Developmental Biology 357 (2011) 532–540

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Genomes and Developmental Control

High-resolution, three-dimensional mapping of gene expression using GeneExpressMap (GEM)

C.J. Flynn [a,b], T. Sharma [a], S.W. Ruffins [c], S.L. Guerra [a], J.C. Crowley [a,b], C.A. Ettensohn [a,*]

[a] Department of Biological Sciences, Carnegie Mellon University, USA
[b] Center for the Neural Basis of Cognition, Carnegie Mellon University, USA
[c] Biological Imaging Center, California Institute of Technology, USA

ARTICLE INFO

Article history:
Received for publication 16 March 2011
Revised 20 June 2011
Accepted 22 June 2011
Available online 29 June 2011

Keywords:
Gene expression
Multiplex fluorescent in situ hybridization
Gene regulatory network
Sea urchin
GeneExpressMap
GEM
Image processing

ABSTRACT

The analysis of temporal and spatial patterns of gene expression is critically important for many kinds of developmental studies, including the construction of gene regulatory networks. Recently, multiplex, fluorescent, whole mount in situ hybridization (multiplex F-WMISH), applied in combination with confocal microscopy, has emerged as the method of choice for high-resolution, three-dimensional (3D) mapping of gene expression patterns in developing tissues. We have developed an image analysis tool, GeneExpressMap (GEM), that facilitates the rapid, 3D analysis of multiplex F-WMISH data at single-cell resolution. GEM assigns F-WMISH signal to individual cells based upon the proximity of cytoplasmic hybridization signal to cell nuclei. Here, we describe the features of GEM and, as a test of its utility, we use GEM to analyze patterns of regulatory gene expression in the non-skeletogenic mesoderm of the early sea urchin embryo. GEM greatly extends the power of multiplex F-WMISH for analyzing patterns of gene expression and is a valuable tool for gene network analysis and many other kinds of developmental studies.

Introduction

Spatial and temporal patterns of gene expression are used to analyze mechanisms of cell specification and patterning in developing tissues. Whole mount in situ hybridization (WMISH), which reveals the distribution of target mRNAs by visualizing the hybridization of complementary, labeled probes, is currently the most commonly used method of assessing spatial patterns of gene expression in fixed tissues. A major recent advance has been the application of fluorescence-based detection methods to WMISH (F-WMISH) (Clay and Ramakrishnan, 2005; Kosman et al., 2004). By labeling probes with different molecular tags and detecting their distributions with different fluorochromes (multiplex F-WMISH), it is now possible to analyze the expression of multiple genes simultaneously (i.e., in a single specimen). By facilitating precise comparisons of the expression patterns of genes, multiplex F-WMISH can reveal subtle differences that are very difficult to detect by other methods.

Transcriptional gene regulatory networks (GRNs) are valuable as conceptual and experimental tools for the analysis of developmental and evolutionary processes (Chan et al., 2009; Davidson, 2009; Ettensohn, 2009; Hinman et al., 2009; Nikitina et al., 2009; Stathopoulos and Levine, 2005). GRN analysis is based, in part, upon the view that the genomic regulatory state of an embryonic cell is defined by the ensemble of transcription factors that is contained within the cell at any particular developmental stage. Therefore, gene network analysis relies on detailed and accurate information concerning patterns of expression of regulatory genes (i.e., genes that encode transcription factors) in specific embryonic territories. Indeed, analysis of regulatory gene expression at single-cell resolution is a useful first step in constructing a GRN.

At present, a limitation of multiplex F-WMISH analysis is that it requires extensive manual manipulation and interpretation of large image data sets (e.g., Croce and McClay, 2010; Peter and Davidson, 2010; Sharma and Ettensohn, 2010). Such analysis is highly labor-intensive and is often difficult to standardize across experiments and between laboratories. F-WMISH data analyzed by manual methods can also be difficult to visualize in three dimensions and at single-cell resolution, both of which are critically important for GRN analysis.

To address these issues, we have developed a MATLAB-based, image analysis pipeline, GeneExpressMap (GEM), which makes possible the rapid, three-dimensional (3D) analysis of multiplex F-WMISH data at single-cell resolution. Here, we describe GEM and demonstrate its basic features. GEM assigns F-WMISH signal to individual cells based upon the proximity of cytoplasmic hybridization signal to cell nuclei. To use GEM, no modifications to standard protocols for preparing and imaging F-WMISH samples are required other than a brief incubation in a nuclear dye and the acquisition of an image stack in this channel. As a test of its utility, we have used GEM

* Corresponding author at: Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA.
E-mail address: ettensohn@andrew.cmu.edu (C.A. Ettensohn).

