Review

Mathematical approaches in studying *bicoid* gene

Zara Ghodsi[1,2], **Hossein Hassani**[3,*] and **Kevin McGhee**[2]

[1] The Statistical Research Centre, Bournemouth University, Bournemouth, BH8 8EB, UK
[2] Translational Genetics Group, Bournemouth University Fern Barrow, Poole, BH125BB, UK
[3] Institute for International Energy Studies, Tehran 1967743 711, Iran
* Correspondence: hassani.stat@gmail.com

It is widely believed that in *Drosophila melanogaster* the pattern of Bicoid protein gradient plays a crucial role in the segmentation stage of embryo development. As a result of its fundamental role, modelling the Bicoid gradient has become increasingly popular for researchers from many different areas of study. The aim of this paper is to bring together the most prominent studies on this maternal gene and discuss how existing techniques for modelling this gradient have evolved over the years.

Keywords: bicoid; *Drosophila melanogaster*; model; segmentation gene

INTRODUCTION

The fundamental question of how a simple fertilized cell develops into a complex multicellular organism has attracted the interest of many biologists over the last three decades. However, the first noteworthy attempt to answer this question was by Driesch in 1891 when he tried to separate two sea urchin blastomeres [1]. Taking Driesch by surprise, each of the separated blastomeres developed to a half-sized blastula rather than a half sea urchin which was the expectation [1]. This result furnished Driesch with the idea of considering a coordinate system which specifies information on a cell's position [1]. Following more extensive research, this initial view by Driesch is now considered as a fundamental step in the discovery of what we know today as morphogen. Today, it is widely accepted that in developmental processes, the pattern of morphogen products play a key role in directing the embryo into a complex multicellular organism.

A prime example of morphogens is *bicoid*[1], which is the first known morphogen identified by Nsslein-Volhard in 1988 [2]. Bcd is a homeodomain transcription factor which plays a crucial role in patterning the head and thorax of *Drosophila* body [2]. There are several studies which proved the important role of this morphogen in developing the anterior structure of *Drosophila* [2–4]. For example, Frohnhfer et al. [5] showed that embryos receiving different doses of Bcd have differently sized anterior structures, Figure 1A, and in the absence of *bcd*, anterior structures of body are replaced with posterior regions.

During the oogenesis *bcd* mRNA which is synthesised in the nurse cells is localised at the anterior end of the egg [4]. After fertilization, translation of the *bcd* begins and consequently Bcd distributes along the Anterior-Posterior (A-P) axis of the egg and forms a concentration gradient of protein [2]. This establishment occurs during the first three hours after fertilisation which *Drosophila* embryo experiences 14 different cleavage cycles [2,6,7]. It is worthy to mention that unlike most of the animal embryos, *Drosophila* develops a syncytial blastoderm of about 6,000 nuclei at the cortical layer. Lack of cell membranes makes Bcd diffusion much easier during this stage [2,8]. After 3 hours, contribution of several genes (among which *bcd* has a significant role) results in developing blastoderm (Figure 1B) [2,8,9].

Moreover, as stated in [2,8] it is believed the

---

[1]The italic lower-case *bcd* represents either the gene or mRNA and Bcd refers to the protein.
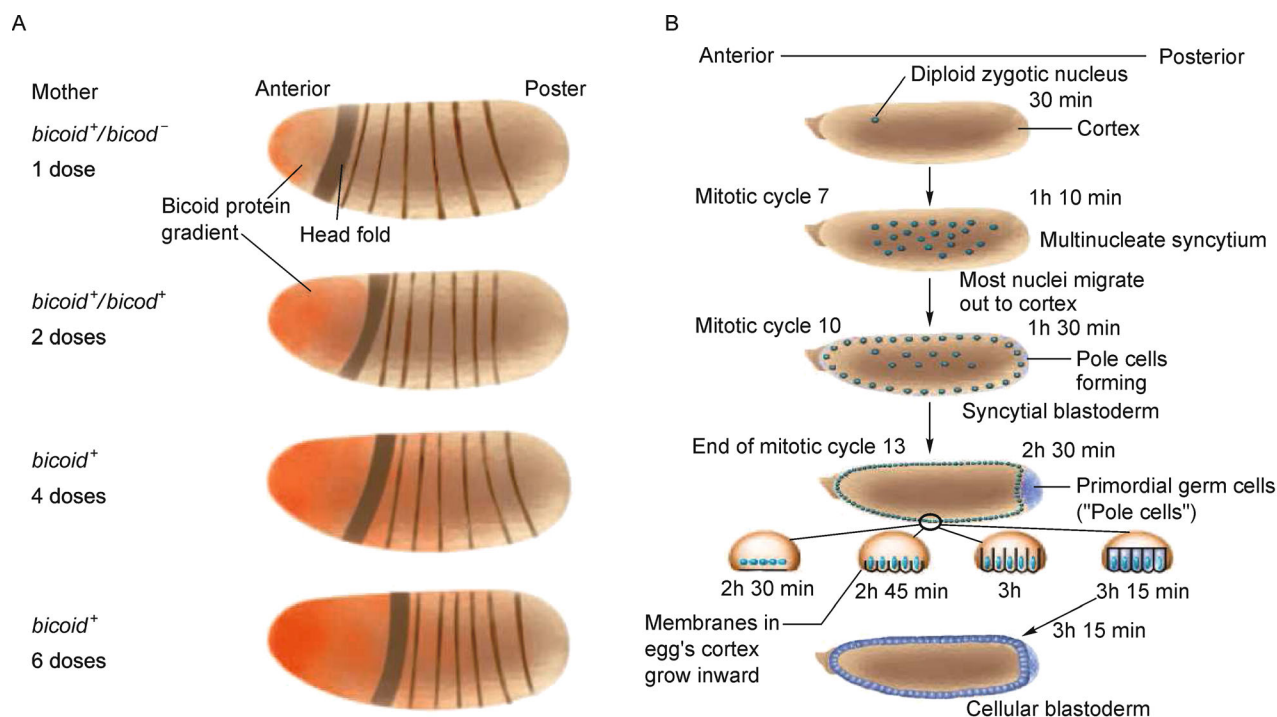
Figure 1. **The BCD protein is a morphogen.** (A) Different doses of *bcd* affects the size of the head in *Drosophila* embryo. (B) Nuclear divisions in 14 cleavage cycles produce a syncytial blastoderm. Figures adapted from [5] with permissoin.

commencement of target gene transcription such as hunchback (*hb*), kruppel (*kr*), and even-skipped (*eve*) is regulated by Bcd gradient in a concentration dependent way.

