THE RECONSTRUCTION OF THE DROSOPHILA SEGMENTATION MECHANISMS FROM EXPERIMENTAL DATA: PROCESSING AND ANALYSIS OF CONFOCAL IMAGES OF EXPRESSION PATTERNS

*1Spirov A.V., 2Timakin D.L., 3Reinitz J., 3Kosman D., 1Spirova O.A.

1The Sechenov Institute of Evolutionary Physiology and Biochemistry, 44 Thorez Ave., St. Petersburg, Russia; spirov@iephb.nw.ru;
2St.Petersburg State Technical University, Computer Science and Engineering Faculty, System e-mail: dtimekin@mail.ru, dave@eve.molbio.mssm.edu
3Dept. of Biochemistry and Molecular Biology, Box 1020 Mt. Sinai Medical School, One Gustave L., USA; reinitz@kruppel.molbio.mssm.edu, spirov@iephb.nw.ru
*Corresponding author

Keywords: genetic networks, genes of segmentation, pattern of expression, confocal microscopy, image analysis, image processing, computer analysis, reconstruction of gene regulation, atlas of expression, Drosophila melanogaster

Resume

Motivation:
Modern large-scale "functional genomics" projects are inconceivable without the automated processing and computer-aided analysis of images. The project we are engaged in is aimed at the automated transformation of gene expression data in confocally scanned images of the fruit fly (Drosophila melanogaster) embryos into an electronic database of expression. For the detailed reconstruction of the fly segmentation mechanisms we need to receive statistically authentic summary picture of detailed pattern dynamics proceeding from a large number of treated embryos.

Results:
Our investigations have leaded us to conclusion about necessity of minimum three procedures of processing of images for compiling of final data set. It is elastic deformation, registration and interpolation procedures. We have developed and tested family of programs for the three image-processing steps. Biologically significant results obtained by these procedures will be discussed.

Availability:
The codes of the programs in C++ and documentation are available via Internet at http://www.mssm.edu/molbio/hoxpro/atlas/atlas.html

Introduction

Computer-Aided Analysis of Biological Images.

The ongoing revolution in molecular genetics has progressed from the large-scale automated characterization of genomic sequence to the characterization of the biological function of the genome. These investigations mark the beginning of the era of "functional genomics" (Lander, 1996). A key feature of genomic scale approaches is the automated treatment of large amounts of data. Both current and future work in the field is impossible without the automated processing and computer-aided analysis of images in connection with updating interactive electronic image databases.

A key aspect of such processing involves the segmentation of individual images, the registration of serial images, and the interpolation of 2D fields of concentrations of segmentation factors. Many problems involving the recognition, classification, segmentation, registration, and interpolation of images can be formulated as optimization problems. Contemporary approaches based on evolutionary computations are a promising avenue for the solution of such problems.

The work reported here is part of a large scale project to construct a model of segment determination in the fruit fly Drosophila melanogaster based on coarse-grained chemical kinetic equations (Reinitz et al., 1998). The acquisition and mapping of gene expression data at a heretofore unprecedented level of precision is an integral part of this project. The current emphasis in our work is on the automated transformation of gene expression data in confocally scanned images into an electronic database of expression. Here we describe our evolutionary computations-based approach to image processing for quantitative atlas of Drosophila genes expression. Biologically significant results obtained by these procedures will be discussed.
Methods and algorithms

Images of Drosophila Genes Expression

Processing of embryonic images begins with data expressed in terms of the average fluorescence level (proportional to gene expression level) of each nucleus, where segmentation proteins exert their biological function. This data was obtained as follows.

Antibodies for 15 protein products of segmentation genes were raised and over 1000 images were prepared and scanned (Kosman and Reinitz, 1998). These images were computationally treated by means of the Khoros package. Embryos were rotated and cropped automatically. Next, the images were segmented (Kosman et al. in preparation). About 2000-2500 segmented and identified nuclei are obtained from each image. Each nucleus is labeled numerically, and the x and y coordinates of its centroid are found, together with the average fluorescence level over that nucleus. The segmented data takes the form of tables in ASCII text format. The result is the conversion of an image to a set of numerical data which is then suitable for further processing.

