Genetics Software

BEST: A Novel Computational Approach for Comparing Gene Expression Patterns From Early Stages of *Drosophila melanogaster* Development

Sudhir Kumar,[*,†,1] **Karthik Jayaraman,**[‡] **Sethuraman Panchanathan,**[*,‡,§]
Rajalakshmi Gurunathan,[*,§] **Ana Marti-Subirana**[**] **and Stuart J. Newfeld**[†]

*Center for Evolutionary Functional Genomics, †Department of Biology, ‡Department of Electrical Engineering, §Department of Computer Science and Engineering, Arizona State University, Tempe, Arizona 85287 and **Phoenix College, Phoenix, Arizona 85013*

ABSTRACT

Embryonic gene expression patterns are an indispensable part of modern developmental biology. Currently, investigators must visually inspect numerous images containing embryonic expression patterns to identify spatially similar patterns for inferring potential genetic interactions. The lack of a computational approach to identify pattern similarities is an impediment to advancement in developmental biology research because of the rapidly increasing amount of available embryonic gene expression data. Therefore, we have developed computational approaches to automate the comparison of gene expression patterns contained in images of early stage *Drosophila melanogaster* embryos (prior to the beginning of germ-band elongation); similarities and differences in gene expression patterns in these early stages have extensive developmental effects. Here we describe a basic expression search tool (BEST) to retrieve best matching expression patterns for a given query expression pattern and a computational device for gene interaction inference using gene expression pattern images and information on the associated genotypes and probes. Analysis of a prototype collection of Drosophila gene expression pattern images is presented to demonstrate the utility of these methods in identifying biologically meaningful matches and inferring gene interactions by direct image content analysis. In particular, the use of BEST searches for gene expression patterns is akin to that of BLAST searches for finding similar sequences. These computational developmental biology methodologies are likely to make the great wealth of embryonic gene expression pattern data easily accessible and to accelerate the discovery of developmental networks.

PATTERNS of gene expression in the fruit fly *Drosophila melanogaster* have been extensively studied by visualizing the presence or absence of gene products or their markers in the developing embryo (visualization methods are reviewed in GOLDSTEIN and FYRBERG 1994). Genetic studies show that genes with similar expression patterns often have mutant alleles that affect the same tissue. In these studies, researchers routinely infer gene interactions by visually comparing gene expression pattern images (*e.g.*, GIESELER *et al.* 2001; TAKAESU *et al.* 2002). For example, the dorsal/ventral polarity of the Drosophila embryo is controlled by many genes, including the secreted factors *decapentaplegic* and *short gastrulation* and the transcription factor *brinker*. *short gastrulation* and *brinker* have very similar expression patterns and their mutant phenotypes are also very similar. Genetic analyses have shown that both of these genes

antagonize *decapentaplegic* signaling (ASHE and LEVINE 1999; JAZWINSKA *et al.* 1999a,b). As this example shows, analysis of similar gene expression patterns is important to understanding the interplay of genes that generate the body plans of fruit flies, humans, and other metazoans (reviewed in CARROLL *et al.* 2000; DAVIDSON 2000; ROUGVIE 2001).

Familiarity with the wealth of images of gene expression patterns gathered over the past two decades is essential for discovering new genetic interactions. However, the burden of becoming familiar with extensive past literature and the rapidly increasing amount of information on gene expression patterns is an impediment to cross-laboratory endeavors and to building a global genetic framework for embryonic development. Computational tools to automatically identify images with similar gene expression patterns from a large collection of images and to predict potential genetic interactions using these images would greatly facilitate developmental biology research. Such tools will become increasingly important with the advent of large-scale *in*

[1]*Corresponding author:* Department of Biology, Life Sciences Center A-351, Arizona State University, Tempe, AZ 85287-1501.
E-mail: s.kumar@asu.edu

situ RNA hybridization studies (http://www.fruitfly.org). The need for a computational system is particularly acute in studies of Drosophila, as scientists move beyond studies of single genes or gene families to generate a global view of development. To address these issues, we have developed *in silico* approaches for automated comparison of gene expression pattern images that mimic some of the visual comparison techniques used by researchers in the laboratory. Our methods enable easy and efficient access to gene expression and interaction data and will likely facilitate new discoveries in developmental biology.

In this article, we describe methodologies for (a) standardizing gene expression pattern images as they are acquired under different illumination conditions and are published in a variety of sizes, orientations, and resolutions; (b) quantifying the amount of dissimilarity between two expression patterns by comparing images containing them; (c) identifying those images from a large collection that contain expression patterns similar to that contained in the given image; and (d) inferring genetic interactions between two genes by comparing the expression pattern of a particular gene in wild-type and mutant backgrounds. In the first phase of our efforts to develop computational methods, we have focused on early embryonic development in *D. melanogaster*. This focus reflects two relevant features of the field of Drosophila developmental genetics. First, there is an emphasis on understanding early developmental events that have global effects, such as the organization of the embryonic dorsal/ventral axis. Second, the deposition of the cuticle after roughly three-fourths of embryonic development prevents the use of several techniques commonly employed for the analysis of gene expression.

We present results showing the performance of our methodologies using a prototype collection of 982 image-searchable gene expression pattern images captured as whole embryos and retrieved from the literature (Gaul and Jackle 1987, 1990; Tautz 1988; Pankratz *et al.* 1989, 1990; Hulskamp *et al.* 1990, 1994; Kania *et al.* 1990; Pankratz and Jackle 1990; Hulskamp and Tautz 1991; Riddihough and Ish-Horowicz 1991; Sommer and Tautz 1991; Steingrimsson *et al.* 1991; Pignoni *et al.* 1992; Gutjahr *et al.* 1993; Lardelli and Ish-Horowicz 1993; Grossniklaus *et al.* 1994; Hartmann *et al.* 1994; Pelegri and Lehmann 1994; Rothe *et al.* 1994; Schulz and Tautz 1994, 1995; Tsai and Gergen 1994; Margolis *et al.* 1995; Rivera-Pomar *et al.* 1995; Sanchez-Herrero 1995; Yu and Pick 1995; Arnosti *et al.* 1996; Klingler *et al.* 1996; Kosman and Small 1997; Vincent *et al.* 1997; Lawrence and Pick 1998; Nibu *et al.* 1998; Toy *et al.* 1998; Tsai *et al.* 1998; Wu *et al.* 1998; Ashe and Levine 1999; Goldstein *et al.* 1999; Jazwinska *et al.* 1999b; La Rosee-Borggreve *et al.* 1999; Niessing *et al.* 1999; Zhang and Levine 1999; Janody *et al.* 2000; Jin *et al.* 2000; Nasiadka *et al.* 2000; Wimmer *et al.* 2000; Casares and Flores-Saaib
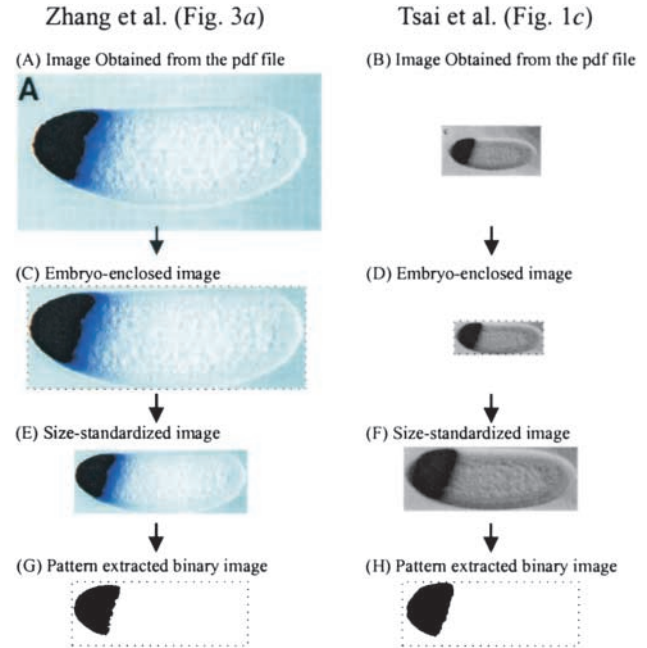

Figure 1.—Gene expression pattern images published by two different groups of investigators (Tsai *et al.* 1998; Zhang and Levine 1999). (A and B) Original gene expression pattern images obtained from the PDF file available online. Relative size differences are due to the differences in sizes of the images as originally published. (C and D) Edge-fitted images with the rectangle bounding the embryo denoted by the dotted line. (E and F) Size-standardized images. (G and H) Expression pattern extracted images.

et al. 2001; Kobayashi *et al.* 2001; Nibu and Levine 2001).