Since the discovery of morphogens and specially the role of Bcd, several models have been proposed for describing the morphogen gradient and analysing the positional information encoded by Bcd (reviewed in [6,10]). The first few models rely only on the production and diffusion of the morphogen and they are mostly based on reaction-diffusion partial differential equations (PDEs). However, the simplicity of the considered models which limited the performance of the mechanism led researchers of many different areas, especially those from developmental fields to introduce more advanced models.

It is imperative to note that even with all these models there are still some fundamental questions to answer. For example, ambiguities regarding the factors which are involved in achieving Bcd precision and establishing the same Bcd concentration gradient in each embryo. Factors like the amount of *bcd* mRNA to be deposited, its localisation and the amount of protease required for degradation. Moreover, this gradient is due to change by environmental changes such as temperature [8].

The remainder of this paper is organised such that the

section of Models describes the different models proposed for describing Bcd gradient which is followed by a brief introduction on the noise reduction methods used for filtering the Bcd profile in order to obtain the quantitative data in the section of Noise reduction methods. The Database section describes the databases available for studying the developmental processes of *Drosophila*. The paper concludes with a concise summary in the last section.

MODELS

Efforts to introduce a model for Bcd gradient has brought together many researchers of different areas including biologists, physicists, and statisticians. Hence, in different studies several aspects of this gradient including molecular, functional and developmental have been considered as the main axis of the research.

Nevertheless, in this paper we attempt to classify those studies mainly focused on Bcd gradient, understanding the basic models which describe not specifically Bcd but morphogens gradient formation is essential. Coming below is a summarised note on the proposed models and in doing so we mainly followed the attributed published studies and [6,10,11].

In 1902 Morgan published a book entitled regeneration and in that book he suggested that cells's fate in an organism is determined by the gradient of "formative substances" which specifies the pattern formation [12]. Later on through an experiment, Spemann [13] discovered that in a gastrola a secondary body axis can be formed by grafting a group of dorsal cells onto the opposite ventral pole. This study indicated that the gradient responsible for pattern formation is released by a group of localised cells [13] which according to Child's study represents a metabolic function [14].

In 1952, Turing named these chemical substances as morphogens [15]. He also introduced the reaction-diffusion model which shows that different morphogens with slightly different diffusion properties will interact with each other and generate spatial patterns of morphogen concentration [10].

The establishment of morphogen gradient is then described via the French Flag Model [16]. According to this model, morphogen is extracted from a group of source cells to provide positional information and different target genes are expressed above the specific concentration thresholds (Figure 2).

Later on according to Source sink model [17], it was proposed that the morphogen molecules are produced in source cells and destroyed at the other end by "sink" cells. However, nowadays it is known considering sink cells is not compulsory since the morphogen gradient can be formed even if it is not degraded at all [10].

The gradient of Bcd for the first time was visualized by antibody staining in [2]. Using this method and the previous models like Differential Equations, authors proposed the Synthesis Diffusion Degradation model (SDD) which is the most widely accepted model for Bcd. According to this model, Bcd gradient in the embryo follows an exponential curve until it reaches a steady state in which the production and degradation of Bcd molecules are in a balance. Since then, studying this gradient has become popular and many quantitative models of gradient formation have been developed.

Most of the models accounting for Bcd gradient establishment rely on three main concepts: Bcd synthesists at the anterior end of the embryo, diffusion and degradation. Interestingly, what makes these models different from each other is the considered assumptions for different parameters.

Differential equation

This equation is the first mathematical model proposed for describing Bcd gradient in 1985 [6]. Based on this model by assuming the space interval as Δx, the concentration of Bcd molecules over the period of Δt will follow the Equation (2).

$$C(x, t + \Delta t) = C(x,t) - kC(x,t)\Delta t$$

$$+ d\left[-C(x,t) + \frac{1}{2}C(x + \Delta x, t) + \frac{1}{2}C(x - \Delta x, t)\right]$$

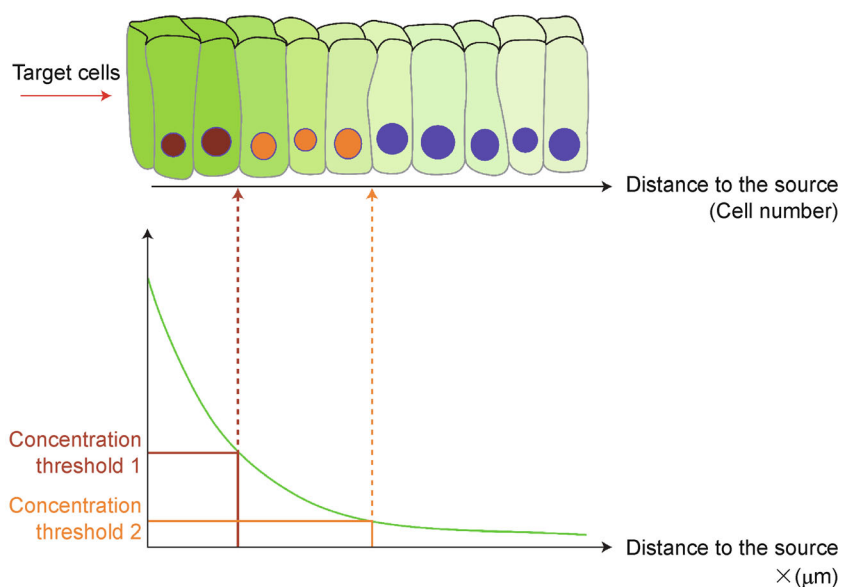$$+ j(x,t)\Delta t. \tag{1}$$

Figure 2. French flag model. Based on this model the concentration is a function of the positions x. The morphogen activates different target genes above different concentration thresholds (brown and orange). Figure adapted from [10] with permission.

where C is the concentration of Bcd [molecules/μm³] and is a function of space and time. It should be noted that for a very small Δt the concentration of Bcd at time $t + \Delta t$ is equal to the concentration at t. Unlike the sink model which considers a localized sink, in this model in order to generate the steady-state gradients morphogens are degraded with a constant degradation rate $k[1/s]$ and the terms inside the bracket describes the Bcd movement which is respectively related to the diffusion, leaving and coming Bcd molecules from the adjacent positions $x \mp \Delta x$. The one-half here shows the equal probability of Bcd movement to the left and right, d is used for scaling and is related to the D which is diffusion coefficient and j is the source function accounts for describing the spatial distribution and the rate of Bcd production [6].