0012-1606/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2011.06.033

to analyze patterns of regulatory gene expression in the non-skeletogenic mesoderm (NSM) of the early sea urchin embryo. Our experimental findings clarify the process of cell specification within this complex domain and reveal that diverse regulatory pathways control epithelial-mesenchymal transition (EMT) during sea urchin gastrulation. Our work demonstrates the usefulness of GEM as a tool for elucidating territories of gene expression in developing tissues and for GRN analysis.

Methods

Animals and embryos

Adult *Lytechinus variegatus* were obtained from Reeftopia (www.reeftopia.com). Gametes were collected by intracoelomic injection of 0.5 M KCl and embryos were raised at 24 °C in temperature-controlled water-baths.

Cloning of genes from L. variegatus

Lvscl, *LvgataC*, *Lvpks*, *Lvets1* and *Lverg* were cloned by degenerate RT-PCR and RACE (random amplification of cDNA ends) using cDNA obtained from mesenchyme blastula stage embryos. *Lvendo16* and *Lvgcm* plasmids were provided by S. Ernst and D. McClay, respectively.

Fluorescent whole mount in situ hybridization (F-WMISH)

Multiplex F-WMISH was carried out according to Sharma and Ettensohn (2010). Digoxigenin-labeled antisense RNA probes were synthesized using the Megascript kit (Ambion, Inc.) and fluorescein-labeled antisense RNA probes were synthesized using the Fluorescein RNA Labeling Mix (Roche). Just before mounting, embryos were incubated in 0.5 μg/ml Hoechst 33342 in PBST (0.1% Tween-20 in phosphate-buffered saline) for 5 min and then rinsed several times with PBST.

Confocal microscopy and image analysis

Specimens that had been processed for F-WMISH were mounted on glass slides and covered with #1 glass coverslips. Small strips of Scotch #665 double-sided tape were used as spacers to ensure that embryos were slightly compressed but were not crushed by the coverslips. Z-stacks with a step size of 1 μm were collected using a Zeiss LSM510 Meta/Duoscan confocal microscope and 40× (N.A. = 1.3) or 100× (N.A. = 1.4) oil immersion lenses.

Results

GeneExpressMap

An overview of GEM operation

The analysis pathway and images that illustrate key steps in the operation of GEM are depicted in Fig. 1. GEM uses co-aligned stacks of Hoechst and F-WMISH images (which are imported as a single LSM file or as separate TIFF stacks) for nuclear localization and mRNA expression assignment, respectively. The pipeline is designed primarily for two-probe analysis, but single-probe analysis can be carried out by simply importing the same F-WMISH dataset into two channels.

All image stacks are thresholded to create binary representations of the data. Binary data are next subjected to noise reduction through the application of a binary erosion followed by a binary dilation, also known as morphological opening. This procedure eliminates small binary objects (noise) while preserving larger objects (signal). Each nucleus is then reconstructed from the binarized nuclear data and a model of all nuclear centers in the data set is constructed. The binarized F-WMISH data are subjected to a morphological dilation in order to induce a slight overlap between F-WHISH label and adjacent nuclei. The percentage of overlapping F-WMISH expression volume to total nuclear volume is then computed for each identified nucleus. A user-defined cutoff for this overlap percentage is used to assign individual cells (nuclei) as positive or negative with respect to the expression of a particular mRNA. Based on these assignments, a model of mRNA expression throughout the imaging volume is constructed. The user may inspect the quality of the

