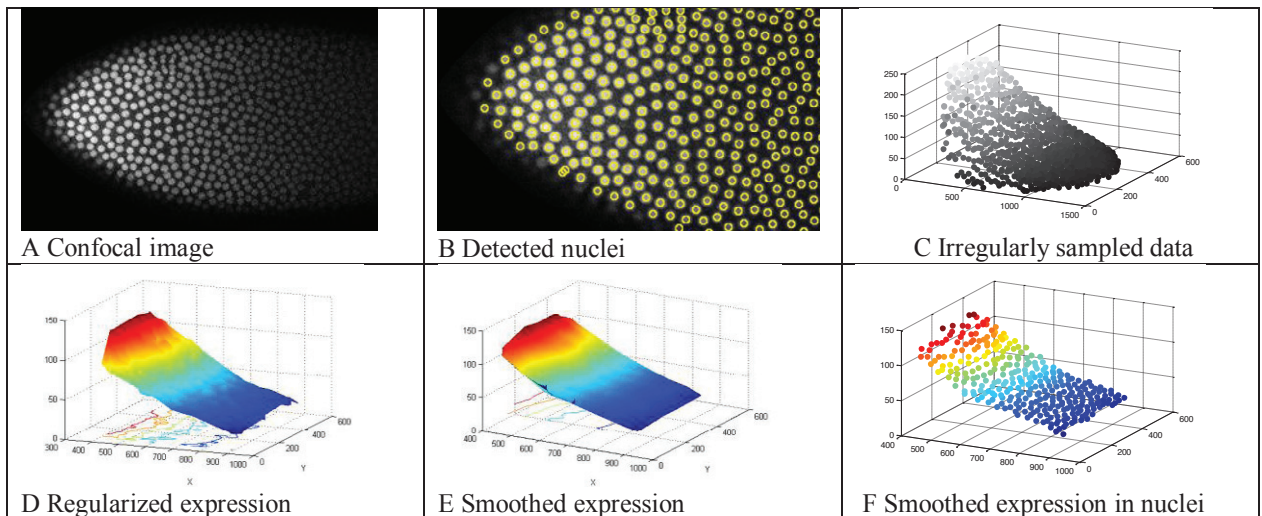


Figure 2. Data processing step results.

### 2.2.1 Segmentation of confocal images

Automatic boundary detection of nuclei can be inaccurate due to varying intensity of nuclei and also due to measurement noise. We propose the use of circular sub-regions that are contained within each nucleus (see Fig 2B). The radius of ROI is chosen manually (in the case of Fig 3, large ROI radius is 7 pixels, small – 4 pixels; visible nuclei radii are 10-12 pixels). The centres of ROI can be defined either manually, or as centroids of nuclei regions detected by a segmentation algorithm.



The centroids of nuclei were found by our original ImageJ plugin (script). The results were then carefully inspected by eye, with the main criterion being whether the centroid corresponded closely enough to the (local) intensity maximum of the nuclear staining (see peaks in Fig 1C). In the vast majority of cases the nuclear intensity looks like a peak with the maximum in the vicinity of the geometric centre of the nucleus. Usually the script finds these correctly, but nuclei with low signal were frequently segmented inappropriately and needed manual correction.

The result of this step are triples  $(x_i, y_i, f_i)$ , where  $(x_i, y_i)$  are the coordinates of centroids and  $f_i$  is the average pixel intensity over the ROI (see Fig 2C).

### 2.2.2 Regularization

The data first is cropped to form a rectangular area  $[x_b, x_u] \times [y_b, y_u]$  with homogeneous placement of nuclear centers. Then the data is regularized to be defined on a rectangular grid with step  $\Delta$ . The values  $(F)_{m,n} = f(x_b + \Delta m, y_b + \Delta n)$  of the regularized image are determined by bilinear interpolation (the function is approximated by a piecewise linear function, defined on triangles given by a Delaunay triangulation of the set of points  $(x_i, y_i)$ ) (see Fig 2D).

