ImageServer, a Tool for On-line Processing and Analysis of Biological Images

Andrei Pisarev[1], Ekaterina Poustelnikova[1], Konstantin Kozlov[1], Ekaterina Myasnikova[1], Maria Samsonova[1] and John Reinitz[2]

[1]Department of Computational Biology, St.Petersburg State Polytechnical University,195251 St.Petersburg, Russia;

[2]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York, 11794-3600, USA

Abstract

We present a novel tool for processing and analysis of images of gene expression patterns stored in a relational database. This tool, known as ImageServer, is portable across different software/hardware platforms and supports both basic and subject domain oriented operations on images. The tool is used to process and analyze images stored in FlyEx database (http://urchin.spbcas.ru/flyex); software and documentation are available for download from http://urchin.spbcas.ru/downloads/IS/IS.htm.

1　　Introduction

In biology visual information has been estimated to represent as much as 70% of all data generated. In recent years the widespread use of computers and digital image capture devices has resulted in accumulation of large amounts of digital images. However, in spite of these technological advances visual information is still for the most part analyzed by qualitative methods. Image processing is usually performed by graphic packages which are installed on a local computer and do not have any built-in tools for image management [1, 2]. This approach requires to download an image from a database to a local computer in order to process it, as well as to insert the processing results in the database again. So far only a few systems provide solution for integrated storage, processing and analysis of image data [3]. Thus the development of software for database-based processing and analysis of visual information is an important task for biology and computer science.

In this work we present a software tool ImageServer, which is designed for on-line processing and analysis of biological images stored in a database. ImageServer performs both basic and subject domain oriented operations on images. Currently the target application of this tool is processing and analysis of images of segmentation gene expression patterns in *Drosophila* stored in FlyEx database [4]. FlyEx is a spatiotemporal atlas of segmentation gene expression. It was developed using IBM DB2 RDMS to answer questions about dynamics of formation of segmentation gene expression domains.

2　　Methods

2.1　　System architecture

ImageServer is an application server (Figure 1), which runs on the server under control of Java Virtual Machine. Use of Java programming language provides for platform independence of the tool. ImageServer implements a subset of standard web server functions,

in particular, GET method. Clients interact with the tool via the HTTP protocol that allows to use both FireWall and Proxy servers. ImageServer is permanently waiting for client requests listening to an IP-port with a given number. To guarantee parallel and independent work of several clients a separate thread is created for each client's request. Clients invoke ImageServer by including the standard tags <image> into the body of a HTML page. The <image> tags have to contain the server URL and the parameters - images (operands), operations and other settings. IS can access images which are stored in a relational database as BLOBs, image files in popular graphic formats (JPEG, GIF, TIFF, BMP, PNG, etc.), as well as the files in RAW format, which represents the byte array of intensities for each image pixel. By default, the target image has the JPEG format, however the output format can be specified explicitly using the corresponding parameter. To provide software independence ImageServer interacts with a database via the Java Database Connectivity (JDBC) protocol.
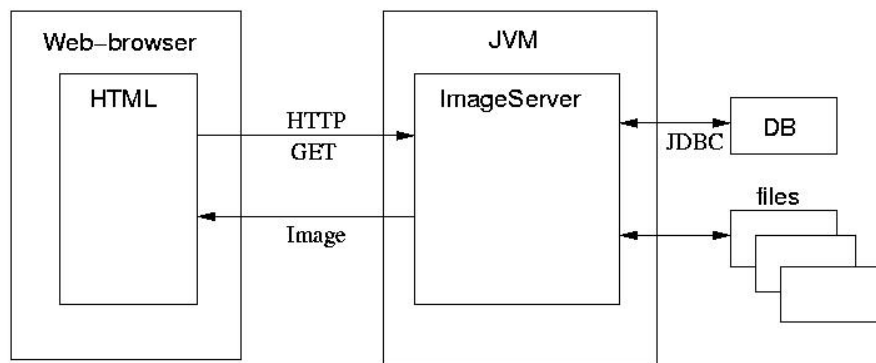

Figure 1: System architecture.

The XML template makes it possible to easily connect the ImageServer to any database. In this template the information about user, password, database, as well as a SQL query should be specified. The SQL query contains the identifiers, which are replaced by the values of parameters present in the HTTP request.

