Mapping organism expression levels at cellular resolution in developing *Drosophila*

David W. Knowles*, Soile Keränen+, Mark D. Biggin+, Damir Sudar*

Departments of Cell and Molecular Biology* and Genome Sciences+, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

ABSTRACT

The development of an animal embryo is orchestrated by a network of genetically determined, temporal and spatial gene expression patterns that determine the animals final form. To understand such networks, we are developing novel quantitative optical imaging techniques to map gene expression levels at cellular and sub-cellular resolution within pregastrula Drosophila. Embryos at different stages of development are labeled for total DNA and specific gene products using different fluorophors and imaged in 3D with confocal microscopy. Innovative steps have been made which allow the DNA-image to be automatically segmented to produce a morphological mask of the individual nuclear boundaries. For each stage of development an "average" morphology is chosen to which images from different embryo are compared. The morphological mask is then used to quantify gene-product on a per nuclei basis. What results is an atlas of the relative amount of the specific gene product expressed within the nucleus of every cell in the embryo at the various stages of development. We are creating a quantitative database of transcription factor and target gene expression patterns in wild-type and factor mutant embryos with single cell resolution. Our goal is to uncover the rules determining how patterns of gene expression are generated.

Keywords: Confocal Imaging, Image Segmentation, Image Registration, Quantitative Fluorescence Analysis, Regulatory Gene Networks, *Drosophila*

INTRODUCTION

The morphology and development of an animal is mediated by transcription factors that bind to and regulate the differential expression of genes. Usually several factors cooperate to activate any specific gene and the abundance of transcription factors themselves is determined by the expression of other genes. As a consequence, transcription factors form a complicated spatial and temporal network of interaction as they orchestrate the development of an organism. The aim of this work is to create a new imaging and image analysis technology that will enable the study of entire transcriptional networks. The work is in response to the need for novel approaches to understand how complex patterns of gene expression arise. We have focused on early *Drosophila* embryogenosis because the genomic sequence of *Drosophila melanogaster* is known[1], an entire *Drosophila* embryo can be imaged with single cell resolution and the early *Drosophila* transcriptional network is controlled by a relatively small number of between 80 and 100 sequence specific factors, most of which have been well characterized.

The acquisition of large scale quantitative data is why microarray analysis has been so successful in identifying transcription factor target genes in microbes and animal cultured cells[2,3,4] but these techniques do not provide spatial information about the spatial distribution of gene expression within an organism. When the expression of randomly selected genes is compared between wild type embryos and embryo mutant for a transcription factor, the patterns change in complex ways[5] that are impossible to accurately describe verbally and would be totally missed by any technique that ignores spatial information [Figure 1]. Previous imaging work has been done to quantify gene expression in *Drosophila* at a per nucleus resolution[6,7] but this work is effectively one dimensional because the analysis was restricted to an anterior/posterior strip of nuclei in images collected in two dimensions (2D). In our technique, a combination of high resolution optical imaging and image analysis is providing a three dimensional (3D) "atlas" of nuclear position and the relative amount of gene produce at cellular and nuclear resolution over the entire organism at various growth stages[8]. Our goal is to provide quantitative data of the expression of many genes in relation to their

spatial pattern and allows temporal and spatial correlations between the expression pattern of a transcription factor protein and that of the mRNA of its target genes.