Accordingly, the rate of alteration in Bcd conccentration can also be obtained by reordering Equation (2):

$$\frac{C(x,\ t + \Delta t) - C(x,\ t)}{\Delta t}$$
$$= -kC(x,t)$$
$$+ \frac{d(\Delta x)^2}{2\Delta t} \frac{-2C(x,\ t) + C(x + \Delta x,\ t) + C(x - \Delta x,\ t)}{(\Delta x)^2}$$
$$+ j(x,t), \tag{2}$$

by assuming the limit $\Delta t, \Delta x \to 0$ equal to a finite value when $d(\Delta x)^2/2\Delta t \equiv D$, Differential equation can be obtained.

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} - kC + j. \tag{3}$$

From now for the simplicity we can show $C(x, t)$ as C.

It is worthy to mention that when using this model different assumptions should be taken into account and also the parameters used in this model are not tractable thorough analytical solutions since the parameters in

Equation (3) are the representor of the macroscopic properties of the embryo which sometimes are not a good representative of Bcd molecules (for more information, see [6,10,17]).

Synthesis diffusion degradation model

As mentioned earlier this model was proposed by Nsslein-Volhard and is widely known as the SDD model. It should be noted that $j(x, t)$ for any other space rather source is equal to zero so having diffusion and degradation together, the diffusion equation with linear degradation can be achieved and the changes in Bcd concentration C over time t can be written as:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} - kC. \tag{4}$$

Now by considering the fact that in the steady state, $\frac{\partial C}{\partial t}$ = 0 diffusion equation with linear degradation can be written as:

$$\frac{\partial^2 C}{\partial x^2} = \frac{k}{D}C. \tag{5}$$

Seeking to find a solution for the second derivative of C either an exponential functions or sums of hyperbolic sines and cosines can be considered since the latter case can also be written as exponentials, the general solution for C follows:

$$C = C_0 e^{-\frac{x}{\lambda}}, \tag{6}$$

where λ is a constant value. Assuming $C = C_0$ at $x = 0$, therefore C_0 is the source concentration and at $x = \lambda$, C will be equal to $C_0 e^{-1}$, thus λ is the distance to the source at which the concentration has dropped to $1/e$ of the maximal value at the anterior (at $x = 0$) (Figure 3A). It
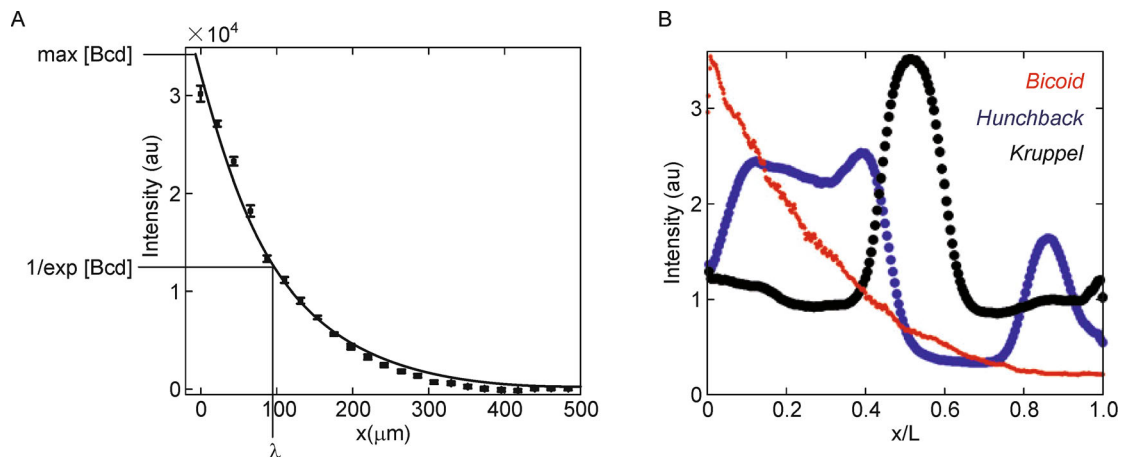

Figure 3. Bcd gradient along the A-P axis. (A) Calculating the length constant [6] λ. (B) Expression the different target genes based on Bcd gradient (find the detailed explanation in text). Figures adapted from [6] with permission.

should be noted that the Equation (6) is only true when the limit of $L \gg \lambda$ and $\lambda = \sqrt{\dfrac{D}{k}}$. This exponential decaying gradient is referred to as the SDD model [2,8,9,18]. This model can also be written as:

$$C = \frac{C_0}{k\lambda} e^{-\frac{x}{\lambda}} \qquad (7)$$

Based on this model the activation of gap genes such as *hb* and *kr* is regulated by the Bcd gradient and their transcription will begin at different concentration of Bcd [2].

In 2005 the SDD model was tested by injecting Dextran in living embryos and this resulted in added support for the SSD model [9]. In that study the relation between changing the egg size during evolution and gene expression pattern scaling was also investigated and the authors showed that diffusion constants are essentially the same in closely related dipteran species with embryos of very different sizes. However, the pattern of gene expression in different species scales with egg length and is dependent on various Bcd lifetime in different species [9].

Among the quantitative approaches Gregor et al. experiment in 2007 [19] on nuclear dynamics and stability of the Bcd gradient must be addressed to achieve a successful study. The issues outlined by Gregor et al. is mentioned below in detail and discussed further using the experiments achieved by other similar studies.

According to [9,19] Bcd gradient is established within the first hour after fertilization and it will not change during blastodermal stages (by 10% precision). However, Bcd is a DNA-binding protein and the nuclear Bcd concentration has an up and down pattern during the mitosis caused by forming and deforming the nuclear membrane [19]. These observations raised the question on how the stable Bcd gradient is established in spite of the repeated dynamical structural change of embryo which happens after mitoses of nuclei.

The important part of this study is measuring the diffusion coefficient of Bcd using fluorescence recovery after photobleaching method which these coefficients turned to be $D \sim 0.3\ \mu m^2/s$ and $\lambda \sim 100\ \mu m$. Therefore, the author concluded that using the SDD model, the time needed for achieving the steady state concentration profile (> 9 h) is much greater than the syncitial embryonic development time (~ 2 h) and protein lifetime τ, and also the length constant λ is much smaller than the length of the embryo [19]. Considering this slow Bcd diffusion how can the gradient be formed so quickly after fertilization?