2.2   Methods for image processing

ImageServer is currently applied to process and analyze images of segmentation gene expression patterns in *Drosophila* stored in the FlyEx database. The basic image processing operations are implemented by means of the JMagick class library, which represents a Java interface to the ImageMagick package. These packages are publicly available [5, 6]. The subject domain oriented methods for image processing, namely background removal and image registration, are described below.

2.2.1   Description of subject domain

FlyEx is a quantitative atlas of segmentation gene expression at cellular resolution [4]. Segments are repeated units of the insect body. In the fruit fly the segmental architecture is determined at the 'syncytial blastoderm' stage, in particular at cleavage cycle 14A which lasts from 130 to 180 minutes after fertilization. [7]. At this stage of development an embryo is a hollow asymmetrical ellipsoid of nuclei which are not separated by cell membranes. The initial determination of the segments is a consequence of the expression of 16 genes which are mainly transcription factors [8, 9].

The FlyEx database contains the images of segmentation gene expression patterns, as well as the numerical data obtained from these images by applying a six step data pipeline [10]. The images of gene expression patterns were obtained by confocal scanning microscopy using

fluorescence tagged antibodies. These images served as a raw material for quantification of gene expression [10, 11, 12]. As a result the reference data on expression of segmentation genes at cellular resolution and at each time point were constructed [12, 13]. Images and quantitative gene expression data from individual embryos, as well as reference gene expression data were used to study the dynamics of formation of segmentation gene expression domains, precision of development and pattern formation, as well as the mechanisms of segment determination [14, 15].

2.2.2   Removal of background signal

At the very first stage of data quantification the non-specific background signal is removed from images of gene expression. The aim of background removal is to bring the data to the unified standard form with a zero background and to get rid of distortions of gene expression patterns caused by the presence of a background signal. Our method [11] is based on the observation that the level of expression of a given gene in a null mutant embryo for that gene is well fit by a very broad two dimensional paraboloid. This paraboloid is automatically determined from the areas of wild type embryos in which a given gene is not expressed and the whole image is then normalized by the paraboloid to remove background from the entire embryo by a linear rescaling of pixels' intensity. The coefficients of the paraboloid are precomputed and stored in the database.

2.2.3   Registration

Image registration is necessary to eliminate small spatial difference between expression patterns of one gene in individual embryos of one age. We have developed a registration method based on the minimization of the squared distance between extrema of expression pattern of *even-skipped* gene in different embryos by affine coordinate transformation. Locations of extrema are determined by two methods: quadratic spline approximation and fast dyadic redundant wavelet transform (FRDWT) [10, 11, 12]. The coefficients of affine transformation computed for both methods are stored in the database for all the images.

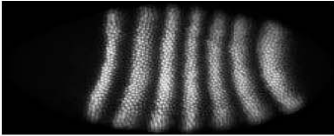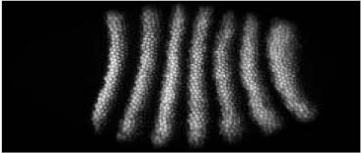## 3   Implementation

3.1   User interface

To retrieve images a user fills query forms, which can be accessed from the main page of the FlyEx database by selecting links "Images of gene expression patterns" or "Analysis tools. Images of gene expression patterns." The control panels for analysis of retrieved images are placed at the bottom of the HTML page containing a query result (Figure 2).

3.2   Analysis of image information

A single image can be subjected to scaling, cutting of rectangular area, filtering of fluorescence intensity, contrast enhancement and background removal. Admissible operations on a set of images are masking of one image by another, combination of up to three greyscale images into the color one, generation of an absolute value of difference between two images, registration of several images. It is possible to combine several operations in a single request.

Figure 2: The result of sample query showing two images of expression of *even-skipped* gene scanned in the embryos dr2 and ds2 belonging to the cleavage cycle 14A, temporal class 7. The control panel for performing various manipulations with these images is presented below on the same page.

3.2.1   Combining of up to three greyscale images into the color one

The confocal microscope at our disposal permits to detect the expression of up to three genes in one embryo, for each gene a greyscale image is obtained and stored in the database. For simultaneous visualization of expression of genes scanned in one embryo the greyscale images can be combined into one color image, in which the expression pattern of each gene is coded by one of the basic colors of the RGB format (see Figure 3).