MATERIALS AND METHODS

To meaningfully compare gene expression images computationally, it is important to standardize them and to remove background information (reviewed in Castleman 1996). Standardization is required because investigators publish digital images in different sizes and orientations and because images are acquired under different illumination conditions. Digital images are composed of many fine dots (called pixels) of different intensities. Thus standardization establishes the pixel-to-pixel correspondence between two images by setting up a uniform size and point of reference for all images. For example, two expression patterns in Figure 1 (A and B) are quite similar to the naked eye; however, they do not have the necessary pixel-to-pixel correspondence for computer-based analysis. The gene expression image standardization procedure involves a number of steps.

Embryo-enclosing algorithm: This procedure starts with fitting the Drosophila embryo boundaries into the smallest possible rectangular area, a process referred to as "edge-fitting" (Castleman 1996; Costa and Cesar 2000). For instance, the dotted lines in Figure 1 (C and D) are the smallest rectangles in which the actual embryonic image is contained. Since early stage embryos have a consistent shape, it is an effective way of removing the noise outside the embryo area. It is actually more effective than a contour-detecting algorithm in the pres-

ent case because many images contain embryo boundaries that are too faint to generate a complete contour.

For this step, we have developed an algorithm in which the boundaries of the initial rectangular image are moved inward until they touch the Drosophila embryo outline on all four sides. In this algorithm, we first consider an area of 5×5 pixels at the four corners of the image and then compute the mean (m) and standard error (s) of the average color intensities for pixels in the four corners of the image. Next, we find the topmost row in the image, where the upper boundary of the embryo is present. For this, we traverse top-to-bottom in each column to identify the pixels closest to the top boundary where the absolute difference between the pixel intensity (p) and m is significant (*e.g.*, $m - p$ is greater than twice the standard error under the assumption of normality). The row coordinate of the pixel identified closest to the top edge of the image is the top boundary of the embryo. To identify bottom, left, and right boundaries, we use the same algorithm; the only difference is that now we traverse from bottom to top in each column, left to right in each row, and right to left in each row, respectively. This results in an embryo enclosed in the smallest rectangular area possible. The embryo boundaries determined using this algorithm for A and B in Figure 1 are shown in C and D. Note that all images are standardized to an anterior (left)–posterior (right)–dorsal (top)–ventral (bottom) orientation.

Size standardization: In the next step, we need to scale all the resultant images to the same size as they are often captured or published in very different sizes (compare Figure 1, A and B). We chose a size of 270×100 pixels, which was the average size of gene expression pattern images acquired from the published literature. For scaling, we perform a geometric transformation (simple scaling) followed by an interpolation to derive the pixel values of the new image. In these studies, we used a gray-level bilinear-interpolation scheme (CASTLEMAN 1996, p. 124). It is a first-order interpolation, which determines the destination pixel intensity value based on the four nearest neighbor pixels of the source image. This is a simple but effective approach for scaling. Results from this transformation are shown in Figure 1, E and F, which are standardized images in which pixel correspondence has been established.

Expression pattern extraction: The next step in the gene expression pattern image standardization is to eliminate background to focus on the foreground containing the expression pattern. This ensures that only actual patterns of gene expression are compared for biologically meaningful analysis. In Figure 1A, the gene expression pattern is the darkly stained region embedded in a lighter-color background. While this can be recognized easily by a trained eye, the process needs to be automated for large-scale data gathering. Therefore, the expression pattern (relevant visual content) first must be extracted from each image.

Extraction of the gene expression pattern from the background requires the use of a threshold value of pixel intensity (GONZALEZ and WOODS 1993; CASTLEMAN 1996). All pixels with intensity less than the threshold are assigned a white color (background) and all others are left as is. Our preliminary results demonstrated, as expected, that the same threshold value cannot be used for all images due to differences in the intensity distribution caused by variations in investigator equipment, gene expression pattern, and other factors. To compensate for these variations, we automatically derive threshold values for each image by using adaptive thresholding methodology (see LIE 1995 for details). This can be accomplished, for example, by using the variance of the pixel intensity values for a given image as the basis for the choice of the specific threshold value for that image. While this method captures the entire expression pattern, the resulting image
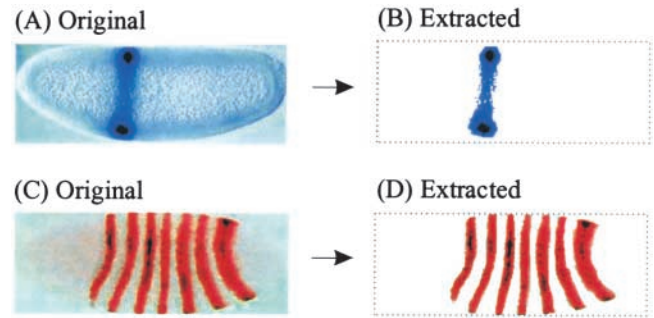

FIGURE 2.—Original and extracted patterns using adaptive thresholding for a simple expression pattern (A and B) and the expression pattern for a pair-rule gene (C and D).

also contains some noise. We therefore filter the images to suppress and/or remove the noise by employing wavelet filters and morphological operators to selectively remove the higher frequency bands contributed by noise (ADAMS and BISCHOF 1994). The result of this extraction process for a simple region of expression (Figure 2A) is shown in Figure 2B. For images containing multiple regions of expression, we employ the region-growing procedure to extract relevant patterns. Region growing is a procedure that groups pixels or subregions into larger regions starting from a seed pixel (or region) and appending the neighboring pixels that have similar properties. This improves data quality and makes the extracted pattern biologically relevant. For instance, an algorithm implementing the region-growing procedure automatically extracted all areas of expression for genes that affect multiple regions (*e.g.*, pair-rule genes) without requiring manual input (Figure 2, C and D). (For multiply stained embryos, use of color-sensitive thresholds corresponding to each stain separately allows for the generation of multiple images.) These methods do not distinguish between quantitative levels of expression; they are meant for identifying spatial similarities in the presence or absence of gene expression.

Finally, we convert the color/gray-scale pixels into simple black-and-white patterns such that only pixels containing gene expression take on a black color and the rest of the image lacks any color (Figure 1, G and H). We do not convert original images directly into black-and-white images prior to conducting the above-mentioned size standardization and pattern extraction procedures because the information contained in differences in color intensities is valuable for reliably separating expression patterns from the background.