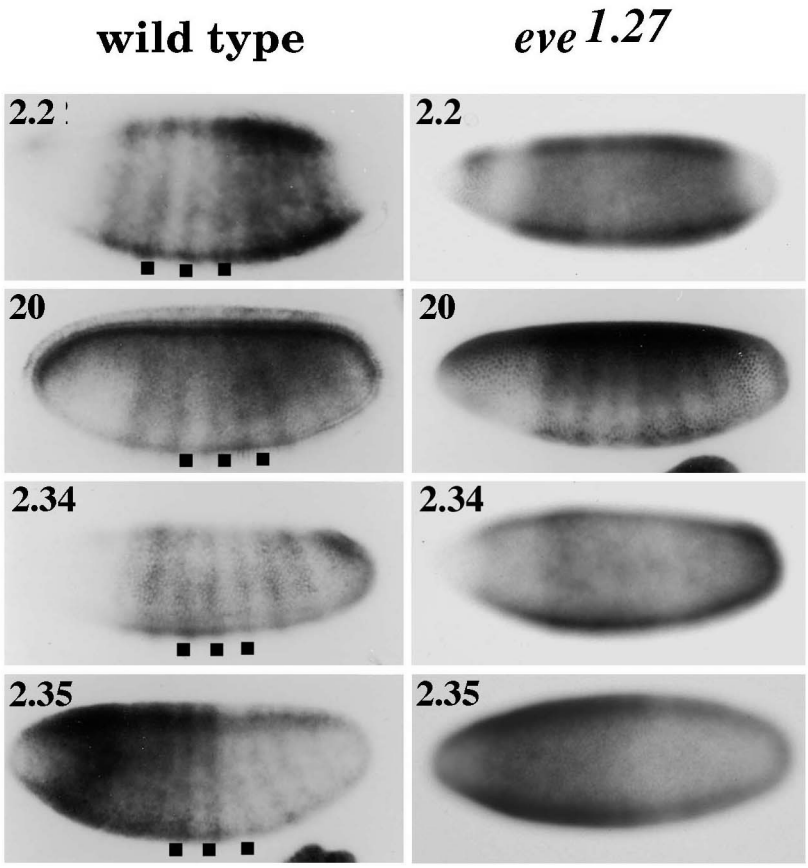

Figure 1: The mRNA expression pattern of 4 randomly selected genes, from clones 2.2, 20, 2.34 and 2.35, just prior to gastrulation in wild type and *eve* minus embryos[5]. Such patterns emphasize the necessity of quantifying the spatial distribution of gene expression.

The work can be divided into three parts; 1) defining embryo topology at nuclear resolution, 2) creating a nuclear "map" for each embryonic stage and 3) quantifying the relative amounts of gene expression. Firstly, embryos at various growth stages are imaged in 3D for fluorescently-labeled DNA using confocal fluorescence microscopy. The images are automatically segmented by an algorithm using a combination of local brightness thresholding and template matching to locate the nuclei. The segmentation produces a enumerated image-mask which defines the location and extent of every nucleus in the embryo. This approach has provided the first measurement of the total number of nuclei in an individual embryo and in turn is used to stage the embryo's growth. This is possible because the first 14 nuclear divisions in *Drosophila* are synchronous and this has allowed us to determine the relationship between the number of nuclei and the number of nuclear divisions from fertilization. Secondly, many embryo at the same stage are analyzed using morphological and registration algorithms to create an average, stereotypical "map" of nuclear distribution at that stage. These algorithms first adjust the gross morphology of individual embryos to fit a predetermined shape. Then the position and number of individual nuclei in the map are adjusted so they best represent those of the individual embryo being used to create the map. A single map is generated for each stage and this not only defines the average morphology of embryo, at that stage, but provides information about the biological variability in nuclei number and position. This information is being used to understand the role of biological variability in development[9,10]. Thirdly, 3D images of embryo at various stages are collected for fluorescently-labeled specific gene-products and total DNA. Automated segmentation of the DNA-stain image produces a segmentation mask for each embryo. The segmentation mask is used to register both the DNA-stain and gene-product images to the embryonic map for that stage, and subsequently the mask is used to quantify the relative brightness per nucleus of both the DNA-stain and gene-product image. We have

demonstrated that the amount of fluorophor bound to DNA per nuclei is constant, and accordingly the brightness per nucleus of the DNA-stain image is a measure of the relative fluorescence collection efficiency of the optical system. These results are then used to normalize the gene-product image to calculate the relative amount of gene produce per nuclei. By systematically adding the results for many different gene-products to the embryonic maps we will create an atlas of the relative amount of the specific gene-products expressed within the nucleus of every cell in embryos at the various stages of development. The atlas defines a quantitative database of transcription factor and target gene expression patterns in wild-type and factor mutant embryos with single cell resolution. Our goal is to uncover the rules determining how patterns of gene expression are generated.

METHODS & RESULTS

Image Collection

Embryos at various stages were stained for total DNA with Sytox-green (Molecular Probes, S-7020, Eugene, OR) and for EVE-protein with Texas-Red monoclonal antibody. The labeled embryos were mounted in Permount for imaging. Single embryo images were collected on a Zeiss 410 Confocal Laser Scanning Microscope with a 20x, 0.75NA objective lens into stacks of 150 2D images, each 1024x1024 pixels square. The stacks were collected at 1μm intervals.