3.2.2   Masking of one image with another

The data quantification includes as an essential step the segmentation of images. To construct a binary nuclear mask each pixel on the raw image is classified as belonging to a nucleus or not, so that on the mask a pixel is equal to one if and only if that pixel is located on a nucleus. Hence the quality of quantitative data is defined by the accuracy of nuclear mask. A user can observe a mask (Figure 4A) and superpose it on the image of expression patterns. This results

in the masked image displaying the localization and shape of nuclei (Figure 4B). Masks of all the embryos are stored in the database.
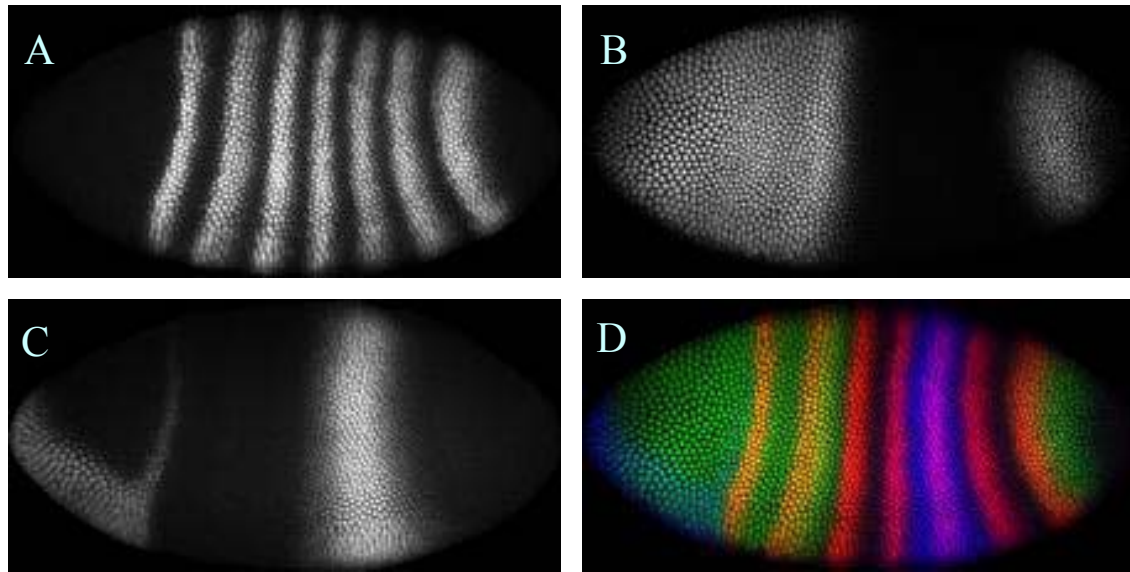

Figure 3. Visualization of expression patterns of genes scanned in embryo dr2. The greyscale images of expression patterns of genes *even-skipped* (A), *hunchback* (B) and *knirps* (C) obtained with confocal microscope. (D) The resultant color image with *even-skipped* in red, *hunchback* in green and *knirps* in blue. Images are from FlyEx database, image in D is generated on-the-fly by ImageServer.
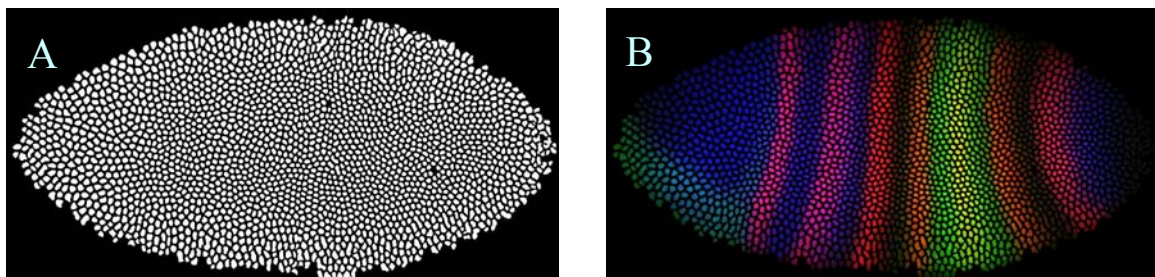

Figure 4: (A) Nuclear mask for embryo dr2, nuclei are visible as white islands. (B) The masked image obtained by superposition of this mask on the image shown in Figure 3D. Here only those pixels are colored that belong to the nuclear mask and are white in Figure 4A.