Digital representation of expression patterns: The binary images representing gene expression patterns are processed to derive a vector of features describing the image content (gene expression patterns). We represent each image in the form of a string of 1's and 0's, where black pixels are denoted by a value of 1 and white pixels by a value of 0. For example, an expression pattern is represented as 0100111000... 00000001111111. This is referred to as the binary sequence vector (BSV) representation. The BSV representation is particularly useful in quantifying image-to-image dissimilarity for finding images with similar or overlapping expression patterns, and it allows for localization of image similarity searches to any section of the embryo.

An alternative set of features can be derived using shape descriptors, which have proven to be effective in establishing the similarity between two images as well as in effecting image retrieval (COSTA and CESAR 2000). The latest multimedia standard proposed by the International Standards Organization (ISO) for indexing and retrieval of images, namely MPEG7,

proposes shape descriptors for efficient and effective image retrieval. These features are specifically designed for natural images and their utility in the gene expression pattern image analysis is generally unclear. We therefore investigated the usefulness of shape descriptors by adapting them to our specific class of expression pattern images. Our analyses showed that they are not more efficient than the binary sequence vector for the early stage embryos considered here (K. JAYARAMAN, S. PANCHANATHAN and S. KUMAR, unpublished results).

Finding best matching expression patterns with the basic expression search tool (BEST): To find the best matching set of images for a given query image, we need to compute the extent of dissimilarity between gene expression patterns. We use the BSV representation for this purpose. The corresponding bit sequences are compared and the number of bits with different values are counted. We define the expression pattern distance (D_E) to be equal to the number of differences between two images divided by the number of pixels depicting the expression pattern in at least one of the two images. In other words, we determine expression pattern similarity by focusing only on the pixels that show gene expression (have value 1) in either image. For example, two images, A and B, with bit sequences of length 30 given below have nine differences (underlined bits).

Expression Pattern A: 000110011111000000000111111111
Expression Pattern B: 0001100111111110000000111000000

Therefore, $D_E = 9/18 = 0.5$. Images showing the highest amount of match (*i.e.*, lowest D_E) with the query image are retrieved as the best matches. This simple metric showed excellent performance for early stage embryos in an analysis of a collection of 982 images (see below).

Inferring gene interactions: Gene interactions are computed by comparing expression patterns of individual genes in wild-type and mutant backgrounds. Normally, investigators conduct this task by simple visual inspection and deduction. However, as the number of images increases, visual inspection becomes cumbersome. Therefore, we have devised an algorithm that uses gene expression pattern images, their genotypes, and their probes to infer the nature of the interaction for that pair of genes. This computational system works well for early stages of embryonic development because gene expression at these stages is often relatively broad, rather than highly localized as in later stage embryos. Also, we find that >80% of the available images are from early stage embryos. Therefore these methods are applicable for large amounts of existing data.

A flowchart for the two-gene case is shown in Figure 3, in which the expression patterns of gene B in the wild-type and mutant backgrounds for gene A are used. By design, these data shed light on how the expression of gene B is influenced by gene A. Figure 3 shows various paths starting from the gene expression images and proceeding to the decisions that correspond to whether or not the expression of gene B is only positively affected (A → B), only negatively affected (A —|B), positively and negatively affected (A →|B), or not affected (A × B) by gene A. Using the BSV representations of gene expression patterns, the actual algorithm works as follows. We begin by defining a new function, termed the gene expression function (*gef*), which evaluates whether a given expression pattern contains all 0's or all 1's or a combination of 0's and 1's in a given image (*i.e.*, no expression, ubiquitous expression, or localized expression pattern, respectively). It produces 1 if the image contains all black pixels. If all pixels are white, then it produces a 0; otherwise, it returns a P (for partial). To analyze all possible interactions between the gene expression patterns, we derive new image patterns based on the two original images, for which the logical exclusive OR (*xor*) operation
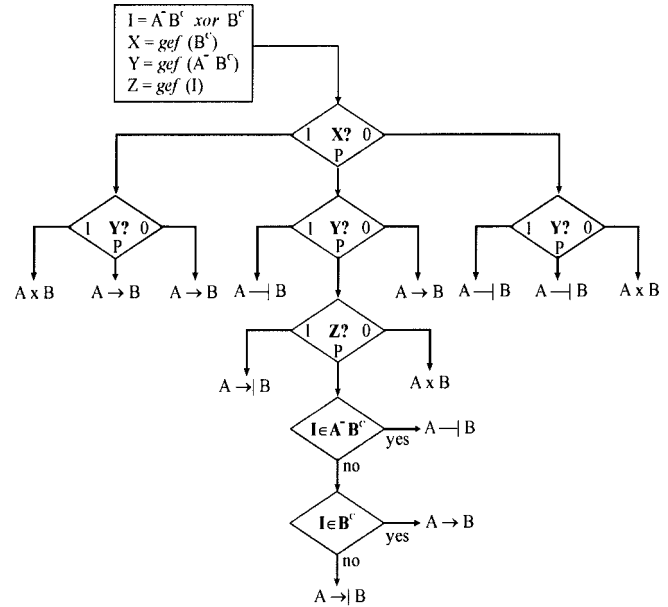

FIGURE 3.—A simple algorithm for inferring gene interaction between genes A and B using gene B's expression pattern images in wild-type and mutant backgrounds for gene A. The wild-type expression pattern B is captured in image B^C and the expression of gene B in A mutant background is captured in image $A^- B^C$. In this image nomenclature, a superscript dash reflects the mutant gene and a superscript C reflects the gene whose expression is captured. Also, the logic expressions containing symbol ∈ test whether the expression pattern in the image on the left is contained completely in the expression pattern of the image on the right.

is used. This operator results in a 0 if the values of the corresponding pixels in the two images under inspection are the same (*i.e.*, if they are both 0 or both 1) or results in a 1 if the values of the corresponding pixels in the two images are different. Outputs from *gef* function for this image as well as the original data are input to the flowchart diagram for inferring interactions.

RESULTS

Performance of data standardization methods: We examined the performance of our methodologies for embryo-enclosing and expression pattern extraction using 97 images from eight articles (LAWRENCE and PICK 1998; TOY *et al.* 1998; TSAI *et al.* 1998; ASHE and LEVINE 1999; GOLDSTEIN *et al.* 1999; JAZWINSKA *et al.* 1999b; ZHANG and LEVINE 1999; JIN *et al.* 2000). We provided the original and the automatic embryo-enclosed images to 16 independent evaluators (graduate students in biology or computer science not involved in the current project) and requested that they give a score of 0.25 for each edge in the edge-fitted image that, in their opinion, enclosed the Drosophila embryo in the tightest fit possible (visually). Therefore, a score of 1.0 refers to perfect fit of all four edges, and a score of 0 represents the poorest fit in all directions. The average edge-fitting score was 0.875; *i.e.*, 3.5 edges fit well on average. While

most edge-fitted images scored perfectly, we identified a few images that scored poorly (score <0.5). This appears to be due to poor or highly varying illumination in the image (*e.g.*, embryo boundaries are too faint in some images) or to artifacts such as panel numbers or arrows. In these difficult cases, we needed to manually set boundaries.

We evaluated the performance of our automatic approach for pattern extraction by again asking 16 independent evaluators to assign a score of 1–4 for each extracted pattern (1 being worst and 4 being best). This tests the performance of the above method to approximate what the human eye can easily recognize. The average success score for the combined region-growing and adaptive-thresholding methodology was found to be 3.5 (87.5%). The observed difficulty with automatic pattern extraction appears to relate to the lack of sufficient contrast between the expression pattern and the background. In addition, in the presence of multiple patterns of interest (with different illuminations) the automatic technique needs to be user guided to extract all biologically significant regions. For this reason, a semiautomatic system was developed in which different threshold values and/or new seed points for different regions of interest in the embryo can be specified manually. This improves data quality and makes the extraction biologically relevant.