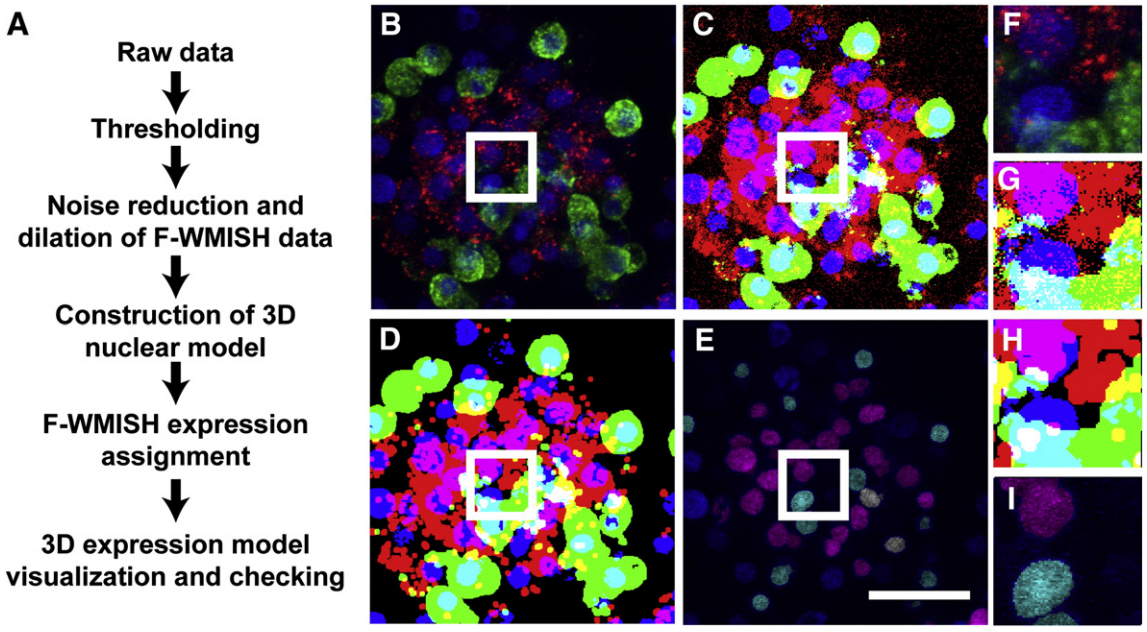


Fig. 1. Overview of GEM operation. (A) Flow chart depicting the key steps in the GEM data processing pathway. (B) Maximum intensity projection of raw data from a sea urchin embryo processed for multiplex F-WMISH using *Lvpks* (green) and *Lverg* (red) probes and counterstained with Hoechst dye to visualize nuclei. (C) Maximum intensity projection after binary thresholding of the raw data. (D) Maximum intensity projection after noise reduction and dilation of F-WMISH data. (E) Maximum intensity projection after construction of a 3D nuclear model and nuclear expression assignment. (F–I) Inset views from the stacks shown in B–E, respectively. Scale bar = 30 μm.

model through the use of built-in tools that assess the accuracy of the nucleus reconstruction process and that allow the user to visualize F-WMISH expression patterns.

Details of GEM operation

Creation of binary images. Each image in the nuclear data set and in the F-WHISH data set is considered independently of other images in the stack in order to control for variation in fluorescence intensity from slice to slice throughout the imaging volume. The user selects a threshold value between the minimum and maximum pixel values in the image as a binary threshold value (Figs. 1C and G). A default threshold value is provided to the user and is calculated using maximum correlation thresholding (Padmanabhan et al., 2010). Briefly, this technique finds the threshold value that maximizes the correlation of the binarized image to the original image by searching an array of threshold values and comparing the resultant image histogram to the original.

Noise reduction and F-WHISH data dilation. Each image in the binarized representation of the nuclear and F-WHISH data sets is again considered independently of other images in the stack. The user specifies an object size for noise reduction. The object size specified in this step corresponds to the size of an object (in microns) that the user chooses to remove from the image. The image plane resolution specified by the user (xy resolution) is used to transform the specified object size to the nearest object size in pixels. Noise reduction is then carried out using binary erosion followed by binary dilation (morphological opening) on the binary image (Supp. Fig. 1). This procedure eliminates objects in the image smaller than the user specified size in microns. Object sizes are specified for the nuclear data set and for each F-WHISH data set separately. Typically, we eliminate only single pixel "shot noise" from all three data sets. Binarized and noise-filtered F-WMISH data sets are further processed by binary dilation in order to slightly expand the F-WMISH data sets and to induce a small overlap between F-WMISH expression and adjacent nuclei (Figs. 1D and H). The user specifies a distance (in micrometers) by which the F-WMISH signal is expanded in this step. The program default for this distance is 1 μm as the resulting expansion is sufficient to induce detectable overlap between F-WMISH signal and adjacent nuclei while minimizing spurious overlap.