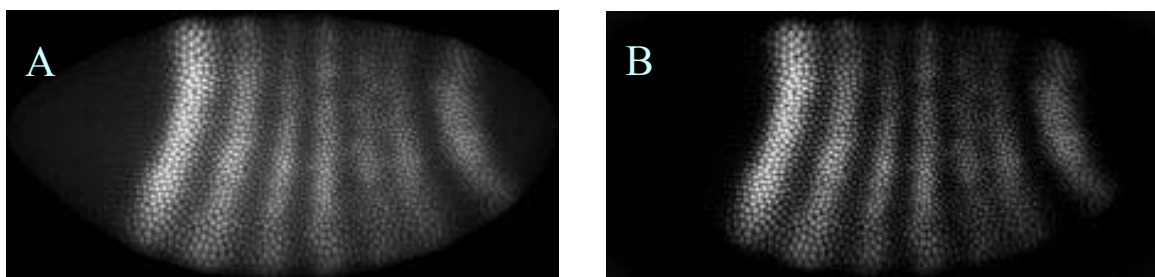### 3.2.3   Estimation of background signal

Figure 5: Expression pattern of gene *even-skipped* in embryo cq3 belonging to temporal class 4. (A) Original image, (B) the same image without background.

Each raw image contains a certain amount of background signal which can be observed visually. To estimate the level of background a user has a possibility to view one and the same

image before (Figure 5A) and after (Figure 5B) background removal. It is evident that the image without background is much more contrast. It should be noted that the image of background alone can not be obtained by simple subtracting the image with no background from the original one because the background removal algorithm requires to rescale the pixel's intensities (2.2.2). The background image reconstructed from the paraboloid coefficients is not informative visually due to its low contrast.

3.2.4   Estimation of spatiotemporal variability of expression of a given gene

Cleavage cycle 10   **Cleavage cycle 13**   **Cleavage cycle 14A**
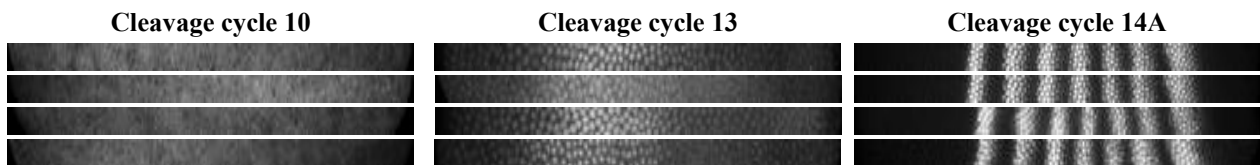

Figure 6: Visual estimation of spatiotemporal variability of gene expression.

Using the web interface shown in Figure 2 the spatiotemporal variability of gene expression can be visually checked. Figure 6 shows the variability of expression of highly dynamic *even-skipped* gene. For each time point (i.e. cleavage cycle). four images representing the gene expression in individual embryos are displayed. The central 10% strip along A-P axis was cut from each image for better observation. The degree of spatial variability can be easily seen by comparing the images of embryos belonging to one cleavage cycle. Temporal variability can be estimated from comparison of images from different cleavage cycles. At cleavge cycles 10 and 13 *even-skipped* is expressed in one broad domain, at late clevage cycle 14A 7 narrow isolated domains of expression are formed.

3.2.5   Temporal dynamics of gene expression

The temporal dynamics of gene expression can be visualized by comparing the images of embryos of different age. ImageServer allows to combine up to three greyscale images of expression patterns into the color one (see 3.2.1). Figure 7D illustrates the result of combining of three images displaying the expression pattern of *even-skipped (eve)* gene at different time points, namely temporal classes 1, 4 and 8 of cleavage cycle 14A. In the resultant image each combined image is coded as one of the basic colors of the RGB format. Thus the areas, where *eve* is expressed at all times, appear as blends of colors, while the areas, where *eve* starts or stops to express, are displayed as one of the basic RGB colors. It is evident that the leftmost *eve* expression domain (the first stripe) has a stable position, while all the rest six stripes of this gene expression move to the left (to the anterior) with time, and the movement of the rightmost (the posterior) stripe is the most pronounced.