Image Segmentation

Nuclei in the DNA-stain images [Figure 2-top left] were segmented by a combination of local intensity thresholding followed by a template matching algorithm written in C and run on a Sun Ultra10 workstation (Sun Microsystems). The local intensity threshold algorithms calculates the absolute values of the difference between two images produced by convolving the original image with two uniform kernels of different volumes. The image convolved with the larger kernel can be thought of as the average brightness in that region and the image convolved with the smaller volume, the local brightness. Subtracting these images and separating the positive and negative portions results in images which define where the local brightness is greater than [Figure2-top right] and less than [Figure 2-bottom left], respectively, the average brightness. Importantly, these images are identically zero everywhere else. To further illustrate how this algorithm works, the local standard deviation of image points within the larger kernel at each image point is calculated [Figure 2-bottom right].

The template matching algorithm then defines a spherical kernel having a radius which approximates that of the nuclei in the image. This kernel is convolved with the positive image resulting from the local intensity threshold and a non-overlapping template is enumerated by color and stamped at the local maxima of this image. This produces the final segmented image [Figure 3, left]. The segmentation accuracy depends on both the growth stage and mitotic stage of the embryo and is demonstrated visually by overlaying an outline of the segmentation mask on the original gray scale image [Figure 3, right]. The segmentation mask is subsequently used to define the nuclear topology, to stage the embryo growth, to register images of different embryos at the same growth stage and to create the stereotypical maps. This area of the work is ongoing and will be presented in the future.

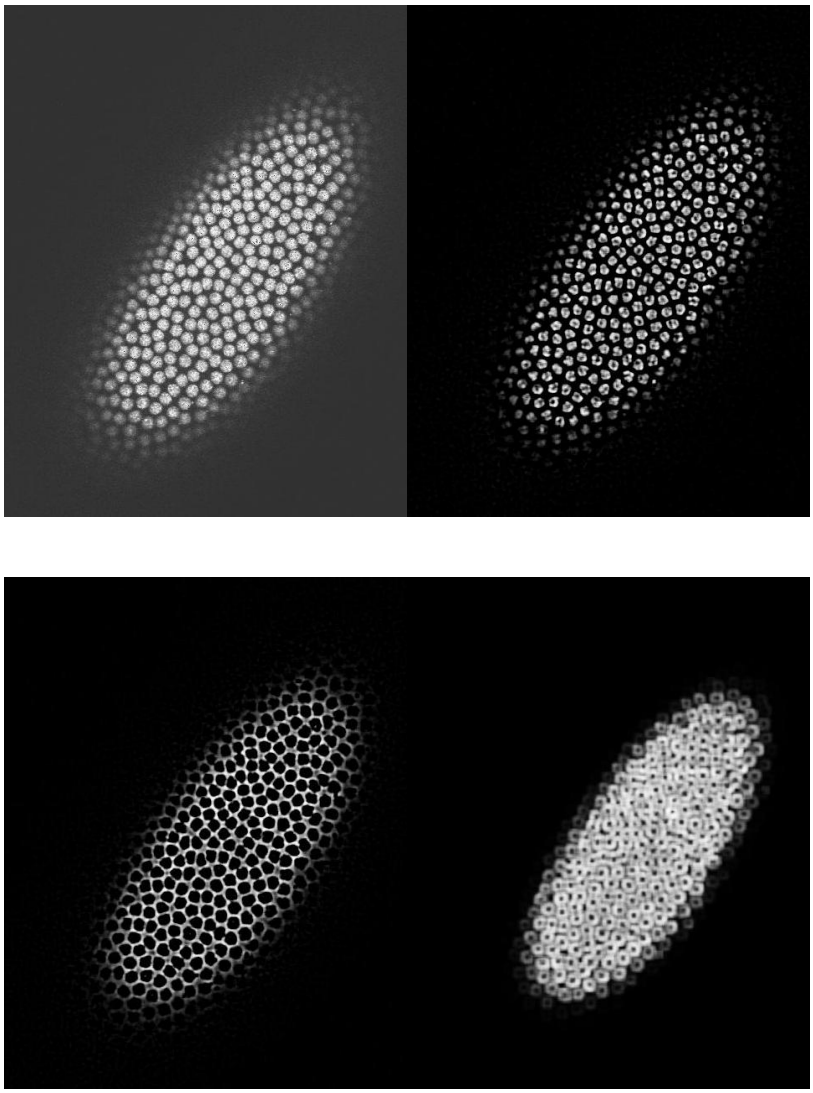

Figure 2: To calculate the position and extent of individual nuclear volumes, a local threshold algorithm analyzes the DNA-stain image (Top Left). The algorithm determines areas in the gray scale image which are brighter than their surrounds (Top Right), darker than their surrounds (Bottom Left), and calculates the local standard deviation (Bottom Right). In each case a single slice from a 3D image is shown.