Nucleus reconstruction. All binarized images in the nuclear data set are first subjected to a distance transform in which the distance from the edge of a binarized nucleus is computed at each pixel in the image (distance image) (Supp. Figs. 2A–D). Only pixels falling inside of the original binarized nucleus data are considered for further processing in this step. The distance image is then segmented into putative nuclear objects using a watershed transform (watershed image) (Supp. Figs. 2E and F). Briefly, the watershed transform finds edges

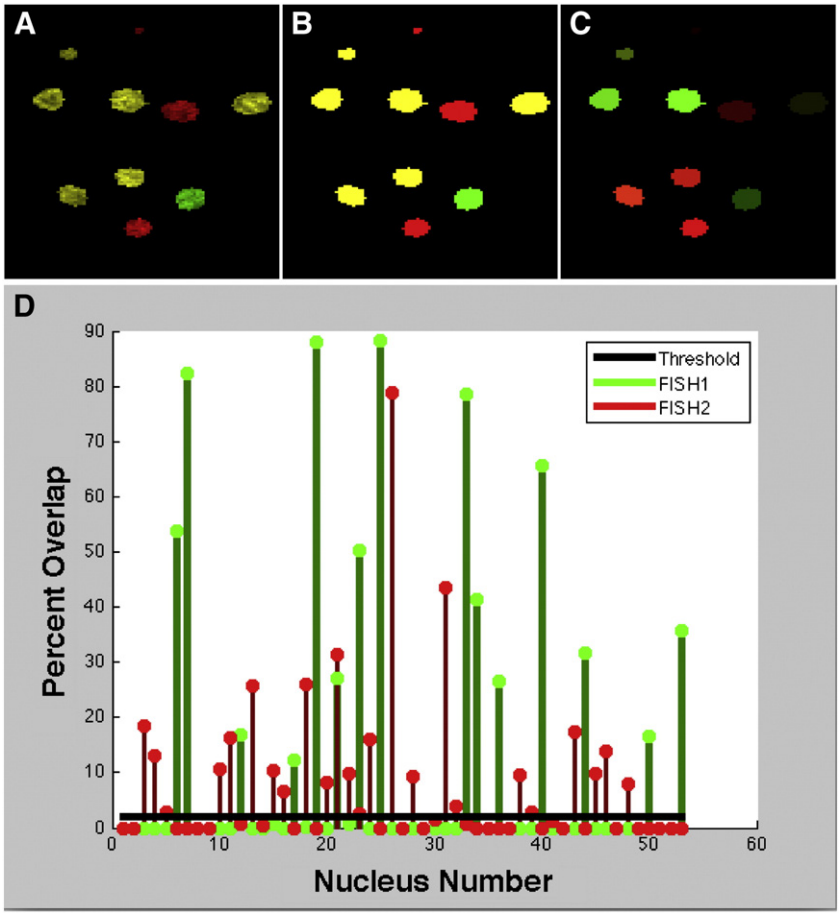


Fig. 2. Tools provided for optimization of the user-defined percent overlap cutoff used in gene expression assignment. (A) Default display mode for positively labeled nuclei. Positive nuclei (i.e., nuclei with percent overlap equal to or greater than a user-defined cutoff) are displayed as a binary value of one multiplied by the raw nuclear data set. (B) Binary visualization of positive nuclei. (C) Intensity map view of positive nuclei. This visualization mode facilitates the rapid evaluation of global patterns of percent overlap levels across the data set. (D) Graphical representation of the percent overlap of each F-WMISH probe with all identified nuclei relative to the user-defined percent overlap cutoff value. This tool allows the user to rapidly evaluate a chosen cutoff value for each unique data set.

between adjacent nuclear objects by locating boundaries between local peaks in the distance image. These edges are then used to isolate individual nuclear centers from one another. The sizes of all putative nuclear objects in the watershed image are next checked to ensure that they do not exceed 10 μm. For putative nuclear objects exceeding 10 μm, the object is further segmented at local minimums in the raw nuclear data set until the size of the putative nuclear object is under 10 μm. This size cutoff was chosen as a reasonable trade-off between prevention of putative nuclear objects containing multiple nuclei and allowance for single nuclei that were slightly larger than normal, given that the typical size of a nucleus is close to 5 μm. After segmentation and checking of putative nuclear objects, the centroid of each object is calculated and these centroids are used to construct a 3D model of all nuclei in the data set. Putative nuclear object centroids are consolidated into 3D nuclear centers by an iterative consolidation procedure. All putative nuclear object centroids are first placed into a list. For each generated putative nuclear object centroid, the 3D Euclidean distance to all other centroids in the list is then computed using xy resolution and z resolution values supplied by the user. Putative nuclear object centers falling within a user-defined distance threshold are consolidated into a single 3D nuclear centroid and are then removed from the list of putative nuclear object centroids. This procedure continues until all putative nuclear object centroids are assigned to a 3D nuclear centroid. Finally, all pixels belonging to putative nuclear objects that are linked together through this process are treated as slices through the same nucleus in different image planes. The resulting model of 3D nuclear pixel assignments is used for F-WMISH expression assignment.