The step  $\Delta$  of the regular grid should be small in comparison to inter-nuclear distances (0.2% embryo length was used for all analyses).

### 2.2.3 2D-SSA nonparametric smoothing

Regularized images are smoothed by the 2D-SSA procedure – a novel nonparametric technique of 2D data analysis. 2D-SSA works on the assumption that the data is a sum of slowly varying “trend”, oscillations (texture) and noise. The decomposition of  $F$  into these summands is performed by the SVD (Singular Value Decomposition) of a structured matrix of higher dimension, which is constructed from data  $F$  using a sliding  $L_x \times L_y$  window.

Trend, oscillations and noise correspond to different singular triplets of the SVD (analogous to principal component analysis), and can be grouped. After grouping, a reconstruction of the original summands is performed. Further details on this method can be found in [14] and [23]. The trend usually corresponds to the first  $r$  principal components, and therefore the parameters of the smoothing method are  $(L_x, L_y, r)$  (see Fig 2E).

### 2.2.4 Reconstruction of smoothed function in original points

To determine the values  $\hat{f}_i$  of the smoothed image on the original irregular grid (points of nuclear centres) a bilinear interpolation is used; the values of  $\hat{F}$  are interpolated separately along X and Y axes. For an appropriately small regularization step, the interpolation method is not significant (see Fig 2F).

The noise (the residuals) is calculated as  $\hat{f}_i = f_i - \hat{f}_i$ .

### 2.2.5 Choice of parameters in 2D-SSA

Window sizes must be large enough to capture distinct nuclei. Smaller windows give more detail in the leading principal components of the SSA decomposition (i.e., the pattern is captured by just one or two leading components), while larger windows give better resolution and decompose the trend into a larger set of components. Choosing the number of components can also affect the fit. Trend principal components should correlate weakly with the residual (noise) set, one can determine the number of trend components by examining the w-correlations between components. For more details see [14]. Due to the above considerations, in this study we chose a window size of 33x33.

## 3. Results and Discussion

### 3.1 Expression noise in the nuclear layer

Fixed immunostained embryos in the FlyEx database were imaged tangentially, that is scanned in a side view of the embryo, capturing a 2D expression surface for the protein of interest (Fig 2A). We conducted tangential scans of Bcd-GFP embryos for comparison. The key methodological problem with tangential scans is to identify the nuclear

positions and areas. This is the process of segmentation, to find a mask or ROI (Region of Interest) that covers each nucleus in the image. This is challenging, since nuclear signals are distorted optically. To the best of our knowledge, nobody has systematically studied the effects of fixation and confocal setting on nuclear segmentation (but see [10,11]). Specifically, the problem is that the overall form of a nuclear image depends on the magnitude setting of the microscope (the lower the magnitude the more nuclei look like peaks, with maximum fluorescence at the center of a nuclear image). Therefore, the lower the intensity of nuclear fluorescence, the smaller the nucleus appears (it does not simply become dimmer, it optically shrinks). In addition, the gap between adjacent nuclei can be so small (and the illumination from the core part of the nucleus can be so high) that intensity in the between-nuclei gaps is probably not representative of the true (lower) cytoplasmic fluorescence or background intensity. It is therefore a general question how to measure the intensity of a nuclear signal – e.g. by i) averaging the sum of pixel intensities by the area of nucleus (defined by pixels above background intensity), or ii) by counting intensity from the bright core only. One of the main methodological problems we have found with nuclear-level variability analysis (for both fixed immunostained and live Bcd-GFP embryos) is the influence of the segmentation (nuclear finding) technique on the character of the extracted nucleus-to-nucleus noise.