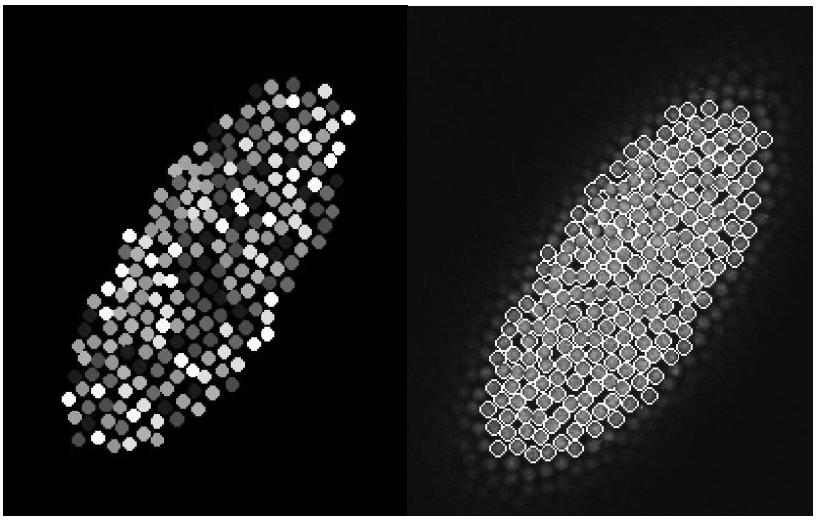

Figure 3: Following the local brightness threshold, a template matching algorithm stamps a non-overlapping, enumerated template of approximate nuclear volume centered on the nuclei. The accuracy of the segmentation is demonstrated here by overlaying an outline of the segmentation mask over the original gray level image (right). In each case only a single slice from a 3D image is shown

Quantifying the DNA-Stain Image

To understand the optical response of the image collection, so that its effects can be accounted for when quantifying the gene-product image, the brightness per nucleus of the DNA-stain image is calculated. The segmentation mask [Figure 3-left] defines the position and extent of each nuclei and allows the average brightness within each nuclear volume of the DNA-stain image to be determined. This data is presented in figure 4, where the average brightness per nuclear volume is plotted versus the average nuclear position along the optical axis. These values are then normalized to form a set of relative nuclear brightness which are enumerated with the number of the nuclei they correspond too.

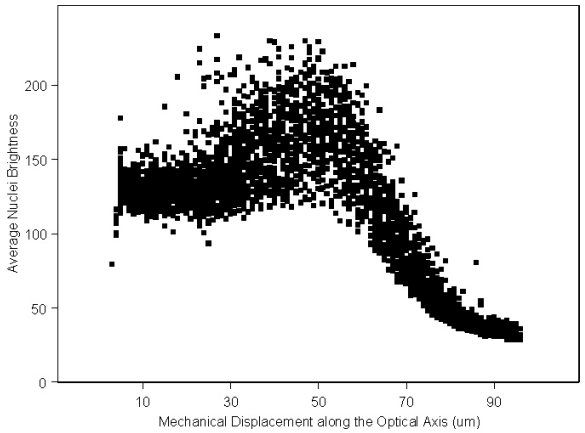

Figure 4: A plot of the average brightness of every nucleus within an embryo stained for DNA, plotted against the average position of the nuclei along the optical axis.

Quantifying Gene Expression

To quantify gene expression, embryo at many stages of growth are fluorescently labeled for DNA and a specific gene-product. A nuclear segmentation mask is generated for each embryo from the DNA-stain image [Figure 5-top left]. The

mask then directs the analysis of the gene-product image which in the case of figure 5 is EVE protein fluorescently labeled with a Texas-Red conjugated antibody [Figure 5- top right]. What results is relative amount of the specific gene-product expressed per nuclei within the entire embryo and this is added to the topological map for that stage of embryo. To demonstrate the results of this analysis the relative EVE expression per nucleus from this image slice is plotted against the average relative position of the nuclei along the anterior/posterior direction of the embryo [Figure 5-bottom].