F-WMISH expression assignment. The nuclear model constructed in the above step is used as a starting point for F-WMISH expression assignment. For each nucleus in the 3D model, the total nuclear volume is calculated by summing the number of pixels contained within the nucleus. Similarly, the volume of overlapping F-WMISH expression is calculated by summing the number of pixels contained with the nucleus that overlap with the expanded binary representation of the F-WMISH data set. A user-defined percent overlap cutoff is then used to determine whether each nucleus should be marked as the nucleus of a cell that expresses a particular mRNA (Figs. 1E and I).

Fine-tuning the percent overlap. In order to ensure that the most accurate gene expression data set is generated, GEM provides two tools that fine tune the user-defined percent overlap cutoff used for F-WMISH expression assignment in the above step (Fig. 2). The first tool is a set of alternate views of the nuclei that fall above the user-defined percent overlap cutoff for positive expression assignment. The default visualization method is to assign a binary expression value of 1 to the nuclei of all cells that have been scored as positive with respect to the expression of a given gene and then to multiply the luminance of the raw nuclear data with this value for display in the appropriate color channel (red or green) (Fig. 2A). Alternately, positive nuclei can be displayed as a binary visualization (Fig. 2B) or as a scaled intensity map (Fig. 2C). The intensity map view of the data is calculated by finding the nucleus in the data set with the highest percent overlap for F-WMISH in each channel and assigning it a value of 1. All nuclei with smaller percent overlap values are displayed with a scaled value according to their percent overlap relative to the nucleus with the highest percent overlap. For example, a positive nucleus will be displayed with a value of 0.5 if its percent overlap is 50% of the value of the highest percent overlap nucleus in the data set. This visualization option makes it possible to quickly and easily assess global patterns of percent overlap values across the entire data set and may help users to gauge their level of confidence that a given cell (or embryonic territory) expresses one or both mRNAs.

The second tool for fine-tuning the percent overlap value is a simple visualization of the percent overlap for each nucleus relative to the user-defined percent overlap cutoff. For each nucleus, the overlap percent is plotted for both probes and the user-defined threshold for positive assignment of gene expression is shown behind these plots (Fig. 2D). This display allows the user to rapidly visualize the expression levels of all nuclei in the data set for both F-WMISH probes and facilitates fine-tuning of the threshold that is used for positive F-WMISH expression assignment.