Finding similar gene expression patterns computationally: We explored the effectiveness of our computational approach in finding similar expression patterns using a prototype collection of 982 patterns from 49 published research articles (Figure 4). These tests were geared toward determining the sensitivity of our computational approach and the biological validity of the image-matching procedure. A computer program was written to automatically retrieve gene expression pattern images showing overlap with a given query pattern based solely on the similarities between images. We refer to the images retrieved by BEST as "BEST hits."

Here we briefly summarize the characteristics of our 10 test-case expression patterns and then we describe BEST-hit results for each case. Of the 10 patterns, 9 are lateral views (Figure 4, A–E, G, and H) and one is a dorsal view (F). Expression patterns in some images differ considerably from each other (*e.g.*, Figure 4, A and B), some differ only slightly (Figure 4, B and C), some have clearly overlapping domains of expression even though they are quite different overall (Figure 4, D and E and G and I), and others have multiple distinct regions of expression (Figure 4, E, F, I, and J). These query images were chosen without regard to their genotype, probe, or any other consideration except for diversity of expression pattern. Also, BEST hits were based solely on gene expression contained in the images; no prior knowledge of gene interactions was used. For each image, the program queried the entire collection of 982 images. The top five BEST hits for each image are shown

in descending order below the query image in Figure 4 with the percentage of similarity $(1 - D_E)$ given in parentheses.

Figure 4A shows an image with expression restricted to the posterior 10% of the embryo (ZHANG and LEVINE 1999). Specifically, the image shows the expression of *forkhead* RNA in a transgenic embryo with maternal expression of a mutant *hairy* cDNA fused to the *bicoid* 3′ untranslated region (UTR). With this as the query image, BEST should retrieve images with posterior restricted expression patterns, especially other panels in the original article. This is indeed the case. Four of the five BEST hits are other panels in the original article. These panels depict *tailless*, *huckebein*, and *forkhead* RNA expression in transgenic embryos with maternal expression of different mutant *hairy* cDNA fused to the *bicoid* 3′ UTR. In addition, the BEST hits contain an image with expression of *brachyenteron* RNA in a *huckebein* mutant embryo (GOLDSTEIN *et al.* 1999).

Figure 4B shows an image with expression restricted to the anterior 25% of the embryo (ZHANG and LEVINE 1999). Here, the expression of *hairy* RNA in a transgenic embryo with maternal expression of a *hairy* cDNA fused to the *bicoid* 3′ UTR is captured. Three of the five BEST hits to this image are other panels in the original article; they depict *hairy* RNA expression in transgenic embryos with maternal expression of different mutant *hairy* cDNA fused to the *bicoid* 3′ UTR. An image from HULSKAMP and TAUTZ (1991), which depicts *bicoid* RNA expression in a wild-type embryo, is also identified. The *bicoid* RNA localization is regulated by sequences in its 3′ UTR (MACDONALD 1990); thus the identification of *bicoid* when using a query image in which gene expression is regulated by *bicoid* 3′ UTR sequences is clearly meaningful. An image from FLORES-SAAIB *et al.* (2001) depicting the expression of Gal4-driven β-galactosidase (lacZ) RNA in a transgenic embryo expressing the C-terminal domain of Dorsal fused to Gal4 was also among the five BEST hits. Results for Figure 4, A and B, clearly show that our search tool effectively retrieves biologically relevant images that match a query image with a simple expression pattern.

Figure 4C shows an image with expression of *orthodenticle* RNA restricted to the anterior 25% of the embryo, except that expression is absent at the anterior terminus (TSAI and GERGEN 1994). Four of the five BEST hits are images with an anterior expression pattern in which expression is also absent at the anterior terminus. Three of these are from another article by the same author (TSAI *et al.* 1998) and depict *orthodenticle* RNA expression in different genotypes. An image from GROSSNIKLAUS *et al.* (1994) depicting the expression of *sloppy paired1* RNA in a wild-type embryo was also among the five BEST hits. The least similar of the five BEST hits is an image with expression throughout the anterior region of the embryo from ZHANG and LEVINE (1999) showing the expression of *hairy* RNA in a transgenic embryo with
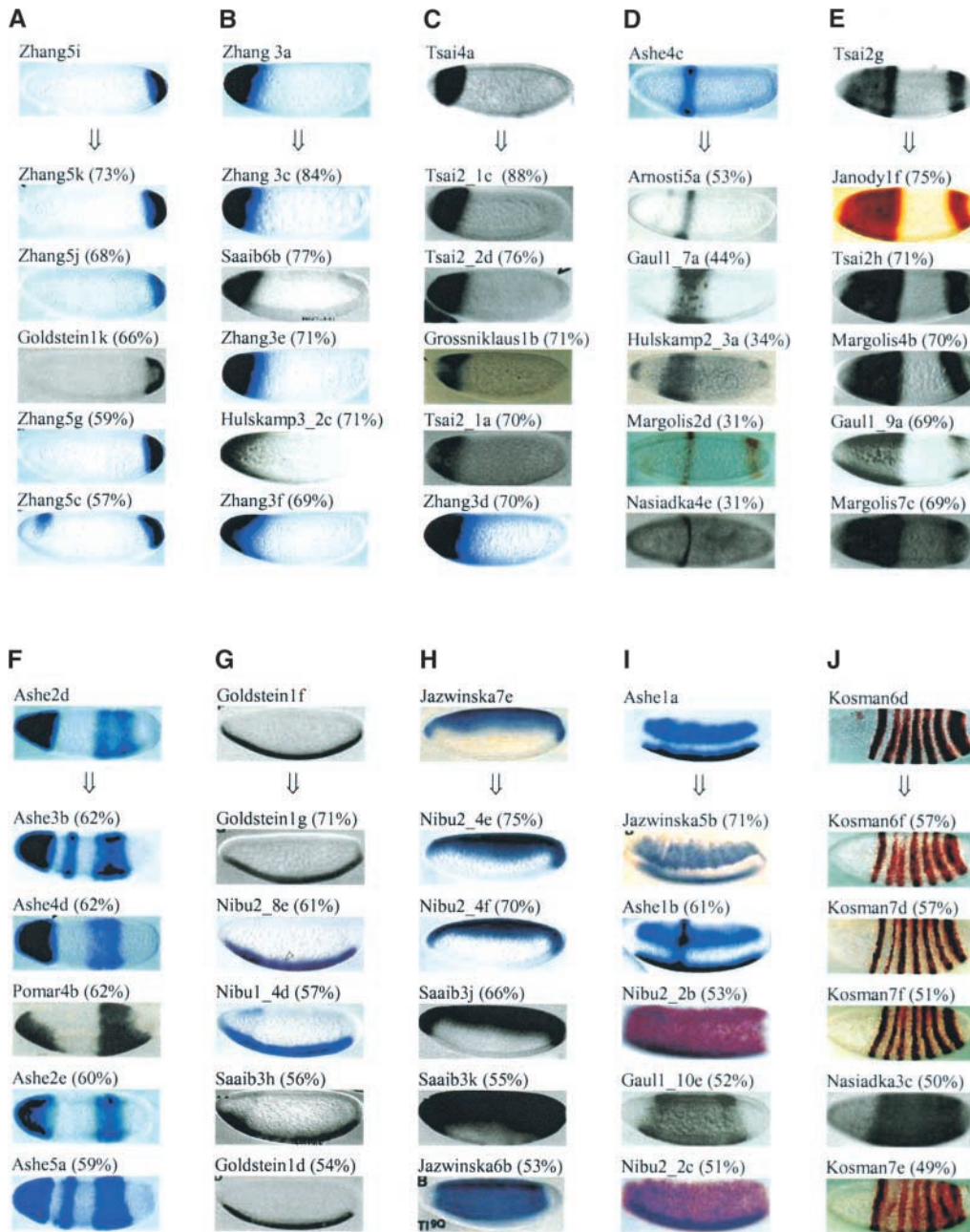
FIGURE 4.—BEST-hit results for a variety of query images. Each image is identified by the last name of the first author of the original research article and the figure number for that article. Values in parentheses are the percentage of similarity score $(1 - D_E)$. BEST hits are arranged in descending order starting with the best hit. Author names and citations are as follows: Zhang (ZHANG and LEVINE 1999); Goldstein (GOLDSTEIN *et al.* 1999); Saaib (FLORES-SAAIB *et al.* 2001); Hulskamp3 (HULSKAMP and TAUTZ 1991); Tsai (TSAI and GERGEN 1994); Tsai2 (TSAI *et al.* 1998); Grossniklaus (GROSSNIKLAUS *et al.* 1994); Ashe (ASHE and LEVINE 1999); Arnosti (ARNOSTI *et al.* 1996); Gaul1 (GAUL and JACKLE 1990); Hulskamp2 (HULSKAMP *et al.* 1990); Margolis (MARGOLIS *et al.* 1995); Nasiadka (NASIADKA *et al.* 2000); Janody (JANODY *et al.* 2000); Pomar (RIVERA-POMAR *et al.* 1995); Nibu1 (NIBU and LEVINE 2001); Nibu2 (NIBU *et al.* 1998); Jazwinska (JAZWINSKA *et al.* 1999b); Kosman (KOSMAN and SMALL 1997).