### 3.2 Fixed immunostained embryos

We picked the best quality cc13 embryo images from the FlyEx database by eye (in contrast with the procedure and parameters used for semi-automatic processing, presented on FlyEx and described in [24]). These were chosen based on the quality of the nuclear images: focus; clearly observable sub-nuclear granular structure; and low background. Only a fraction of the database images fit these criteria. Our selection criteria were constrained by needing to match, as far as possible, fixed and immunostained images with clear live Bcd-GFP images. We found the ms18 embryo, used as the representative exemplar in Figure 1 of Wu et al. [9] not to be the best one according to these criteria. (The ms18 embryo belongs to early cc13 and its nuclei are smaller than for later embryos; as a result, they have more optical distortion.) The bd3 embryo is analyzed here as an exemplar.

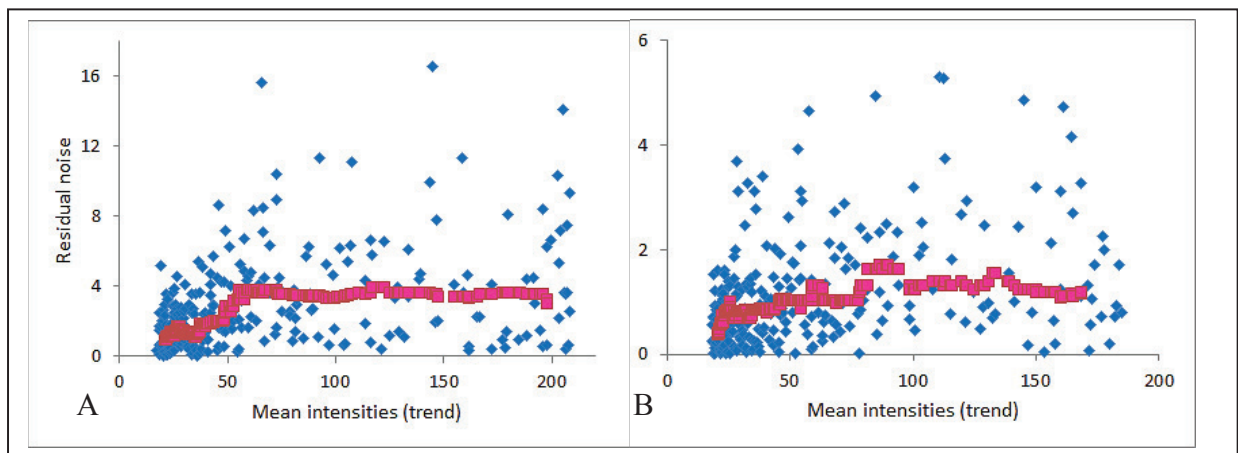


Figure 3. Dependence of the between-nucleus noise on signal at different ROI, tangential data for Bicoid factor, fixed and immunostained embryo bd3 from the FlyEx database ([25]; <http://urchin.spbcas.ru/flyex/>). The between-nucleus noise is largely independent of the signal level (trend). A) Nuclear intensity of Bicoid, measured via small ROI. Median smoothing of absolute values of the noise (y-axis) vs. means (x-axis). Moving medians are in magenta. The between-nucleus noise is independent of signal level, particularly at moderate intensity (value ~60) and above. B) Bicoid nuclear intensity with large ROI. Magenta – as in A); again the between-nucleus noise is largely independent of signal level (in this case for intensities above value ~25).

Since the quality of the nuclear images in tangential scans depends dramatically on magnitude / resolution of the microscope, we varied and optimized both the nuclear segmentation and SSA parameters before proceeding with decomposition into 2D trend and residual noise.

Fig 3 shows the results for a good fixed immunostained image (embryo bd3, Fig 2A). The nuclear centers were

found manually and mean nuclear intensity was calculated for two radius values (small ROI, large ROI). Large ROI is shown in Fig 2B. The results of the noise analysis are shown in Fig 3. The main conclusion is that the between-nucleus noise for the Bcd factor for fixed immunostained embryos is largely independent of the signal level. This characteristic of the noise was reproduced with different image processing parameters (particularly, different ROI radii; Fig 3CD) and different SSA parameters (size of the sliding 2D window, the number of components in the SSA decomposition corresponding to the trend; not shown). As can be seen ( Fig 3), the residual noise with small ROI is much higher than for large ROI.