Elastic Deformations: "Stripe Straightening" Procedure

Early in the development the fruit fly embryo is shaped roughly like a hollow prolate ellipsoid, composed of a shell of nuclei which are not separated by cell membranes. Deviations from the ellipsoidal shape reveal the future polarity of the animal's body: The more pointed end on the long axis makes anterior (head) structures, and the rounder end posterior (tail) structures. From a lateral (side) perspective, one long edge of the embryo is flat and will make dorsal ("back") structures, while the other long edge is rounded and makes ventral ("underside") structures. But what is more, so called pair-rule stripes (early markers of the future segmental pattern (Akam, 1987) are not parallel and straight, but have a crescent-like form. The curvature of the stripes is highest at the termini, and minimal at the central part. Each stripe specifies an anterior-posterior location, and these stripes can be regarded as contours in an intrinsic coordinate system (Spirov et al., 2000) that is being created by the embryo itself.

Our data processing begins with a smooth transformation of spatial coordinates. If the image is smoothly transformed such that the curvilinear coordinates are plotted orthogonally, the stripes appear straight, so the determination of these coordinates can be viewed as a "stripe straightening" procedure (Spirov et al., 2000).

Registration of Serial Images

Our next procedure is registration of serial images. We need this treatment for full-scale quantitative comparisons and analysis of pattern dynamics at a single nucleus resolution. Registration cannot be performed on a direct nucleus by nucleus basis because of individual differences among embryos. Moreover, close inspection of the edge of a well-demarcated expression domain shows irregularity due to the arrangement of nuclei, which do not lie on a rectangular or hexagonal grid. What is more, any two embryos of the same age can differ in size and form.

Our preliminary at hand computations demonstrated that registration of Drosophila early blastoderm images takes elastic deformations. So we can use practically the same approach, as is the case of the stripe straightening problem.

Interpolation

Because blastoderm nuclei don't form either regular square or regular hexagonal mesh one-dimensional and two-dimensional interpolation of expression patterns are non-trivial computational problems. However resolution of this problem is essential requirement for the dynamical modeling and statistical analysis of the expression data. Hence all of the following calculations drastically depend on the correct identification of interpolation function for this irregular mesh.

We used and compared several standard approaches for 1D and 2D interpolation. Particularly it was 1D and 2D spline (cubic spline) and 1D and 2D Fourier interpolations. However all these procedures require either regular mesh or are based on transition to a regular mesh. And it has appeared completely unacceptable for the level of precision, which is pursued in our project.

All this has motivated us to take advantage of an interpolation by truncated two-dimensional Fourier polynomials. The power of series was chosen empirically. The Fourier coefficients were found by optimization techniques, while comparison of interpolation result was performed on given irregular mesh of each image under treatment.

Availability

The codes in C++ for the programs for elastic deformations, registration and interpolation, as well appropriate documentation, are available via Internet at http://www.mssm.edu/molbio/hoxpro/atlas/atlas.html.
Results and Discussion.

**Experimental Determination of Drosophila Embryonic Coordinates**

This elastic deformation procedure called stripe straightening gives possibility to found following:
1. The intrinsic coordinates found for primary pair-rule gene *even-skipped* fits for other gap and pair-rule genes;
2. Curvilinearity of coordinates in anterior and posterior parts of an embryo is controlled by autonomous mechanisms;
3. Curvilinearity of coordinates in anterior part probably is under control of the morphogen *bicoid* gradient.

**Statistical Features of Expression of the Segmentation Genes at Single-Nucleus Resolution**

It is shown, that
1. The space distribution and the level of a fractional error of fluorescence of the segmentation factors correlates with a position of the factor in the cascade of the genes-controllers of segmentation;
2. In limits of an investigated time interval (from early to late cleavage cycle 14) the irregularity of the boundaries of domains of the genes expression increases;
3. For genes of pair-rule group at least up to late 14 division stage, the evident increase of a proportion of the nuclei is observed, the level of an expression of these genes in which is essentially (~ 5-20 %) differs from average for the nearest neighbors.

**Temporal Ranking, Registration and Classification of the Time-Series of Expression Patterns**

The developed methods of serial registration of images make possible to realize the following procedures:
1. Automatic ranking of images according to “maturity” of pattern expression resulting in temporally ordered series beginning from earliest up to mature patterns;
2. Automatic subdivision of obtained time series according to known temporal stages (in accordance with accepted classification);
3. Registration of images on the basis of obtained time-series;
4. Visual methods of comparison of pattern dynamics of different genes with high precision;
5. Comparison of stages and rate of the expression patterns maturation for any pairs of studied genes.

**Acknowledgements**

This work is supported by RFBR, grant No 00-04-48515; INTAS, grant No 97-30950; USA National Institutes of Health, grant RO1-RR07801; and GAP awards RBO-685 and RBO-895.

**References**