Taking the nuclear degradation into account, Gregor et al. [19] proposed a variant of the SDD model and suggested that this gradient is established by the intra-nuclear degradation of protein.

In that model, transporting the molecules into the nucleus depends on the Bcd concentration in the cytoplasm ($k_{in}C_{out}$). However the molecules which disappear from nucleus are both transported back into the cytoplasm (at k_{out}) and degradation (at k_{d-nuc}) are accounted for. This dynamic for $n(t)$ nucleus can be written as:

$$\frac{\partial n(t)}{\partial t} = k_{in}C_{out} - \frac{1}{\tau_n} n(t), \qquad (8)$$

$$\frac{1}{\tau_n} = k_{out} + k_{d-nuc}, \qquad (9)$$

where τ_n is the life time related to the Bcd molecules in the nucleus. As cytoplasmic concentration approximated to be constant, hence we have:

$$n(t) = n_{\infty} - \Delta n_0 e^{(-t/\tau_n)}, \qquad (10)$$

where Δn_0 is the molecules reduction, and $n_{\infty} = k_{in}\tau_n C_{out}$ is the number of molecules at nucleus steady state. Following this equation τ_n is approximated to be around 1 or 2 minutes. Considering this equation along with the physical factors of the nucleus Equation (11) can be achieved:

$$\frac{C_{in}}{C_{out}} = \frac{3D\tau_n}{r_n^2}. \qquad (11)$$

where r_n is the nucleus radius.

Gregor et al. [19] also suggested three possible reasons for the paradox seen between the slow obtained diffusion constant and rapid formed steady state; either diffusion coefficient is due to change in different temporal and spatial conditions or nuclear contribution function and active transportation should be taken into account.

Seeking to find an answer to the question above, several explanations were then proposed in different studies:

Pre-steady-state decoding

Based on the previous studies, it was generally assumed that cell fates are induced only after Bcd steady state. However, in 2007 the pre-steady state decoding hypothesis raised by Bergmann et al. stated that Bcd gradient is decoded before reaching a steady state [18]. Based on their experiment, changes in Bcd dosage will induce small shifts in gap and pair-rule gene expression domains, which can be concluded that the observed shifts are related to the gap genes positions to the source because those genes at the posterior part found to be less sensitive to Bcd changes which is in consistent with steady state assumption. Accordingly this model is able to explain small shifts in gap and pair-rule gene expression domains [18]. Based on this model solution attributed to the

Equation (3) is time independent and Bcd concentration changes with time:

$$C = \frac{C_0}{k\lambda}\left[e^{-\frac{x}{\lambda}} - \frac{e^{-\frac{x}{\lambda}}}{2} erfc\left(\frac{2Dt/\lambda - x}{\sqrt{4Dt}}\right)\right.$$

$$\left. - \frac{e^{\frac{x}{\lambda}}}{2} erfc\left(\frac{2Dt/\lambda - x}{\sqrt{4Dt}}\right)\right]. \tag{12}$$

Based on this model the λ is not constant and is due to change by both degradation and diffusion time.

Nuclear trapping

The idea of nuclear trapping by Coppey proposed a substitute for degradation of the Bcd [10,20] and can explain the temporarily stable Bcd gradients in the absence of degradation. Based on this idea nuclear divisions increase the nuclear density in the syncytium, as well as shuttling the Bcd molecules in and out of nuclei at each cleavage cycle which in turn decreases the effective diffusivity of Bcd. Hence, the nuclear Bcd concentration will be stable over several cleavage cycles, although the total concentration continuously increases. In this model the stable Bcd concentration in the nuclei will be read out by target genes.

$$C_{(tot)} = \frac{C_0}{D}\left[\frac{\lambda(t)}{\sqrt{\pi}} e^{-\frac{x^2}{\lambda(t)^2}} - x\cdot erfc\left(\frac{x}{\lambda(t)}\right)\right], \tag{13}$$

bcd mRNA based approaches

As discussed before based on [21,22] *bcd* mRNA is transcribed in the mother and localised at the anterior end of the egg and encodes Bcd. These molecules then by diffusion establish a gradient along A-P axis. However, Surdej and Jacobs-Lorena in 1998 suggested that the stability of maternal *bcd* mRNA is under a systematically regulation [23] and following this experiment a completely different approach for describing Bcd gradient establishment proposed by Spirov et al. states that Bcd gradient is formed mainly by transporting *bcd* mRNA along the cortex rather than by Bcd diffusion. Using FISH method and confocal microscopy authors showed that the gradient related to *bcd* mRNA and Bcd are virtually identical at all time classes. Moreover, the proposed Active RNA Transport and Synthesis (ARTS) model could successfully explain small measured diffusion coefficient by Gregor et al.

According to ARTS model Berezhkovskii et al. stated that since there is no production of mRNA at the anterior part and all the *bcd* is provided by the mother, the *bcd* diffusion will follow a Gaussian rather than exponential distribution. Taking this distribution into account the source function of Equation (3) can be rewritten as:

$$j = \frac{C_0}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \tag{14}$$

where σ is the length constant attributed to the mRNA Gaussian distribution (i.e., $\lambda_{RNA} \equiv \sigma$), now another solution for the Equation (3) can be obtained:

$$C(x) = \frac{C_0}{2k\lambda} e^{\frac{\sigma^2}{2\lambda^2}}\left(e^{\frac{x}{\lambda}} erfc\left(\frac{\sigma^2 + \lambda x}{\sqrt{2}\lambda\sigma}\right) + e^{-\frac{x}{\lambda}} erfc\left(\frac{\sigma^2 - \lambda x}{\sqrt{2}\lambda\sigma}\right)\right) \tag{15}$$

Following this solution Bcd gradient is not just a simple exponential curve and both λ and σ should be considered in gradient formation model.

According to these findings another computational model proposed by Dilao and Muraro [24] which considers that *bcd* mRNA diffuses along the A-P axis and Bcd protein production happens in the ribosomes localized near the syncytial nuclei. In that model a scaling relation has been considered between *bcd* diffusion coefficient, degradation rate and embryo length and the steady state obtained when the *bcd* translation is completed. Hence, the idea of Bcd degradation is not a necessary step in this model.

However, in 2011 Little et al. by using a novel method of fluorescent *in situ* hybridization found that approximately 90% of all *bcd* mRNA is still localised within the anterior 20% of the embryo. Hence the result is in consistent with the Dilao and Muraro work because the mRNA spatial distribution is not sufficient to provide the Bcd protein gradient specially at the posterior end of the egg therefore either the active or passive Bcd movement is necessary for gradient formation [25].