### 3.3 Live embryos, Bcd-GFP

Two of the most serious problems in quantitatively studying the dynamics of GFP-tagged proteins in live *Drosophila* embryos are i) low signal due to low GFP concentration, and ii) GFP bleaching. Ideally, we would like to adjust the scanning time to collect as many photons per pixel as possible. However, extension of the exposure time bleaches the GFP, and we cannot use anti-fog agents. By careful comparison of the Bcd-GFP dynamics under different scanning protocols, we found confocal settings optimal for our purposes of studying within- and between-embryo noise (see 2.1).

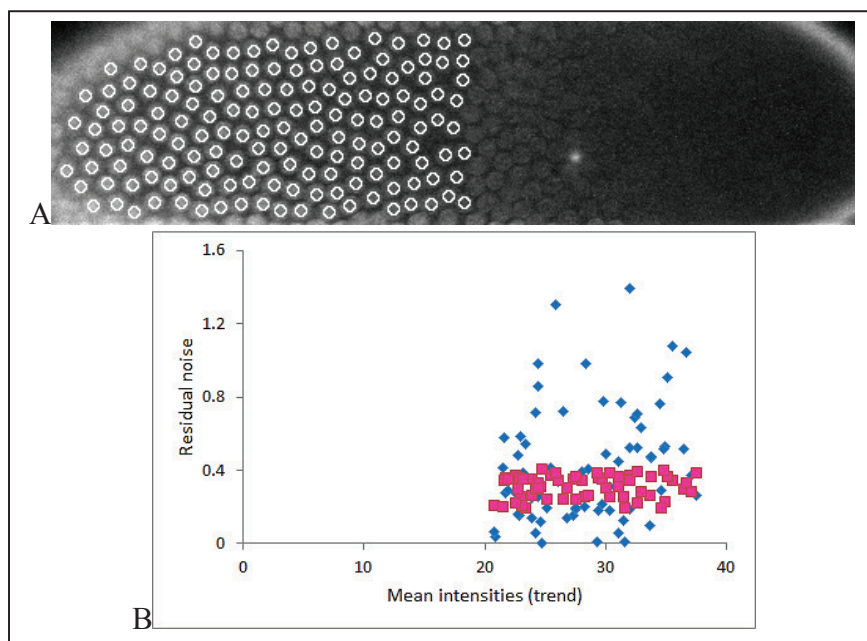


Figure 4. Nucleus-to-nucleus noise for a live embryo, Bcd-GFP signal. A) Live embryo image, scanned 430 times in 10 minutes. Each nucleus was masked by an ROI of 37 pixels (small ROI). B) Dependence of between-nucleus variation on trend. Moving medians are in magenta. The between-nucleus noise is independent of signal level. (Cf Fig 3)

Fig 4A shows a live embryo image which was scanned 430 times in 10 minutes. Each nucleus is masked by a 37 pixel ROI, the small ROI in 3.2 (see Fig 3). For live imaging, the large ROI causes greater optical distortions of the nuclear images. Overall intensity for Bcd-GFP signal in live embryos is lower than with fixed immunostained embryos. Live embryo nuclear intensities are generally between 18 – 48 units on a 0-255 scale, about 5 times lower than for fixed embryos.

Following segmentation, the data from Fig 4A was processed by 2D-SSA, as in 3.2. The results are shown in Fig 4B. The between-nucleus noise is independent of the signal level. The residuals are about 0.5-1 on an intensity scale of 0-255; the noise is quite low. Our procedure for scanning makes GFP signal of comparable noise level to fixed immunostained embryos. The main conclusion of this communication is that both procedures, live GFP and fixed immunostained, display signal independent noise.