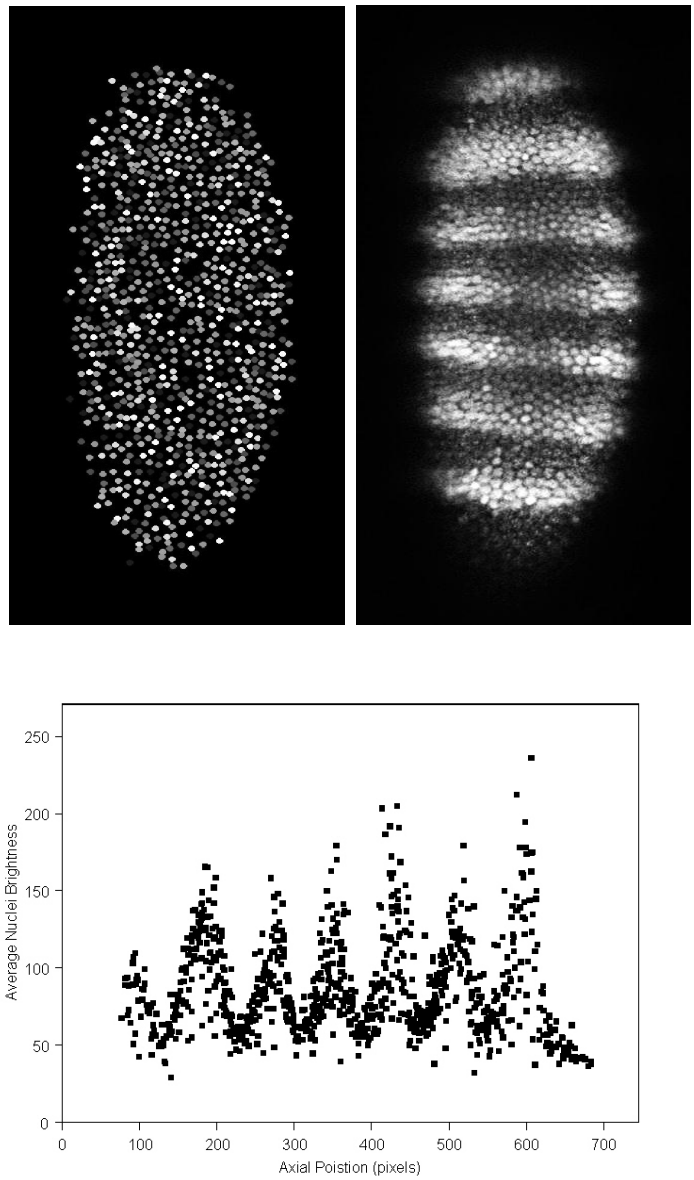

Figure 5: The segmentation mask (top, left) and corresponding gene-product image of Texas-Red stained EVE protein (top, right) are combined to calculate the relative amount of EVE expression per nuclei within an entire embryo. The relative amount of EVE expression from this image is plotted against the average relative anterior/posterior position of the nuclei within the embryo (bottom).

DISCUSSION

We are creating imaging and image analysis techniques to study the transcriptional network of early *Drosophila*, which despite its morphological simplicity, is complex.

In the first part of this work we have created novel automated segmentation algorithms which are uncovering *Drosophila* embryo topology at nuclear resolution. Because there are upwards of 8000 nuclei in early *Drosophila*, the need for automated segmentation is evident. The presented method, which uses a combination of local brightness thresholding and template matching, is just that. However, since the template is a fixed size, it does not segment the nuclear boundary as accurately as other techniques. Although this is not critical for the calculation of average brightness within nuclear volumes, work is being done to increase the segmentation accuracy of the nuclear boundaries while maintaining the speed and automation of the present algorithm. Work is also being done so that the segmentation accuracy can be automatically tested on a per embryo basis. This is important for the high throughput aspect of this work so that quality of embryo preparation can be assured. With these segmentation techniques we are able to class the growth of embryos based on the number mitotic divisions and produce topological maps showing number and the relative nuclear packing density. These techniques are not only providing the basis for studying gene expression in 3D images at nuclear resolution but are shedding light on the role of morphological variability in development. For example one important question being answered is whether morphological variability increases or decreases as the embryo develops.

The segmentation mask is the key to quantitative analysis. In combination with the DNA-stain image, the segmentation mask provides information about the fluorescence collection efficiency of the confocal microscope which is in turn used to quantify the gene-product images. One of the key concerns in microscopy is deteriorating image quality with increased optical depth into an opaque object. This is of particular concern in confocal microscopy as it results in decreased image brightness along the optical axis. This has been demonstrated in figure 4 which clearly shows that the average brightness per nuclei falls off with increased optical depth. The graph also shows another important aspect of confocal microscopy which has to do with the size of the pinhole though which the image is collected. The pinhole size determines the effective optical thickness within the object from which light is collected. The smaller the pinhole the thinner the optical section but the less light is collected. Figure 4 clearly shows more light is collected from nuclei at the mid plane along the optical axial of the embryo [Figure 6]. However, this is due to the optical collection depth of the microscopy and the embryo geometry because nuclei at the mid plane have other nuclei above and below them, which is not the case for nuclei in the top and bottom planes of the embryo. This effect can be almost eliminated by collecting the images through a smaller pinhole but to the detriment of being able to image nuclei all the way through the embryo.