maternal expression of a *hairy* cDNA fused to the *bicoid* 3′ UTR. Notably, even though the query image in Figure 4C is quite similar to the query image in Figure 4B, the top-five BEST hits for these query images are not the same. This example demonstrates the sensitivity of the image search system.

Figure 4D shows an image with a single stripe of expression of *short gastrulation* RNA driven by the *even skipped* stripe 2 enhancer in a transgenic embryo (ASHE and LEVINE 1999). The top five BEST hits include RNA and protein expression patterns. The top hit is an image showing Gal4-driven lacZ RNA in an embryo expressing a Knirps-Gal4 fusion protein under the control of the *even skipped* stripe 2 enhancer (ARNOSTI *et al.* 1996).

The next two BEST hits are images of Krüppel protein expression in a *hunchback* mutant embryo (GAUL and JACKLE 1990) and Krüppel RNA expression in a *bicoid* and *hunchback* double-mutant embryo (HULSKAMP *et al.* 1990). The next BEST hit is an image of *hunchback* RNA expression in a wild-type embryo (MARGOLIS *et al.* 1995). The *engrailed* RNA expression in an embryo with ubiquitous *fushi tarazu* expression (NASIADKA *et al.* 2000) rounds out the top five. While the query image and the BEST hits are from six different articles, all of the genes involved are part of a well-characterized, spatially localized hierarchy of gene regulation. The query image and the top BEST hits depict expression of a transgene containing the stripe 2 enhancer that normally directs

the expression of the pair-rule gene *even skipped*. As reviewed by Gaul and Jackle (1990) *even skipped* expression is regulated by the gap genes *hunchback* and *Krüppel* (BEST hits 2, 3, and 4). Finally, the segment polarity pattern of *engrailed* expression (BEST hit 5) is regulated by the pair-rule gene *even skipped* (Fujioka *et al.* 1995). This example demonstrates the use of the expression search system in retrieving data that help in understanding components of developmental pathways.

Figure 4E shows an image of the two domains of *hunchback* RNA expression in a wild-type embryo. *hunchback* is expressed in a wide stripe toward the anterior and a narrow stripe toward the posterior of the embryo (Tsai and Gergen 1994). All five BEST hits are images showing *hunchback* expression (visualized in a variety of ways) in wild-type or mutant backgrounds. The five BEST hits include overlapping expression patterns from the original article and from three other articles (Gaul and Jackle 1990; Margolis *et al.* 1995; Janody *et al.* 2000).

Figure 4F shows an image, in dorsal view, with three domains of expression. Expression is seen in the anterior 25% and in two stripes of expression toward the posterior of the embryo. Specifically, the image shows *Race* RNA expression in a *goosecoid* mutant embryo with *short gastrulation* expression driven by the *even skipped* stripe 2 enhancer (Ashe and Levine 1999). Four of the five BEST hits are images from the same article showing dorsal views of *Race* or *Race* and *short gastrulation* RNA expression in *goosecoid* mutant embryos with *short gastrulation* expression driven by the *even skipped* stripe 2 enhancer. An image, in lateral view, showing the expression of *knirps* RNA in a maternal and zygotic *caudal* mutant embryo, is also identified (Rivera-Pomar *et al.* 1995).

Figure 4G shows an image with expression restricted to the ventral 10% of the embryo. The expression of *snail* RNA in an embryo derived from a *groucho* germline clone is shown (Goldstein *et al.* 1999). Two of the five BEST hits are images from the same article showing *snail* RNA expression in wild-type and *huckebein* mutant embryos. In addition, two images of lacZ RNA expression driven by a fusion of *rhomboid* and *twist* enhancers are identified (Nibu *et al.* 1998; Nibu and Levine 2001). The identification of *twist* expression when using a query image of *snail* expression is biologically meaningful. Both *twist* and *snail* are expressed in the ventral ectoderm and are important for ventral mesoderm formation (Ray *et al.* 1991). The expression of Gal4-driven lacZ RNA in a transgenic embryo expressing Dorsal fused to Gal4 is also retrieved in this search (Flores-Saaib *et al.* 2001).

The expression of *decapentaplegic* RNA is restricted to the dorsal 40% of a wild-type embryo as shown in Figure 4H (Jazwinska *et al.* 1999b). One of the five BEST hits is an image from the same article showing *brinker* RNA expression in an embryo derived from a female bearing a dominant *Toll* mutant allele. Two images of lacZ RNA driven by wild-type and modified versions of the *decapentaplegic* dorsal enhancer (Flores-Saaib *et al.* 2001) in wild-type embryos are also identified. The other two top hits are *zernault* RNA expression in wild-type and *dCtBP* mutants (Nibu *et al.* 1998). In this case, the query and all five BEST hits depict the expression of genes involved in direct interactions with Decapentaplegic signaling during embryonic dorsal/ventral patterning.

Figure 4I shows an image with stripes of expression in ventral and lateral regions. Specifically, the image shows the expression of *short gastrulation* in a wild-type embryo (Ashe and Levine 1999). The top BEST hit is an image of *brinker* RNA expression in a wild-type embryo (Jazwinska *et al.* 1999b). The fact that the top hit for *short gastrulation* RNA expression is *brinker* is consistent with the roles of *brinker* and *short gastrulation* in the Decapentaplegic signaling pathway (Jazwinska *et al.* 1999a,b). One of the five BEST hits is an image from Ashe and Levine (1999) showing the expression of *short gastrulation* RNA in a transgenic embryo in which *short gastrulation* is also expressed from the *even skipped* stripe 2 enhancer. Two images of *rhomboid* RNA expression in a *dCtBP* and a *snail* mutant embryo are also identified (Nibu *et al.* 1998). All of these genes are involved in embryonic dorsal/ventral patterning. Even skipped protein expression in an embryo derived from a *bicoid oskar* double-mutant female is also identified (Gaul and Jackle 1990).