The precision and reproducibility of positional information of Bcd were also studied by Gregor et al. in [26] where a 10% concentration difference was reported to be necessary to distinguish individual nuclei from their neighbours. However, the absolute concentration of Bcd molecules at half-embryo length, where Bcd targets Hb, found to be too low to be distinguished by nuclei [26]. Thus, the author attempted to quantify three other measures of precision; limits set by the random arrival of Bcd molecules at their targets, input-output relationship between Bcd and Hb and reproducibility in the spatial profile of Bcd from embryo to embryo. Using a combination of different experiments, all of the quantities mentioned above found to be 10% [26]. This quantification suggests that mechanisms other than mere noise reduction should be involved to achieve a 10% accuracy and it was concluded that the pattern formation in embryo is not faced with noisy input signals and variable readouts, but is instead precisely and reproducibly controlled [26].

To understand the mechanisms that control this extreme precision and reproducibility, several studies were conducted by other researchers [27–29]. For example, a fundamental dynamic mechanism in which transcription apparatus coordinates the reliable transcriptional responses to cellular signals were introduced in [27]. Using stochastic simulations in that study, Wang et al. challenged the traditional model for gene regulation system, which suggests that activators regulate gene transcription through recruiting proteins such as general transcription factors and RNA polymerase II (Pol II), and proposed instead that this system is through controlling circumstance where Pol IIs reinitiate transcription that activators mediate the responses to the upstream signals [27].

It is imperative to note that although the approaches employing statistics and probability theories are considered to be powerful in this field, still an experimentally elucidation based on the dynamic mechanisms of molecular machines is needed to be developed. The stochastic approaches conducted to explain the dynamics of Bcd gradient are discussed in more details in the following section.

Stochastic models

The deterministic models presented so far have limitations when the uncertainties analysis is required. Thus, a number of researchers attempted to study a discrete chemical reaction system, based on stochastic simulation to provide a more detailed understanding of the chemical reactions in morphological systems [11,30–33].

According to [11], the procedure of stochastic simulation is to answer two fundamental questions: "which reaction occurs next and when does it occur?" and is considered as a new successful method for simulating evolution at the genomic level. However, when the system contains many reactions, the stochastic simulation can be computationally expensive [11].

In order to understand the dynamics of Bcd gradient formation at the molecular number level and identify the source of the nucleus-to-nucleus expression variation Wu et al. formulated a chemical master equation by considering the embryo as a finite number of sub volumes and presented stochastic simulations for molecules undergoing transitions between those compartments [34]. The simulations then obtained using publicly available software MesoRD [35]. Another stochastic approach can be found in [36] where a bayesian inference approach presented in order to solve both the parameter and state estimation problem for stochastic reaction-diffusion systems.

To address that how Bcd gradient is generated and read out precisely to distribute Hb with a small embryo-to-
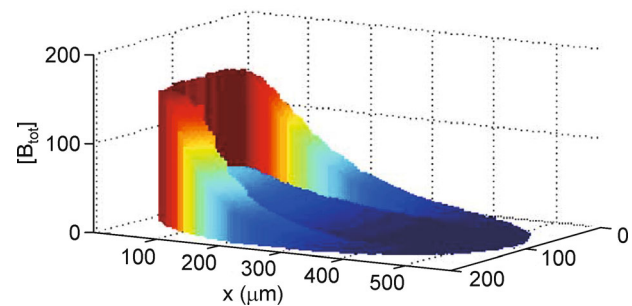

Figure 4. **A simulated embryo at cleavage cycle 14 showing the local total Bcd concentration [B$_{tot}$] (arbitrary units).** Figure adapted from [51] with permission.

embryo fluctuation, Okabe-Oho et al. developed one- and three-dimensional stochastic models in which in the three-dimensional model the effects of dynamically changing nuclei are examined by considering a cylinder of length L ≈ 490 m extending along the AP axis of the embryo [37].

Following the obtained results by these two models, it is reconfirmed that the stable profile of Bcd is established by the stochastic processes of synthesis, diffusion and degradation as well as the rapid movements of the Bcd molecules to nuclei [37]. Moreover, the authors claimed that the random arrival of Bcd at the *hb* enhancer causes the fluctuations in *hb* gene profile [37].

In another model presented in [38], two forms of Bcd molecules were considered in the model; free-diffusing molecules, Bfree, that are in the cytoplasm, and immobile molecules, Bbound, that are bound to low-affinity DNA sites inside the nuclei. According to this two dimensional stochastic model, Bbound plays a significant role in formation of the Bcd gradient [38] confirming the results of [26] in which showed that the random diffusion of Bcd and Hb are dominant sources to limit the precision of readout (Figure 4).

The quantitative and theoretical analysis of Bcd gradient allowed to focus attention on several key issues: the generation mechanism of Bcd profile, the readout system of this gradient for downstream genes and the important parameters that needed to be measured. However, still a simpler quantitative model with more molecular and cell biological depth is encouraged to be developed. Such a less detailed mechanism, in which avoids using too many equations and parameters gives a more reliable image of *Drosophila* morphogene system. Furthermore, the fundamental questions of how morphogens and their receptors are moving inside cells and at the extracellular matrix and how precision and robustness of developmental control is achieved at the molecular number level throughout development are yet to be addressed.

NOISE REDUCTION METHODS

As discussed earlier, accurate quantitative data has a huge importance for prosperity of a developmental study. To extract the quantitative data from the confocal images several steps should be taken into account which among them removing noise is a crucial step since the efficiency of the proposed method largely depends on the accuracy of the data used as the source. In this section we introduce two different procedures which have been used so far for filtering Bcd profile.

The noise removal stage applied in FlyEx database mainly relies on removing noise by approximating the signal using symmetrical 2D parabloid plus a linear mapping of intensity. This process applied for most segmentation genes [39]. However, this procedure is not applicable for Bcd because the size of the Bcd nonexpressing areas needed for applying this method procedure is not large enough, so as an exception, removing noise for this gene is simply achieved by fitting the following exponential model plus a constant value [40]:

$$C_0 e^{-\frac{x}{\lambda}} + B. \tag{16}$$

where the constant value of B is added to make the further correction.