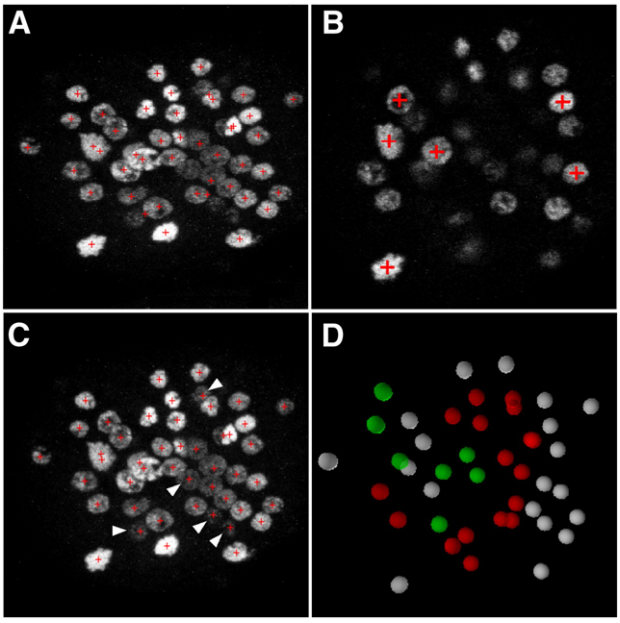
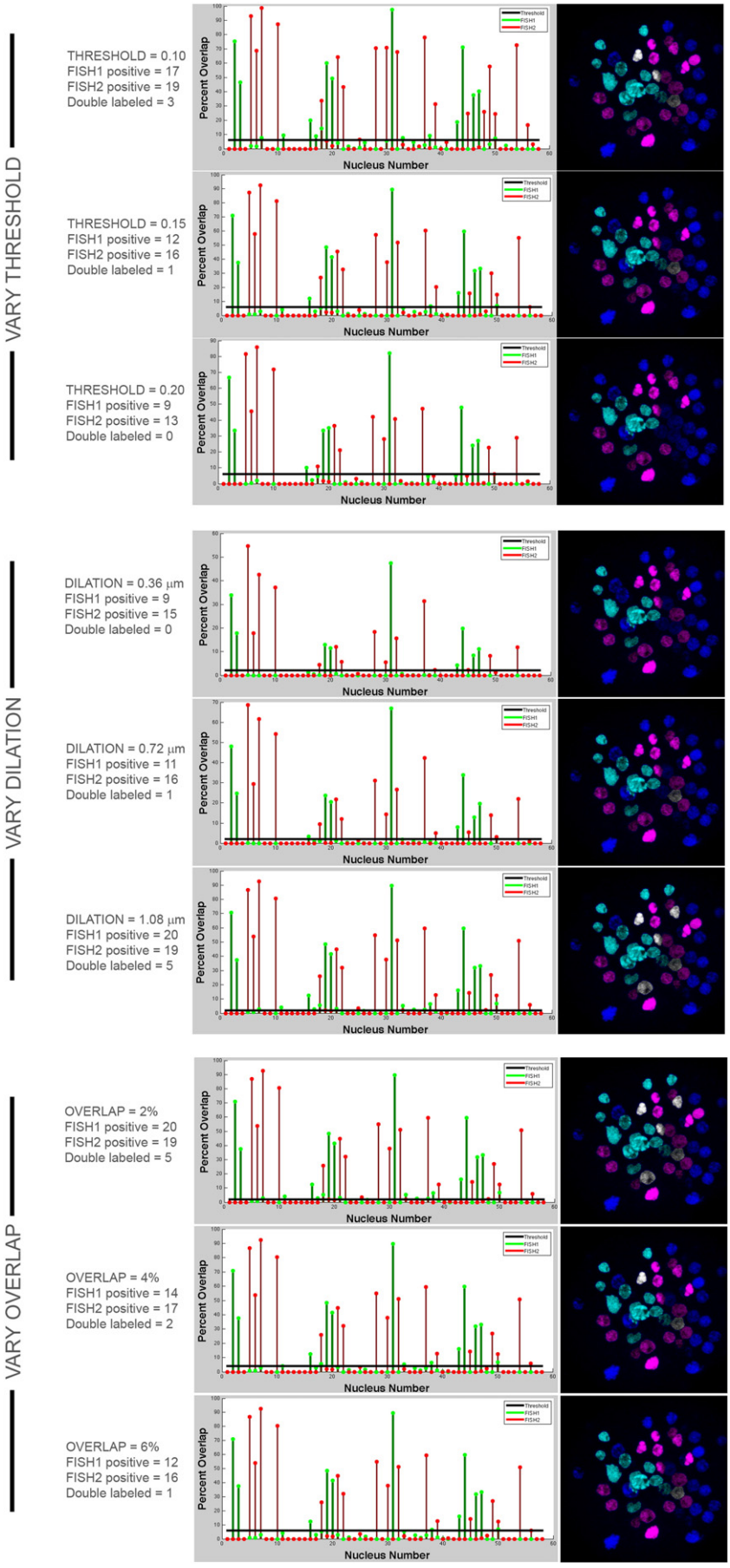


Fig. 3. Visual validation of nuclear assignment and F-WMISH expression. (A) Maximum intensity projection of a Hoechst-labeled sea urchin embryo with overlaid nucleus center estimates (red crosses). (B) Visualization of a single slice through the same data set with only the center positions in the slice visible. (C) Modified nuclear assignments after minor adjustments to z-axis resolution and distance threshold. Note the improved assignments of five nuclei (arrowheads). (D) 3D model of nuclear locations and their F-WMISH expression assignment. Gray spheres represent the nuclei of cells with no F-WMISH signal. Red and green spheres represent the nuclei of cells that express *Lvpks* and *Lvalx1*, respectively. Yellow spheres ordinarily represent the nuclei of cells that express both genes; no such cells are found in this data set.