Finally, we consider an embryo with seven stripes of expression, a pattern typical of pair-rule genes, for finding BEST hits (Figure 4J). Figure 4J shows the expression of *runt* RNA in an embryo expressing low levels of *knirps* under the control of the *even skipped* stripe 2 enhancer (Kosman and Small 1997). Four of the five BEST hits are images from the same article. The top BEST hit is an image showing the expression of *runt* RNA in an embryo expressing intermediate levels of *knirps* under the control of the *even skipped* stripe 2 enhancer. The other three BEST hits from this article show the expression of *fushi tarazu* RNA in embryos expressing low or intermediate levels of *knirps* under the control of the *even skipped* stripe 2 enhancer. *fushi tarazu* is a pair-rule gene expressed in the same pattern as *even skipped* (Lawrence and Johnston 1989). An image of *fushi tarazu* RNA expression in an embryo ubiquitously expressing a Fushi tarazu-VP16 fusion protein is also identified (Nasiadka *et al.* 2000). In this image, the *fushi tarazu* expression pattern appears continuous due to its reduced size. In the original figure (as published in Nasiadka *et al.* 2000) it is possible to distinguish *fushi tarazu*'s normal striped pattern above the background of ectopic *fushi tarazu* expression induced by the fusion protein.

Overall these results suggest that the basic expression search tool is (i) able to successfully retrieve expression patterns of the same gene from the same article, (ii)

sensitive to relatively small changes in expression pattern, (iii) able to retrieve expression patterns of different genes that have similar functions (*e.g.*, dorsal/ventral patterning), and (iv) able to recover expression patterns of different genes within a spatially localized regulatory hierarchy (*e.g.*, gap, pair-rule, and segment polarity genes). Our examples identify numerous known, biologically meaningful matches (*e.g.*, *short gastrulation* and *brinker*) and several potentially new matches (*e.g.*, *orthodenticle* and *sloppy paired1*) from a large data set. This is exactly what a researcher would do when comparing their images to those in the published literature manually by eye. The use of BEST hits will expedite the search for finding comparable RNA and protein expression patterns (as is the case for the BLAST search for molecular sequences). As when searching the literature for matching gene expression patterns, a BEST-hits user must determine the biological meaningfulness of matches retrieved by consulting the original articles.

Inferring gene interactions: Here we present results from five examples in which genetic interactions were computed employing the flowchart diagram in Figure 3. Our algorithm compares the expression pattern of a gene in wild-type and mutant embryos. In each case, inferred interactions are expected to match those that an investigator would deduce on the basis of their visual inspection of the same images.

Figure 5A shows the effect of *bicoid* on *Krüppel* (*Kr*) expression. On the left, we show *Krüppel* RNA expression in a wild-type embryo (Hulskamp *et al.* 1990) and on the right we show Krüppel protein expression in a maternal *bicoid* mutant (*bcd⁻*) embryo (Gaul and Jackle 1990). Our algorithm compares the pattern of Krüppel expression in each image and determines that there is more extensive Krüppel expression in the *bcd⁻* embryo than in the wild-type embryo. The program suggests that Krüppel expression is negatively affected by *bicoid* activity, when we consider the extent of the region of gene expression. This is consistent with the reported refinement of Krüppel expression by *bicoid*-mediated repression (Gaul and Jackle 1990; Hulskamp *et al.* 1990).

Figure 5B shows the effect of *bicoid* on *tailless* (*tll*) expression by comparing *tailless* RNA expression in wild-type and *bcd⁻* embryos (Pignoni *et al.* 1992). The algorithm suggests that *tailless* expression is also negatively affected by *bicoid* activity. This is consistent with visual comparison of these images as reported by Pignoni *et al.* (1992).

In Figure 5C, we examine the interaction between *caudal* and *knirps* (*kni*). For this purpose, we use images containing *knirps* RNA expression in a wild-type embryo (Arnosti *et al.* 1996) and in a maternal and zygotic *caudal* mutant (*cad⁻*) embryo (Rivera-Pomar *et al.* 1995). The algorithm determines that there is more extensive *knirps* expression in the *cad⁻* embryo than in the wild-type embryo and concludes that there is a
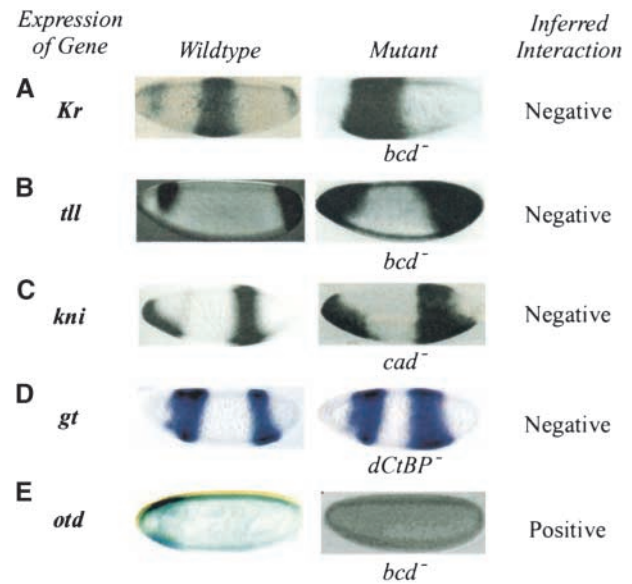

FIGURE 5.—Application of the algorithm for inferring gene interaction between genes using expression patterns in wild-type and mutant backgrounds. (A) Effect of *bicoid* (*bcd*) on *Krüppel* (*Kr*) expression (Gaul and Jackle 1990; Hulskamp *et al.* 1990). (B) Effect of *bcd* on *tailless* (*tll*) expression (Pignoni *et al.* 1992). (C) Effect of *caudal* (*cad*) on *knirps* (*kni*) expression (Rivera-Pomar *et al.* 1995; Arnosti *et al.* 1996). (D) Effect of *dCtBP* on *giant* (*gt*) expression (Nibu *et al.* 1998). (E) Effect of *bcd* on *orthodenticle* (*otd*) expression (Tsai *et al.* 1998; Janody *et al.* 2000).

negative relationship between *caudal* activity and *knirps* expression. This is consistent with the Rivera-Pomar *et al.* (1995) report that *giant*, a *caudal*-dependent repressor, refines the expression of *knirps*.

The posterior refinement of *giant* expression by *dCtBP*-mediated repression is shown in the next example (Figure 5D). Here *giant* RNA expression in a wild-type embryo and a *dCtBP* germline clone mutant embryo (*dCtBP⁻*; Nibu *et al.* 1998) are compared. The algorithm suggests that *giant* expression is negatively affected by *dCtBP* activity as reported by Nibu *et al.* (1998).

Finally, we consider the effect of *bicoid* on *orthodenticle* (*otd*) expression (Figure 5E). The algorithm compares the expression patterns of Orthodenticle protein in a wild-type embryo (Janody *et al.* 2000) and *orthodenticle* RNA in a maternal *bicoid* mutant (*bcd⁻*) embryo (Tsai *et al.* 1998; Janody *et al.* 2000). Our algorithm suggests that *orthodenticle* expression is positively affected by *bicoid* activity, because there is less extensive *orthodenticle* expression in the *bcd⁻* embryo than in the wild-type embryo. This result is consistent with the known activation of *orthodenticle* expression by *bicoid* (Tsai *et al.* 1998).

These examples illustrate the ability of our computational techniques to infer genetic interactions using image data in a manner similar to that employed visually by researchers. The simplicity of the approach is that the entire process can be described by a few logic functions and is therefore easily automated for analyzing all

available pairs of images that are suitable for inferring genetic interactions (results from those analyses will be published elsewhere).