Another useful method for filtering Bcd and extracting the signal is Singular Spectrum analysis (SSA). SSA is a powerful nonparametric method for time series analysis and forecasting [41] and was introduced for reducing noise in Bcd in 2006 [42]. Although SSA is normally used for analysing time series data, for the data traced in space like Bcd data, different intensities along the A-P axis can be defined as the time reported data. For example, considering a set of intensities related to the A-P axis of a given time class as $Y_N = (y_1, ..., y_N)$ with length of N, the main purpose of SSA is to decompose this series into a sum of a small number of series, so that each subseries can be identified as either a trend, periodic or quasi-periodic component or noise. This is followed by a reconstruction of the original series.

Unlike the previous method, SSA does not require any assumption regarding the gene expression pattern and its applicable for all the segmentation genes. It also does not rely on any assumptions about the data or residuals. These features make SSA as a valuable method for signal extraction and since then it has been used in several other studies, see for example [42–47].

Even though the signal extraction by SSA appears to be simple, in practice it is a complicated task since for some cases the trend cannot be separated from noise or cyclic components just by choosing the first eigenvalue [48]. Hence, in addition to the basic SSA approach, two theoretical developments of this method have also been used for Bcd noise reduction: two dimensional SSA (2D-SSA) and SSA based on minimum variance. The data points used in 2D-SSA study attributes to the intensity levels for the positions along both AP and Dorsoventral (DV) axis and are considered as a sequenced series [49,50].

Recently, SSA based on minimum variance was examined for filtering and extracting the *bcd* gene expression signal [51] and the results illustrated that SSA based on minimum variance can significantly outperform the previous methods used for filtering noisy Bcd [51]. An important contribution from this work is shown in Figure 5 which depicts the normal distribution of the estimated parameters A and λ before and after filtering. As it appears, parameters show different values in noisy and noise-free series, which confirms that noise reduction does help to obtain a better model. Thus, if the user is inclined to apply a parametric model such as SDD, it is strongly suggested that they first reduce the noise with any noise reduction method (here for example SSA) and then fit a model to the noise reduced Bcd, or they may simply apply SSA, which is recognized for both noise reduction and trend extraction.

It is imperative to note that even though achieving a clear signal using SSA, the extracted signal profiles do not follow the expression data satisfactorily when the data series changes sharply. Therefore, in order to solve this issue and capture the peaks of the profiles using sequential SSA is recommended. Such a recursive SSA application
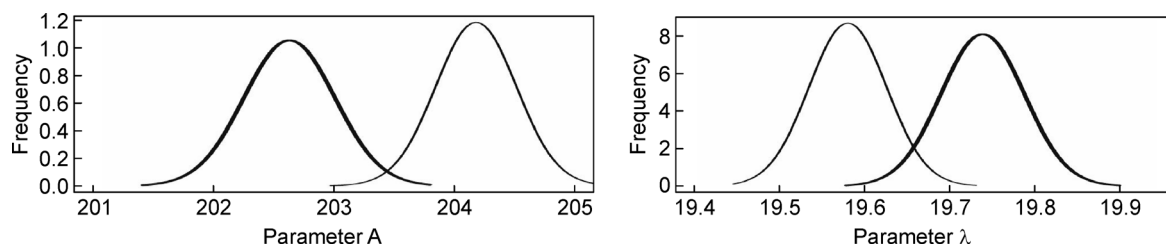
Figure 5. **Distribution of the estimated parameters of *A* and λ for noisy *bcd* and noise-reduced *bcd* (thick line) the actual values of *A* and λ are equal to 200 and 20 respectively.** Figure adapted from [51] with permission.

produces a gradual extraction of the signal present in the noise. Moreover, since identifying the number of eigenvalues needed for signal reconstruction plays a significant role in extracting signal using SSA, developing an algorithm which optimises signal and noise separation by identifying the number of eigenvalues needed separately for each gene is essential.

DATABASE

Along with increasing number of databases in genetics, availability of a source which contains accurate information and quantitative data specified to *Drosophila* plays a significant role in developing new models for analysing Bcd gradient and the regulatory network between several segmentation genes.

There are several databases which provide comprehensive information on genomic and developmental process in *Drosophila*, like FlyBase [52,53], Flymove [54] and FlyEx [39,55], BDGP [56], DroID [57], FlyTF [58], FlyMine [59], FlyAtlas [60], FlyCircuit [61]. Each of these databases provide different aspects of information needed for understanding the *Drosophila* development. Since FlyEx is the most popular database for providing quantitative data, we will briefly touch upon the methodology used for gathering this quantitative data.

FlyEx

FlyEx [62] is a large database on *Drosophila* segmentation genes expression in cycles 11-13 and all temporal classes of cycle 14A. This database contains 4,716 images of 14 segmentation gene expression patterns obtained from 1,579 embryos and 9,500,000 quantitative data records.

FlyEx is designed as a spatiotemporal atlas on the following 14 maternal coordinate genes: bicoid (*bcd*) and caudal (*cad*), gap genes Kruppel (*Kr*), knirps (*kni*), giant (*gt*), hunchback (*hb*) and tailless (*tll*), pair-rule genes even-skipped (*eve*), fushi-tarazu (*ftz*), hairy (*h*), runt (*run*), odd-skipped (*odd*), paired (*prd*) and sloppypaired (*slp*).

The quantitative gene expression data was also obtained from 1,580 wild-type (OregonR) *Drosophila melanogaster* fixed embryos immunostained for segmentation proteins (Figure 6).

A short description on the methodology used for gathering the Bcd data in FlyEx is given below, and in doing so we mainly follow [39,55] where a more detailed description is made available.

Confocal imaging: The images of gene expression patterns are obtained by confocal scanning microscopy. As described in [63], gene expression was measured by using fluorescence tagged antibodies technique, and the final image of a single gene expression pattern in the
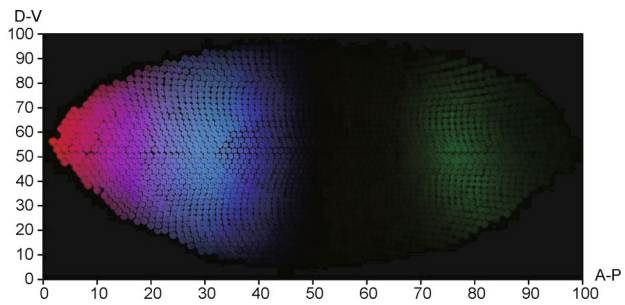

Figure 6. An example of a gene expression image from Flyex. Different colors of pink, blue and green are respectively related to *bcd*, *Hb* and *Gt*. Figure adapted from [62] with permission.

embryo is achieved by averaging, cropping and rotating the obtained image.