### 3.4 Bcd between-nucleus noise: the effect of data processing

Wu et al. [9] previously reported on Bcd nucleus-nucleus variation, using the same dataset (cc13 FlyEx images [24]). They reported a heterogeneity in their results, with signal-independence of Bcd variation in some embryos and signal-dependence in others. Here we discuss the differences in data processing which likely led to the heterogeneous outcome in Wu et al. [9], and how our process produces more robust, consistent results.

First, as discussed above, the segmentation process, to find nuclei, can have a strong effect on acquired nuclear intensities. (These embryos were not stained with a specific nuclear stain; nuclei must be inferred from transcription factor expression [24,25].) Using the Khoros/VisiQuest segmentation package with publicly available workspaces and recommended settings [24 and <http://urchin.spbcas.ru/flyex/>], we found poor quality masks. In Fig 5, we show the Khoros/VisiQuest segmentation for embryo bd3 (shown with ROI segmentation in Fig 2B). (FlyEx does not display the mask images that were used as data in [9].) This procedure tends to be ‘space-filling’, and extend irregular boundaries far beyond the real nuclear boundaries. We have found use of the small ROI’s, smaller than the observable nuclear boundaries, to provide much more consistent data regarding nuclear intensities.

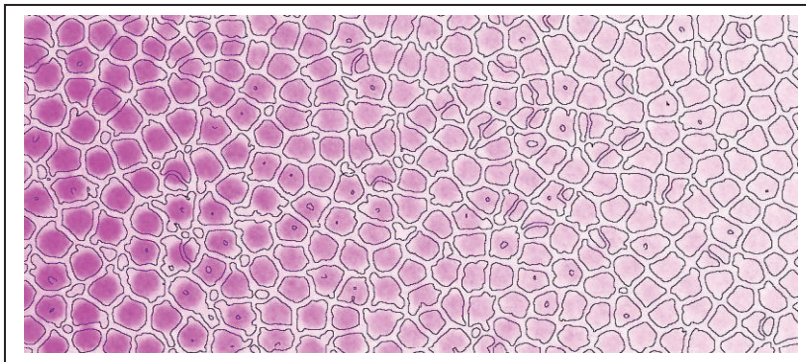


Figure 5. Segmentation (nuclear masking) of the bd3 embryo image by tools and settings recommended in [24; <http://urchin.spbcas.ru/flyex/>]. Bcd immunostaining is shown in magenta, black contours are the nuclear masks found by Khoros/VisiQuest software. This is approximately the same embryo region illustrated in Fig 2B.

Second, the choice of a trend, from which to find the noise (residuals left from subtracting the trend from the data), is critical. Along the anterior-posterior (AP) axis, the Bcd trend is surprisingly exponential looking in central portions of the embryo, and can be described by a simple exponential,  $a \exp(-j/\lambda)$ . Wu et al. [9] calculated residuals as the difference between this exponential and the data. However, a non-parametric decomposition of the data, via 2D-SSA, demonstrates that the full 2D surface of Bcd intensity (AP by dorso-ventral plane) is not a simple exponential and is not smooth – it has low frequency disturbances or texture. 2D-SSA shows that these components statistically belong to the trend (they have a low correlation with the higher frequency noise components of the data; see 2.2.5). Fig 6 shows the 2D Bcd trend extracted via 2D-SSA (same embryo as Fig 3). The non-smooth components of the trend may have a biological basis in the short range correlations between neighbouring nuclei (a component with a correlation length  $5 \pm 1$  nuclei) reported by [19 19].

Wu et al. [9] discussed their results in terms of the number of primary and secondary antibody molecules which bind to proteins in the fixed immunostaining process. They discussed that fluctuations in these numbers could lead to a rescaling noise (between numbers of protein molecules and observed fluorescence) which could produce the signal-dependent noise observed in many of the embryos. This mechanism seems quite reasonable and warrants further study.