DISCUSSION

A major problem facing biologists today is that data generation has consistently outpaced advancements in computational techniques for developing biological insights and generating hypotheses. This problem has plagued molecular sequence data for over a decade. With the advent of large-scale high-throughput *in situ* efforts, we are headed toward biological data overload in developmental biology as well. For example, Berkeley Drosophila Genome Project has released a large collection of images (>30,000) generated with a standardized RNA *in situ* hybridization method using whole embryos and cDNA probes from the Drosophila UNIGENE set (http://www.fruitfly.org). About 15,000 of these images are expected to depict gene expression from early stage embryos. These data are released in raw form to the public. How will we view these data efficiently? How will we find images with similar expression patterns, other than by manually browsing through each and every image? How will we compare the newly acquired knowledge to all existing knowledge published in the literature? How will the images be standardized? How can we use all the data to generate a global view of the early stages of Drosophila embryonic development? No computational frameworks that will facilitate efficient analyses of these new images or integrate these new images with existing image data from the literature exist.

We have tackled this challenge by developing methods and algorithms for standardization, extraction, and similarity search of gene expression patterns. We have demonstrated their utility in finding similar and overlapping expression patterns for a wide variety of genes from early stages of Drosophila embryonic development. These bioinformatic methodologies are likely to pave the way for establishing a public computational resource to find similar expression patterns based on the visual content of the query image, a facility that will parallel the BLAST search (Altschul *et al.* 1990) for finding homologous sequences. It will make the great wealth of embryonic gene expression pattern data easily accessible. While we have focused on methods for early stage embryos here, these methods are adaptable for analysis of images capturing later stages as well as imaginal disc gene expression. Toward this end, our initial results are encouraging and are in preparation for publication elsewhere.

We have also proposed a method for elucidating genetic interactions computationally by comparing expression patterns of a given gene in wild-type and mutant backgrounds. This method will provide researchers for the first time with a tool to exhaustively examine every possible pair of genes for which wild-type and mutant expression patterns are available. It therefore will facilitate construction of developmental networks and help explore the combinatorial nature of genetic interactions that provide the exquisite specificity and diversity of gene expression patterns. These algorithms provide specific predictions for researchers, who can then test the validity of these *in silico* inferences. In the future, we plan to develop similar algorithms for simultaneous analysis of multiple genes.

In summary, methods presented in this article are efforts to fulfill the need for computational technology for large-scale analysis of gene expression data. For instance, our approach will facilitate the construction of similarity-based clusters of expression pattern images. Examination of similarities and differences in their spatial and temporal contexts and of the genetic background of images within each similarity cluster will likely reveal new genetic interactions and thus accelerate the discovery of hidden links in developmental networks. Therefore, an accelerated, expanded understanding of the interconnected nature of gene function may emerge from *in silico* analysis.

LITERATURE CITED

Adams, R., and L. Bischof, 1994 Seeded region growing. IEEE Transactions on Patterns Analysis and Mach. Intell. **16:** 641–647.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Arnosti, D. N., S. Gray, S. Barolo, J. Zhou and M. Levine, 1996 The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. EMBO J. **15:** 3659–3666.

Ashe, H. L., and M. Levine, 1999 Local inhibition and long-range enhancement of Dpp signal transduction by Sog. Nature **398:** 427–431.

Carroll, S. B., J. K. Grenier and S. D. Weatherbee, 2000 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.* Blackwell Science, Oxford.

Casares, F., and E. Sanchez-Herrero, 1995 Regulation of the infra-abdominal regions of the bithorax complex of Drosophila by gap genes. Development **121:** 1855–1866.

Castleman, K. R., 1996 *Digital Image Processing.* Prentice-Hall, Englewood Cliffs, NJ.

Costa, L. F., and R. M. Cesar, 2000 *Shape Analysis and Classification: Theory and Practice.* CRC Press, Cleveland.

Davidson, E., 2000 *Genomic Regulatory Systems: Development and Evolution.* Academic Press, New York.

Flores-Saaib, R. D., S. Jia and A. J. Courey, 2001 Activation and repression by the C-terminal domain of Dorsal. Development **128:** 1869–1879.

Fujioka, M., J. B. Jaynes and T. Goto, 1995 Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. Development **121:** 4371–4382.

GAUL, U., and H. JACKLE, 1987 Pole region-dependent repression of the Drosophila gap gene Krüppel by maternal gene products. Cell **51**: 549–555.

GAUL, U., and H. JACKLE, 1990 Role of gap genes in early *Drosophila* development, pp. 239–275 in *Genetic Regulatory Hierarchies in Development*, edited by T. R. F. WRIGHT. Academic Press, San Diego.

GIESELER, K., E. WILDER, M. C. MARIOL, M. BURATOVITCH, H. BERENGER *et al.*, 2001 DWnt4 and wingless elicit similar cellular responses during imaginal development. Dev. Biol. **232**: 339–350.

GOLDSTEIN, L. S. B., and E. A. FYRBERG (Editors), 1994 *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology*. Academic Press, San Diego.

GOLDSTEIN, R. E., G. JIMENEZ, O. COOK, D. GUR and Z. PAROUSH, 1999 Huckebein repressor activity in *Drosophila* terminal patterning is mediated by Groucho. Development **126**: 3747–3755.

GONZALEZ, R., and R. WOODS, 1993 *Digital Image Processing*. Addison-Wesley, Reading, MA.

GROSSNIKLAUS, U., K. M. CADIGAN and W. J. GEHRING, 1994 Three maternal coordinate systems cooperate in the patterning of the Drosophila head. Development **120**: 3155–3171.

GUTJAHR, T., E. FREI and M. NOLL, 1993 Complex regulation of early paired expression: initial activation by gap genes and pattern modulation by pair-rule genes. Development **117**: 609–623.

HARTMANN, C., H. TAUBERT, H. JACKLE and M. J. PANKRATZ, 1994 A two-step mode of stripe formation in the Drosophila blastoderm requires interactions among primary pair rule genes. Mech. Dev. **45**: 3–13.

HULSKAMP, M., and D. TAUTZ, 1991 Gap genes and gradients—the logic behind the gaps. Bioessays **13**: 261–268.

HULSKAMP, M., C. PFEIFLE and D. TAUTZ, 1990 A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Krüppel and knirps in the early Drosophila embryo. Nature **346**: 577–580.

HULSKAMP, M., W. LUKOWITZ, A. BEERMANN, G. GLASER and D. TAUTZ, 1994 Differential regulation of target genes by different alleles of the segmentation gene hunchback in Drosophila. Genetics **138**: 125–134.

JANODY, F., J. REISCHL and N. DOSTATNI, 2000 Persistence of Hunchback in the terminal region of the Drosophila blastoderm embryo impairs anterior development. Development **127**: 1573–1582.

JAZWINSKA, A., N. KIROV, E. WIESCHAUS, S. ROTH and C. RUSHLOW, 1999a The Drosophila gene brinker reveals a novel mechanism of Dpp target gene regulation. Cell **96**: 563–573.

JAZWINSKA, A., C. RUSHLOW and S. ROTH, 1999b The role of *brinker* in mediating the graded response to Dpp in early *Drosophila* embryo. Development **126**: 3323–3334.

JIN, S., S. MARINKE, W. S. JOO, J. R. WORTMAN, N. MIRKOVIC *et al.*, 2000 Identification and characterization of a p53 homologous in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **97**: 7301–7306.

KANIA, M. A., A. S. BONNER, J. B. DUFFY and J. P. GERGEN, 1990 The Drosophila segmentation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Genes Dev. **4**: 1701–1713.