Image segmentation: This step can be considered as the compensation method for the size differences among embryos. At a given time each embryo is scanned for the expression of three segmentation genes and this process continues by combining three embryo images and the resultant image is then segmented to make a binary nuclear mask. The advantage of using the binary nuclear mask lies in determining the average x and y related to each nucleus in order to estimate the average fluorescence level of each three gene expressions. A description on the characteristics of the mask can be found in [64].

The following two steps are applied in order to improve the quality and characterize the obtained quantitative data for each single embryo.

Data normalization: In this step, the background noise is minimized by re-scaling the *bcd* expression data via fitting an exponential curve [40]. The reason behind selecting the exponential curve was further explained in the noise reduction section.

Temporal characterization: In order to get a better estimation of quantitative data, determining the developmental age of each embryo and temporal dynamics of gene expression data is mandatory. To this end, for the embryos prior to cycle 14A the calculation of nuclei number are performed. Since cleavage cycle 14A is about 50 min long, for staging embryos other morphological markers such as examination of the blastoderm morphology and the membrane/cortex ratio are also applied.

Registration: In this step the small size variation among embryos within a specific temporal class will be removed. As discussed in [65,66] there are two registration methods, the spline (SpA) and or wavelet method (FRDWT). They mainly differ in the procedure applied to extract the ground control points. The data obtained by this step is called integrated data.

CONCLUSION

In this review which is focussed on the Bcd and the related gradient, we have initially outlined a brief history on identifying Bcd's role as well as a general concept regarding this maternal gene. Thereafter we endeavour to bring together and discuss different mathematical models which have been exploited for describing Bcd gradient. As can be seen most of the theoretical analysis are based on the three main concepts: Bcd synthesists at the anterior end of the embryo, diffusion and degradation, and the main difference is between parameters measured by different models. The review also unearthed that recent models mostly have taken *bcd* mRNA into account. However, what is interesting here is that even after three decades none of the proposed models are fully consistent with all characteristics of the Bcd gradient and it is clear that most models are proposed to give an answer to several questions raised by previous models.

COMPLIANCE WITH ETHICS GUIDELINES

REFERENCES

1. Driesch, H. (1908). The science and philosophy of the organism. Aberdeen: Aberdeen University

2. Driever, W. and Nusslein-Volhard, C. (1988) The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. Cell, 54, 95104

3. Frohnhöfer, H. G. and Nüsslein-Volhard, C. (1986) Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. Nature, 324, 120–125

4. Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., Nüsslein-Volhard, C. (1988) The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. EMBO J., 7, 1749

5. http;//highered.mheducation.com

6. Grimm, O., Coppey, M. and Wieschaus, E. (2010) Modelling the Bicoid gradient. Development, 137, 2253–2264

7. Grimm, O. and Wieschaus, E. (2010) The Bicoid gradient is shaped independently of nuclei. Development, 137, 2857–2862

8. Houchmandzadeh, B., Wieschaus, E. and Leibler, S. (2002) Establishment of developmental precision and proportions in the early *Drosophila* embryo. Nature, 415, 798–802

9. Gregor, T., Bialek, W., de Ruyter van Steveninck, R. R. and Tank, D. W., Wieschaus. E.F.(2005). Diffusion and scaling during early embryonic pattern formation. Proc. Natl. Acad. Sci. USA., 102, 18402–18407

10. Wartlick, O., Kicheva, A. and Gonzlez-Gaitn, M. (2009) Morphogen gradient formation. Cold Spring Harb. Perspect. Biol., 1, a001255

11. Liu, W. (2013). Machine learning approaches to modelling bicoid morphogen in *Drosophila melanogaster*. Ph. D. Thesis, University of Southampton.

12. Morgan, T. H. (1901). Regeneration. New York: The Macmillan Company

13. Spemann, H. and Mangold, H. (1924) Introduction of embryonic primordia by implantation of organizers from a different species. Rouxs Arch. Entw. Mech, 100, 599638

14. Child, C. M. (1941). Patterns and Problems of Development. Chicago: The university of chicago press

15. Turing, A. M. (1952) The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B Biol. Sci., 237, 37–72

16. Wolpert, L. (1969) Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol., 25, 1–47

17. Crick, F. (1970) Diffusion in embryogenesis. Nature, 225, 420–422

18. Bergmann, S., Sandler, O., Sberro, H., Shnider, S., Schejter, E., Shilo, B. Z. and Barkai, N. (2007) Pre-steady-state decoding of the Bicoid morphogen gradient. PLoS Biol., 5, e46

19. Gregor, T., Tank, D. W., Wieschaus, E. F. and Bialek, W. (2007) Stability and nuclear dynamics of the bicoid morphogen gradient. Cell, 130, 153–164

20. Coppey, M., Berezhkovskii, A. M., Kim, Y., Boettiger, A.N. and Shvartsman, S. Y. (2007) Modeling the bicoid gradient: diffusion and reversible nuclear trapping of a stable protein. Dev. Biol., 312, 623–630

21. St Johnston, D., Driever, W., Berleth, T., Richstein, S. and Nusslein-Volhard, C. (1989) Multiple steps in the localization of bicoid RNA to the anterior pole of the *Drosophila* oocyte. Development, 107, 13

22. Ephrussi, A. and Johnston, D. S. (2004) Seeing is believing: the bicoid morphogen gradient matures. Cell, 116, 143–152

23. Surdej, P. and Jacobs-Lorena, M. (1989) Developmental regulation of bicoid mRNA stability is mediated by the first 43 nucleotides of the 3 untranslated region. Mol. Cell. Biol., 18, 28922900

24. Dilão, R. and Muraro, D. (2010) mRNA diffusion explains protein gradients in *Drosophila* early development. J. Theor. Biol., 264, 847–853

25. Little, S. C., Tkačik, G.Kneeland, T. B.Wieschaus, E. F., and Gregor T. (2011) The formation of the bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. PLoS Biol., 9, e1000596

26. Gregor, T., Tank, D. W., Wieschaus, E. F. and Bialek, W. (2007) Probing the limits to positional information. Cell, 130, 153–164

27. Wang, Y., Liu, F. and Wang, W. (2012) Dynamic mechanism for the tran-scription apparatus orchestrating reliable responses to activators. Sci. Rep., 2422,