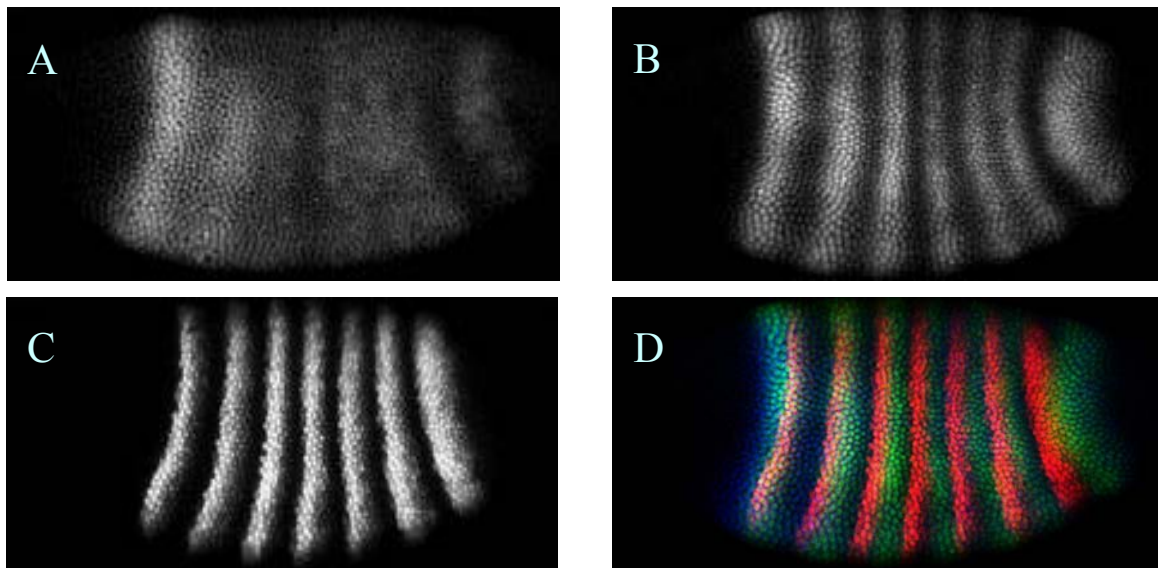### 3.2.6   Estimation of the accuracy of registration

Figure 7: Temporal dynamics of expression of *even-skipped* gene. Expression patterns from 3 embryos of different age: (A) temporal class 1, (B) temporal class 4 and (C) temporal class 8. (D) is obtained by combining A, B and C, which are color coded as blue, green and red, correspondingly.

To estimate the accuracy of registration (2.2.3) up to three images of expression patterns before and after registration can be combined in one color image (Figure 8). The registered expression domains have similar position and thus appear as blends of basic colors of the RGB format (Figure 8B).
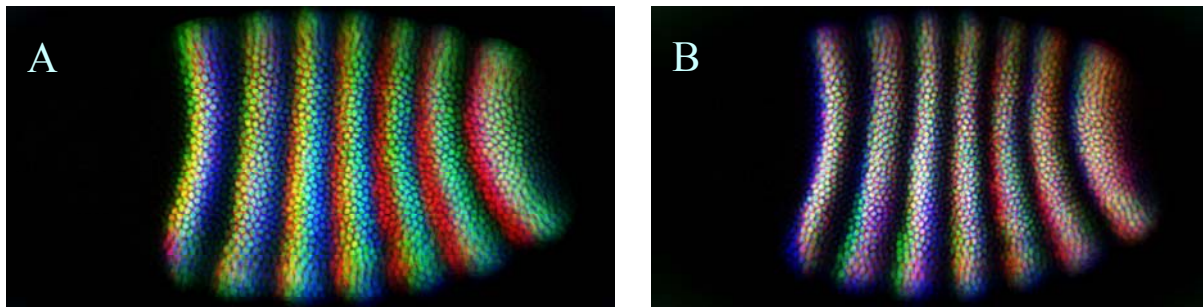

Figure 8: Three expression patterns of *even-skipped* gene before registration (A) and after registration (B).

4   Discussion

Image management and processing of image information are usually performed by different types of software. Images are processed by graphic packages which do not keep track of data and images in a rigorous way, while databases simply present selected images and metadata to a user [16, 17]. The solution, which we propose here, allows to integrate image processing and analysis with the information storage. We have designed the application server, which on the one hand can access images stored in any database or file management system and on the other hand supports different operations on images ranging from image scaling to registration

of several images. Due to application of Java programming language ImageServer can be ported to any software/hardware platform.