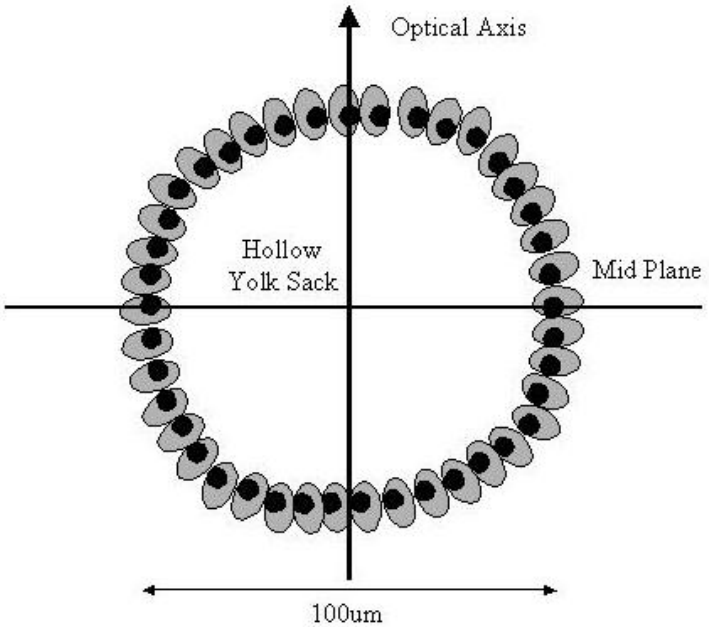

Figure 6: In pregastrula *Drosophila* the first 14 mitotic divisions result in a single layer of nuclei surrounding a yolk sack. This cartoon is of a cross section thought an embryo and demonstrates its geometry relative to the optical axis of the microscope.

This work is part of a multidisciplinary effort at Berkeley to uncovering the rules determining how patterns of gene expression are generated and illustrates the power of quantitative optical bio-imaging. Our goal is to combine 3D expression data at cellular resolution of transcription factors and mRNAs of many genes, at multiple time point during development, to create a database of gene expression in pregastrula embryo.

ACKNOWLEDGEMENTS

This work has been supported by the Director, Office of Energy Research, Office of Health and Environmental Research of the U.S. Department of Energy under contract number DE-AC03-76SF00098, and NIH R01 grant GM42387

REFERENCES

1. Adams et al "The genome sequence of Drosophila melanogaster", *Science* **287**:2185-2195, 2000

2. DeRisi, J.L., Iyer, V.R. and Brown, P.O., "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science*, **278**, 680-686, 1997

3. Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O., "Genomic binding sites of the yeast cell-cycle trnascription factors SBF and MBF", *Nature*, **409**, 533-538, 2001

4. Ross, M.T., LaBrie, S., McPherson, J. and Stanton Jr, V.P. (eds.). "Screening Large-Insert Libraries by Hybridization", Wiley, New York, 1999

5. Liang Z, Biggin M.D., "Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila", *Development* **125**:4471-4482, 1998

6. Kosman, D., Reinitz, J. and Sharp, D.H., "Automated assay of gene expression at cellular resolution", *Pacific Symposium on Biocomputing*, **NA**, 6-17, 1998

7. 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**, 3-12 2001

8. David W. Knowles Mark D. Biggin, Stephen Richards, Damir Sudar "Mapping organism expression levels at cellular resolution in developing Drosophila" *Microsc. Microanal.* **7**: 10-11, 2001

9. David W. Knowles, Soile Keranen, Suman Dasgupta, Mark D. Biggin, Damir Sudar, "Quantitative 3D imaging to analyze developmental variability in pregastrula Drosophila embryos", *Biophysical J.* ***in press*** 2002

10. Soile V. E. Keränen, Mark D. Biggin, Damir Sudar, David W. Knowles, "Measuring morphological variability in pregastrula *Drosophila* embryos", *43rd Annual Drosophila Research Conference* ***in-press*** 2002