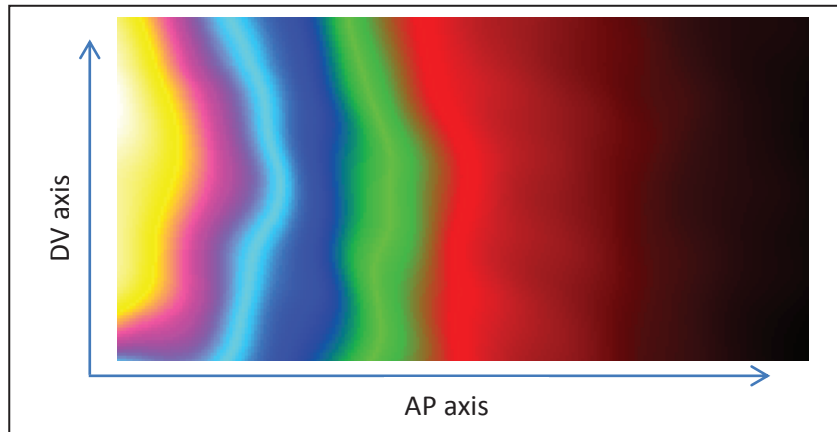


Figure 6. Bcd trend found by 2D SSA processing of the two-dimensional dataset shown in Fig 3. The first three components in the SSA decomposition were chosen to plot the expression surface on the basis of the  $w$ -correlations between components (see 2.2.5 for details; Cf Fig 2E). (AP, anteroposterior; DV, dorsal-ventral.)

However, we have found, in the present communication, that i) nuclear segmentation and ii) determination of the signal trend have strong effects on any conclusions about residual between-nucleus noise. With a more consistent nuclear masking technique and non-parametric determination of the trend, we find that both live and fixed embryos show signal-independent between-nucleus noise. This suggests that the immunochemical amplification of the signal does not change the nature of the noise drastically.

#### 4. Conclusions

- Nucleus-to-nucleus variability (noise) of the Bicoid transcription factor intensity, both for live embryos (Bcd-GFP) and for fixed & immunostained embryos is signal-independent.
- The character of the noise is sensitive to the segmentation method used to find nuclei (masks).
- The technique for finding the signal (expression surface) from raw intensity data affects the residual noise extracted, and affects conclusions regarding the character of the noise.
- The immunochemical amplification of the signal in fixed embryos (due to usage of secondary antibodies) does not drastically change the character of the noise.

#### Acknowledgements

This work was supported by Joint NSF/NIGMS BioMath Program, 1-R01-GM072022 and the National Institutes of Health, 2R56GM072022-06, 2-R01-GM072022.