Other validation tools. In order to validate the gene expression model constructed in the above steps, the software provides tools to assess the accuracy of the nuclear reconstruction steps and to examine the assignment of F-WMISH expression to nuclei in the 3D model. Two tools are provided to assess the accuracy of nuclear center estimates (Fig. 3). The first is an overlay of the maximum intensity projection of the raw nuclear data set (i.e., the Hoechst staining data) and the nuclear center estimates (Fig. 3A). This tool allows the user to quickly compare (i.e., in a single image) the positions of the centers of all predicted nuclei with the positions of nuclei as they appear in the original Hoechst data set. The second tool is a slice-by-slice viewer of the raw nuclear data, again with nuclear center estimates overlaid (Fig. 3B). This tool provides users with a more detailed view of the nuclear center estimates by allowing them to navigate through the 3D volume of the data set and visualize the precise location of nuclear center estimates in three dimensions. These two tools are extremely important, as they allow the fine-tuning of parameters and the optimization of nuclear center estimates. For example, we have found that small adjustments to the z-axis resolution, the threshold level for the nuclear data, and/or the distance threshold can improve the estimation of nuclear centers (Fig. 3C).

To visualize the assignment of F-WMISH expression to individual nuclei in the constructed model, the full data set is displayed in three dimensions and positive nuclei are highlighted (Fig. 3D). The user can manipulate this simple visualization tool to gain an intuitive understanding of the F-WMISH expression pattern in three dimensions. These tools allow the user to rapidly evaluate the effects of changing


VARY THRESHOLD
THRESHOLD = 0.10
FISH1 positive = 17
FISH2 positive = 19
Double labeled = 3
THRESHOLD = 0.15
FISH1 positive = 12
FISH2 positive = 16
Double labeled = 1
THRESHOLD = 0.20
FISH1 positive = 9
FISH2 positive = 13
Double labeled = 0
VARY DILATION
DILATION = 0.36 μm
FISH1 positive = 9
FISH2 positive = 15
Double labeled = 0
DILATION = 0.72 μm
FISH1 positive = 11
FISH2 positive = 16
Double labeled = 1
DILATION = 1.08 μm
FISH1 positive = 20
FISH2 positive = 19
Double labeled = 5
VARY OVERLAP
OVERLAP = 2%
FISH1 positive = 20
FISH2 positive = 19
Double labeled = 5
OVERLAP = 4%
FISH1 positive = 14
FISH2 positive = 17
Double labeled = 2
OVERLAP = 6%
FISH1 positive = 12
FISH2 positive = 16
Double labeled = 1
Percent Overlap
Nucleus Number
Threshold
FISH1
FISH2

parameters in the program and to fine-tune its operation to each unique data set. It should be noted that the 3D visualization tools included in GEM are intended only for the initial assessment of gene expression patterns and for parameter fine-tuning. GEM output can easily be imported into stand-alone 3D visualization programs, including Avizo (Visualization Sciences Group) and the public V3D program (Peng et al., 2010), in order to generate high-quality 3D renderings (see below).

Effects of altering user-defined parameters. Fig. 4 illustrates the effects of altering each user-defined parameter (other than noise reduction, which is typically 1 pixel) on the analysis of a sample data set. In general, increasing the threshold value for F-WMISH data results in an attenuation of the F-WMISH signal and therefore a decrease in the number of positive nuclei (Fig. 4, upper panels). We have found that the default threshold value is appropriate for most F-WMISH data. Increasing the dilation value has the opposite effect; it results in an expansion of the F-WMISH signal and therefore an increase the number of positive cells (Fig. 4, center panels). Increasing the overlap percent (the amount of overlap required to score a nucleus as positive) results in fewer nuclei that are scored as positive (Fig. 4, lower panels). In practice, the dilation and percent overlap parameters effectively counterbalance one another.

Saving and sharing data. GEM allow users to save an entire session, which can be reopened to continue the analysis from any point, or only the parameter settings of a session, which can be imported into other sessions (e.g., if similar data sets are being analyzed). A valuable feature of GEM is that it allows researchers at different locations to collaborate on the analysis of the same F-WMISH dataset by sharing GEM session files.

Evaluation of GEM using control data sets

To assess the accuracy of GEM in defining territories of gene expression at the single-cell level, we visualized the patterns of expression of known marker genes in the sea urchin (*L. variegatus*) embryo (Fig. 5). We chose two combinations of genes that challenged GEM in complementary ways: a) *Lvalx1* (*aristaless-like-1*) and *Lvtbr* (*t-brain*), which are co-expressed by skeletogenic primary mesenchyme cells (PMCs) at the blastula and gastrula stages, and b) *Lvalx1* and *Lvpks* (*polyketide synthase*), which are expressed in adjacent but non-overlapping territories in the vegetal plate at the late blastula stage.