28. Crauk, O. and Dostatni, N. (2005) Bicoid determines sharp and precise target gene expression in the *Drosophila* embryo. Curr. Biol., 15, 1888–1898

29. Lewis, J. (2008) From signals to patterns: space, time, and mathematics in developmental biology. Science, 322, 399–403

30. Gibson, M. A. and Bruck, J. (2009) Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem., 104, 18761889

31. Andrews, S. S. and Bray, D. (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Phys. Biol.,1, 137–151

32. Hattne, J., Fange, D. and Elf, J. (2005) Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics, 21, 2923–2924

33. Erban, R. and Chapman, S. J. (2009) Stochastic modelling of reaction diffusion processes: Algorithms for bimolecular reactions. Phys. Biol., 6, 046001

34. Wu, Y. F., Myasnikova, E. and Reinitz, J. (2007) Master equation simulation analysis of immune stained Bicoid morphogen gradient. BMC Syst. Biol., 1, 52

35. Hattne, J., Fange, D. and Elf, D. (2005) Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics, 21, 2923–2924

36. Dewar, M. A., Kadirkamanathan, V., Opper, M. and Sanguinetti, G. (2010) Parameter estimation and inference for stochastic reaction-diffusion systems: application to morphogenesis in *D. melanogaster* BMC Syst. Biol., 4, 21

37. Okabe-Oho, Y., Murakami, H., Oho, S. and Sasai, M. (2009) Stable, precise, and reproducible patterning of bicoid and hunchback molecules in the early *Drosophila* embryo. PLoS Comput. Biol., 5, e1000486

38. Deng, J., Wang, W., Lu, L. J. and Ma, J. (2010) A two-dimensional simulation model of the Bicoid gradient in *Drosophila*. PLoS One, 5, e10275

39. Pisarev, A., Poustelnikova, E., Samsonova, M. and Reinitz, J. (2009) FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. Nucleic Acids Res., 37, D560–D566

40. Myasnikova, E., Samsonova, M., Kosman, D. and Reinitz, J. (2005) Removal of background signal from *in situ* data on the expression of segmentation genes in drosophila. Dev. Genes Evol., 215, 320–326

41. Hassani, H. (2007) Singular spectrum analysis: Methodology and comparison. J. Data Sci., 5, 239–257

42. Holloway, D. M., Harrison, L. G., Kosman, D., Vanario-Alonso, C. E. and Spirov, A. V. (2006) Analysis of pattern precision shows that *Drosophila* segmentation develops substantial independence from gradients of maternal gene products. Dev. Dyn., 235, 2949–2960

43. Alexandrov, T., Golyandina, N. and Spirov, A. (2008) Singular spectrum analysis of gene expression profiles of early *Drosophila* embryo: Exponential-in-distance patterns. Res. Lett. Signal Process., 2008, 825758

44. Alexandrov, T. (2009) A method for trend extraction using Singular Spectrum Analysis. Rev. Stat., 7, 1–22

45. Golyandina, N. E., Holloway, D. M., Lopesc, F. J. P., Spirov, A. V., Spirova, E. N., and Usevich, K. D. (2012). Measuring gene expression noise in early *Drosophila* embryos: nucleus-to-nucleus variability. Procedia. Comput. Sci.9, 373–382

46. Spirov, A. V., Golyandina, N. E., Holloway, D. M., Alexandrov, T., Spirova, E. N., Lope, F. (2012) 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. Evol. Comp. Mach. Learning and Data Mining in Bioinformatics, 7246, 177–188

47. Holloway, D. M., Lopes, F. J. P., da Fontoura Costa, L., Travençolo, B. A. N., Golyandina, N., Usevich, K. and Spirov, A. V. (2011) Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in *Drosophila* segmentation. PLoS Comput. Biol., 7, e1001069

48. Ghodsi, Z., Silva, E. S. and Hassani, H. (2015) Bicoid signal extraction with a selection of parametric and nonparametric signal processing techniques. Genomics Proteomics Bioinformatics, 13, 183–191

49. Holloway, D. M., Lopes, F. J. P., da Fontoura Costa, L., Travençolo, B. A. N., Golyandina, N., Usevich, K. and Spirov, A. V. (2011) Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in *Drosophila* segmentation. PLoS Comput. Biol., 7, e1001069

50. Golyandina, N. E., Holloway, D. M. and Lopes, F. J. P., Spirov, A. V., Spirova, E. N. and Usevich, K. D. (2012) Measuring gene expression noise in early *Drosophila* embryos: nucleus-to-nucleus variability. Procedia Comput. Sci., 9, 373–382

51. Hassani, H. and Ghodsi, Z. (2014) Pattern recognition of gene expression with singular spectrum analysis. Med. Sci., 2, 127–139

52. Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., et al. (2009). FlyBase: Enhancing *Drosophila* gene ontology annotations. Nucleic Acids Res., 37, D555–D559

53. Gelbart, W., Bayraktaroglu, L., Bettencourt, B. and Campbell, K. (2003) The FlyBase database of the *Drosophila* genome projects and community literature. Nucleic Acids Res., 31, 172–175

54. Weigmann, K., Klapper, R., Strasser, T., Rickert, C., Technau, G., Jackle, H., Janning, W. and Klambt, C. (2003) FlyMove a new way to look at development of *Drosophila*. Trends Genet., 19, 310–311

55. Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M. and Reinitz, J. (2004) A database for management of gene expression data *in situ*. Bioinformatics, 20, 2212–2221

56. http://www.fruity.org/index.html

57. http://www.droidb.org/Index.jsp

58. http://www.ytf.org/

59. http://www.ymine.org/

60. http://yatlas.org/atlas.cgi

61. http://www.flycircuit.tw/

62. http://urchin.spbcas.ru/yex

63. Kosman, D., Reinitz, J. and Sharp, D. H. (1998). Automated assay of gene expression at cellular resolution. In Proc. 1998 P SB, 617

64. Janssens, H., Kosman, D., Vanario-Alonso, C. E., Jaeger, J., Samsonova, M. and Reinitz, J. (2005) A high-throughput method for quantifying gene expression data from early *drosophila* embryos. Dev. Genes Evol., 215, 374–381

65. Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M. and Reinitz, J. (2001) Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. Bioinformatics, 17, 3–12

66. Kozlov, K., Myasnikova, E., Pisarev, A., Samsonova, M. and Reinitz, J. (2002) A method for two-dimensional registration and constrution of the two-dimensional atlas of gene expression patterns *in situ*. In Silio Biology, 2, 125–141