## References

1. M.B. Elowitz, A.J. Levine, E.D. Siggia, and P.S. Swain, Stochastic gene expression in a single cell, *Science* 297 (2002) 1183–1186.
2. E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman, A. van Oudenaarden, Regulation of noise in the expression of a single gene, *Nature Genetics*, 31 (2002) 69-73.
3. J.M. Raser and E.K. O'Shea, Control of stochasticity in eukaryotic gene expression, *Science* 304 (2004) 1811–1814.
4. A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Noise in protein expression scales with natural protein abundance, *Nat. Genet.* 38 (2006) 636–643.
5. H. Kitano, Biological robustness, *Nature Reviews Genetic* 5 (2004) 826-837.
6. D. Longo and J. Hasty, Dynamics of single-cell gene expression, *Mol. Syst. Biol.* 2 (2006) 64.
7. J.R. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. Derisi, and J.S. Weissman, Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise, *Nature* 441 (2006) 840–846.
8. N. Fedoroff and W. Fontana, Small Numbers of Big Molecules, *Science, New Series* 297 (2002) 1129-1131.
9. Y. Wu, E. Myasnikova, and J. Reinitz, Master equation simulation analysis of immunostained Bicoid morphogen gradient, *BMC Syst. Biol.* 1 (2007) 52.
10. E. Myasnikova, S. Surkova, L. Panok, M. Samsonova and J. Reinitz, Estimation of errors introduced by confocal imaging into the data on segmentation gene expression in *Drosophila*, *Bioinformatics* 25 (2009) 346-352.
11. E. Myasnikova, S. Surkova, G. Stein, A. Pisarev, and M. Samsonova A regression system for estimation of errors introduced by confocal imaging into gene expression data in situ, *BMC Bioinformatics* 12 (2011) 320.
12. L. Zamparo and T.J. Perkins, Statistical lower bounds on protein copy number from fluorescence expression images, *Bioinformatics* 25 (2009) 2670–2676.
13. A.V. Spirov, D.M. Holloway, F.J.P. Lopes, N.E. Golyandina, and E.N. Spirova, Measuring gene expression noise in early *Drosophila* embryos: the highly dynamic compartmentalized micro-environment of the blastoderm is one of the main sources of noise, In: M. Giacobini, L. Vanneschi, and W.S. Bush (Eds.), *EvoBIO 2012, LNCS 7246*, Springer-Verlag Berlin Heidelberg, 2012, pp. 177–188.
14. D.M. Holloway, F.J.P. Lopes, L. da Fontoura Costa, B.A.N. Travençolo, N. Golyandina, K. Usevich, and A.V. Spirov, Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in *Drosophila* segmentation, *PLoS Computational Biology* 7 (2011) e1001069.
15. V.E. Foe and B.M. Alberts, Nuclear and Cytoplasmic Interactions during the Five Mitotic Cycles that Precede Gastrulation in *Drosophila* Embryogenesis, *J. Cell Science* 61 (1983) 31-70.
16. E.M. Lucchetta, M.E. Vincent, and R.F. Ismagilov A Precise Bicoid Gradient Is Nonessential during Cycles 11–13 for Precise Patterning in the *Drosophila* Blastoderm, *PLoS ONE* 3 (2008) e3651.
17. T. Hazelrigg, N. Liu, Y. Hong, and S. Wang, GFP expression in *Drosophila* tissues: time requirements for formation of a fluorescent product, *Dev. Biol.* 199 (1998) 245-249.
18. E. Kiseleva, S. Rutherford, L.M. Cotter, T.D. Allen, and M.W. Goldberg, Steps of nuclear pore complex disassembly and reassembly during mitosis in early *Drosophila* embryos, *J. Cell Sci.*, 114 (2001) 3607–3618.
19. T. Gregor, D.W. Tank, E.F. Wieschaus and W. Bialek, Probing the limits to positional information, *Cell* 130 (2007) 153-164.
20. T. Gregor, E.F. Wieschaus, A.P. McGregor, W. Bialek and D.W. Tank, Stability and nuclear dynamics of the Bicoid morphogen gradient, *Cell* 130 (2007) 141-152.
21. S. MacArthur, X.-Y. Li, J. Li, J.B. Brown, H.C. Chu, L. Zeng, B.P. Grondona, A. Hechmer, L. Simirenko, S.V.E. Keränen, D.W. Knowles, M. Stapleton, P. Bickel, M.D. Biggin, and M.B. Eisen, Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions, *Genome Biology* 10 (2009) R80.
22. J.L. Epps and S. Tanda, The *Drosophila* *semushi* mutation blocks nuclear import of bicoid during embryogenesis, *Curr. Biol.* 8 (1998) 1277-1280.
23. N. Golyandina and K. Usevich 2D-extension of Singular Spectrum Analysis: algorithm and elements of theory. In: V. Olshevsky and E. Tyrtshnikov (Eds.), *Matrix Methods: Theory, Algorithms, Applications*, World Scientific Publishing, Singapore (2010) pp. 450–474.
24. H. Janssens, D. Kosman, C.E. Vanario-Alonso, J. Jaeger, M. Samsonova, and J. Reinitz, A high-throughput method for quantifying gene expression data from early *Drosophila* embryos, *Dev. Genes Evol.* 215 (2005) 374–381.
25. E. Poustelnikova, A. Pisarev, M. Blagov, M. Samsonova, and J. Reinitz, A database for management of gene expression data in situ, *Bioinformatics* 20 (2004) 2212-2221.