The efficiency of on-line processing and analysis of images stored in a database to a great extent depends on the size of images. Currently ImageServer is applied to process and analyze images of segmentation gene expression patterns in *Drosophila* stored in FlyEx database. These images are not large: the typical size of an image of expression pattern of one gene scanned in one embryo is at maximum about 1300 650 pixels and the typical size of an image file in JPEG format is about 60K. While in our system the joint operation of image retrieval, conversion to JPEG format and visualization takes about 300 msec, the execution of the same operation, as well as other image processing/analysis operations on larger images could take a considerable time. To speed up the performance of ImageServer on images of larger size we plan to modify this tool to support processing and analysis of images subdivided into tiles.

ImageServer can be easily extended to support processing and analysis of new images by addition of new basic and subject domain specific operations implemented as program modules written in C++ or Java. This tool is the core of the Laboratoty Management System for processing and analysis of images of gene expression patterns *in situ,* which is currently being developed by authors. The ImageServer software and a test database can be downloaded from http://urchin.spbcas.ru/downloads/IS/IS.htm.

Acknowledgements

References

[1]   Marzola, P., Da Pra, A., Sbarbati, A., Osculati, F. A PC-based workstation for processing and analysis of MRI data. MAGMA, 7, 1, 16–20, 1998.

[2]   Rasure, J. and Young, M. In SPIE/IS&T Symposium on Electronic Imaging Vol.1659, of SPIE Procedings, 1992.

[3]   Swedlow,J.R., Goldberg,I., Brauner,E., Sorger,P.K. Informatics and quantitative analysis in biological images.Science, 300, 100–102, 2003.

[4]   Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M., Reinitz, J. A database for management of gene expression data *in situ*. Bioinformatics, 20, 2212–2221, 2004.

[5]   Yeo, E. (2003) JMagick http://www.yeo.nu/jmagick/.

[6]   Still, M. Graphics from the command line. IBM developerWorks, July 16, 2003 http://www-106.ibm.com/developerworks/library/l-graf/?ca=dnt-428.

[7]   Foe, V.A. and B.M. Alberts Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. Journal of Cell Science, 61: 31-70, 1983.

[8]   Akam, M. The molecular basis for metameric pattern in the *Drosophila* embryo. Development, 101: 1-22, 1987

[9]   Ingham, P.W. The molecular genetics of embryonic pattern formation in *Drosophila*. Nature, 335: 25-34, 1988.

[10]   Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M. And Reinitz, J. Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. Bioinformatics,17(1), 3–12, 2001.

[11]   Myasnikova, E., Samsonova, M. and Reinitz, J. Removal of background signal from in situ data on the expression of segmentation genes in *Drosophila*. Devel. Genes *&* Evol., in press.

[12]   Kozlov K., Myasnikova E., Pisarev A., Samsonova M., Reinitz J. A method for two-dimensional registration and construction of the two-dimensional atlas of gene expression patterns in situ. In Silico Biol., 2, 125-141, 2002, http://www.bioinfo.de/isb/2002/02/0011.

[13]   Kozlov, K., Myasnikova, E., Samsonova, M., Reinitz, J., and Kosman, D. (2000). Method for spatial registration of the expression patterns of *Drosophila* segmentation genes using wavelets. Computational Technologies, 5:112-119.

[14]   Jaeger, J., Blagov, M.,Kosman, D.,Kozlov, K., Manu, Myasnikova, E.,Vanario-Alonso, C.,Samsonova, M.,Sharp, D., and Reinitz, J. Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila* melanogaster. Genetics, 167: 1721-1737, 2004.

[15]   J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. Kozlov, Manu, E. Myasnikova, C. Vanario-Alonso, M. Samsonova, D. Sharp, and J. Reinitz Dynamic control of positional information in the early *Drosophila* embryo, Nature, 430: 368-371, 2004.

[16]   Gonzalez-Couto, E., Hayes, B., Danckaert, A. The life sciences global image database (GID). Nucl. Acids Res., 29, 1,:336–339, 2001.

[1]   Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S. Q., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., Rubin, G. M. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. Genome Biology, 3(12), research0088.1–0088.14, 2002.