KLINGLER, M., J. SOONG, B. BUTLER and J. P. GERGEN, 1996 Disperse versus compact elements for the regulation of runt stripes in Drosophila. Dev. Biol. **177**: 73–84.

KOBAYASHI, M., R. E. GOLDSTEIN, M. FUJIOKA, Z. PAROUSH and J. B. JAYNES, 2001 Groucho augments the repression of multiple Even skipped target genes in establishing parasegment boundaries. Development **128**: 1805–1815.

KOSMAN, D., and S. SMALL, 1997 Concentration-dependent patterning by an ectopic expression domain of the Drosophila gap gene knirps. Development **124**: 1343–1354.

LARDELLI, M., and D. ISH-HOROWICZ, 1993 Drosophila hairy pair-rule gene regulates embryonic patterning outside its apparent stripe domains. Development **118**: 255–266.

LA ROSEE-BORGGREVE, A., T. HADER, D. WAINWRIGHT, F. SAUER and H. JACKLE, 1999 hairy stripe 7 element mediates activation and repression in response to different domains and levels of Krüppel in the Drosophila embryo. Mech. Dev. **89**: 133–140.

LAWRENCE, P. A., and P. JOHNSTON, 1989 Analysis of function of the pair-rule genes *hairy, even-skipped* and *fushi tarazu* in mosaic Drosophila embryos. Development **107**: 847–853.

LAWRENCE, P. A., and L. PICK, 1998 How does the *fushi tarazu* gene activate *engrailed* in the *Drosophila* embryo? Dev. Genet. **23**: 28–34.

LIE, W.-N., 1995 Automatic target segmentation by locally adaptive image thresholding. IEEE Transactions on Image Processing **4**: 1036–1041.

MACDONALD, P. M., 1990 bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure. Development **110**: 161–171.

MARGOLIS, J. S., M. L. BOROWSKY, E. STEINGRIMSSON, C. W. SHIM, J. A. LENGYEL *et al.*, 1995 Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. Development **121**: 3067–3077.

NASIADKA, A., A. GRILL and H. M. KRAUSE, 2000 Mechanisms regulating target gene selection by the homeodomain-containing protein Fushi tarazu. Development **127**: 2965–2976.

NIBU, Y., and M. S. LEVINE, 2001 CtBP-dependent activities of the short-range Giant repressor in the Drosophila embryo. Proc. Natl. Acad. Sci. USA **98**: 6204–6208.

NIBU, Y., H. ZHANG, E. BAJOR, S. BAROLO, S. SMALL *et al.*, 1998 dCtBP mediates transcriptional repression by Knirps, Krüppel and Snail in the Drosophila embryo. EMBO J. **17**: 7009–7020.

NIESSING, D., N. DOSTATNI, H. JACKLE and R. RIVERA-POMAR, 1999 Sequence interval within the PEST motif of Bicoid is important for translational repression of caudal mRNA in the anterior region of the Drosophila embryo. EMBO J. **18**: 1966–1973.

PANKRATZ, M. J., and H. JACKLE, 1990 Making stripes in the Drosophila embryo. Trends Genet. **6**: 287–292.

PANKRATZ, M. J., M. HOCH, E. SEIFERT and H. JACKLE, 1989 Krüppel requirement for knirps enhancement reflects overlapping gap gene activities in the Drosophila embryo. Nature **341**: 337–340.

PANKRATZ, M. J., E. SEIFERT, N. GERWIN, B. BILLI, U. NAUBER *et al.*, 1990 Gradients of Krüppel and knirps gene products direct pair-rule gene stripe patterning in the posterior region of the Drosophila embryo. Cell **61**: 309–317.

PELEGRI, F., and R. LEHMANN, 1994 A role of polycomb group genes in the regulation of gap gene expression in Drosophila. Genetics **136**: 1341–1353.

PIGNONI, F., E. STEINGRIMSSON and J. A. LENGYEL, 1992 bicoid and the terminal system activate tailless expression in the early Drosophila embryo. Development **115**: 239–251.

RAY, R. P., K. ARORA, C. NUSSLEIN-VOLHARD and W. M. GELBART, 1991 The control of cell fate along the dorsal-ventral axis of the Drosophila embryo. Development **113**: 35–54.

RIDDIHOUGH, G., and D. ISH-HOROWICZ, 1991 Individual stripe regulatory elements in the Drosophila hairy promoter respond to maternal, gap, and pair-rule genes. Genes Dev. **5**: 840–854.

RIVERA-POMAR, R., X. LU, N. PERRIMON, H. TAUBERT and H. JACKLE, 1995 Activation of posterior gap gene expression in the Drosophila blastoderm. Nature **376**: 253–256.

ROTHE, M., E. A. WIMMER, M. J. PANKRATZ, M. GONZALEZ-GAITAN and H. JACKLE, 1994 Identical transacting factor requirement for knirps and knirps-related gene expression in the anterior but not in the posterior region of the Drosophila embryo. Mech. Dev. **46**: 169–181.

ROUGVIE, A. E., 2001 Control of developmental timing in animals. Nat. Rev. Genet. **2**: 690–701.

SCHULZ, C., and D. TAUTZ, 1994 Autonomous concentration-dependent activation and repression of Krüppel by hunchback in the Drosophila embryo. Development **120**: 3043–3049.

SCHULZ, C., and D. TAUTZ, 1995 Zygotic caudal regulation by hunchback and its role in abdominal segment formation of the Drosophila embryo. Development **121**: 1023–1028.

SOMMER, R., and D. TAUTZ, 1991 Segmentation gene expression in the housefly Musca domestica. Development **113**: 419–430.

STEINGRIMSSON, E., F. PIGNONI, G. J. LIAW and J. A. LENGYEL, 1991 Dual role of the Drosophila pattern gene tailless in embryonic termini. Science **254**: 418–421.

TAKAESU, N. T., A. N. JOHNSON, O. H. SULTANI and S. J. NEWFELD, 2002 Combinatorial signaling by an unconventional Wg pathway and the Dpp pathway requires Nejire (CBP/p300) to regulate dpp expression in posterior tracheal branches. Dev. Biol. **247**: 225–236.

TAUTZ, D., 1988 Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. Nature **332**: 281–284.

TOY, J., J.-M. YANG, G. S. LEPPERT and O. H. SUNDIN, 1998 The *Optx2* homeobox is expressed in early precursors of the eye and

activates retina-specific genes. Proc. Natl. Acad. Sci. USA **95:** 10643–10648.

Tsai, C., and J. P. Gergen, 1994 Gap gene properties of the pair-rule gene runt during Drosophila segmentation. Development **120:** 1671–1683.

Tsai, C. C., S. G. Kramer and J. P. Gergen, 1998 Pair-rule gene runt restricts orthodenticle expression to the presumptive head of the Drosophila embryo. Dev. Genet. **23:** 35–44.

Vincent, A., J. T. Blankenship and E. Wieschaus, 1997 Integration of the head and trunk segmentation systems controls cephalic furrow formation in Drosophila. Development **124:** 3747–3754.

Wimmer, E. A., A. Carleton, P. Harjes, T. Turner and C. Desplan,

2000 Bicoid-independent formation of thoracic segments in Drosophila. Science **287:** 2476–2479.

Wu, X., R. Vakani and S. Small, 1998 Two distinct mechanisms for differential positioning of gene expression borders involving the Drosophila gap protein giant. Development **125:** 3765–3774.

Yu, Y., and L. Pick, 1995 Non-periodic cues generate seven ftz stripes in the Drosophila embryo. Mech. Dev. **50:** 163–175.

Zhang, H., and M. Levine, 1999 Groucho and dCtBP mediate separate pathways of transcriptional repression in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA **96:** 535–540.