Gastrula stage embryos that were labeled with *Lvalx1* and *Lvtbr* probes and analyzed using GEM showed that these mRNAs were co-expressed by PMCs, as expected (yellow-gray nuclei in Fig. 5B). Furthermore, analysis of late blastula stage embryos that were labeled with *Lvalx1* and *Lvpks* probes showed that these two mRNAs were expressed in non-overlapping domains within the vegetal plate (Figs. 5C and D). Significantly, GEM was able to distinguish *Lvalx1*(+) and *Lvpks*(+) cells, even though cells were very closely packed within the pseudostratified epithelium of the vegetal plate and many *Lvalx1*(+) and *Lvpks*(+) cells were immediate neighbors. Therefore, analysis of these known marker mRNAs provided strong evidence that GEM produces an accurate representation of territories of gene expression at single-cell resolution.

We analyzed GEM control data sets using a variety of 3D visualization programs, including Avizo (Visualization Sciences Group) and the public V3D program (Peng et al., 2010). Using these 3D visualization tools, output data from GEM provided striking, high-resolution images of gene expression patterns in the sea urchin embryo, as illustrated in Figs. 6A and B. We also found that GEM was useful in analyzing patterns of gene expression in starfish (*Asterina miniata*) embryos (Fig. 6C). GEM will facilitate the analysis of gene expression patterns in any experimental model that can be used to generate high-quality F-WMISH data.

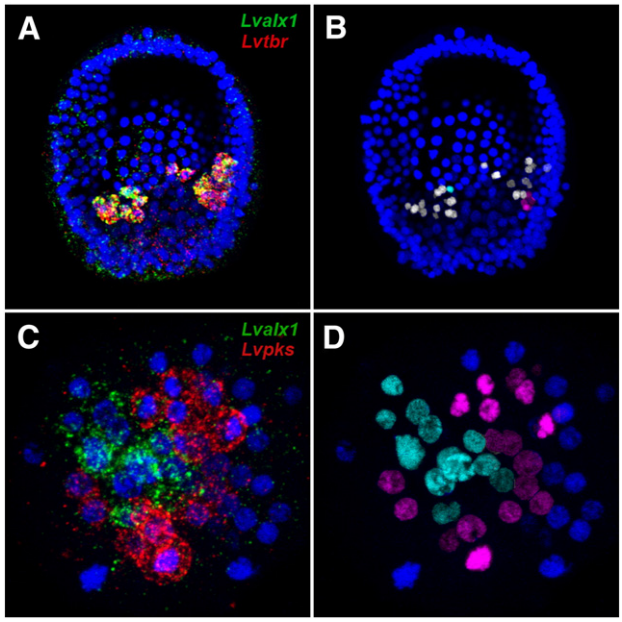


Fig. 5. GEM analysis of the expression patterns of known marker genes in the sea urchin embryo. Panels at left show z-projections of raw F-WMISH image stacks. Panels at right show z-projections of the same data processed using GEM. (A,B) *Lvalx1* (green) and *Lvtbr* (red) expression in a late gastrula stage embryo. Regions where red and green signals overlap appear yellow-gray in the image. (C,D) *Lvalx1* (green) and *Lvpks* (red) expression in a late blastula stage embryo (prior to PMC ingression). These genes are expressed in non-overlapping populations of cells in the vegetal plate.

Analysis of gene expression patterns in the sea urchin NSM using GEM

We next used GEM to reveal new patterns of gene expression during sea urchin gastrulation. We focused on the non-skeletogenic mesoderm (NSM), a complex territory that gives rise to several different cell types, including pigment cells, blastocoelar cells, coelomic pouch cells, and muscle cells (Ruffins and Ettensohn, 1996). Two of these cells types, pigment cells and blastocoelar cells, undergo EMT during gastrulation, as do PMCs earlier in development. *Ets1* plays a critically important role in PMC specification and inputs from *ets1* are also required for PMC ingression, although these regulatory connections may be indirect (Kurokawa et al., 1999; Oliveri et al., 2008; Röttinger et al., 2004). One of our goals was to explore the possible role of *ets1* in the specification and EMT of pigment cells and blastocoelar cells during gastrulation.

As markers of prospective pigment cells, we used the regulatory gene, *gcm (glial cells missing)*, and a terminal differentiation gene, *pks*. *Gcm* exhibits a dynamic pattern of expression in the vegetal region; it is expressed transiently by both Veg2U and Veg2L layers at the early blastula stage but soon becomes restricted to the Veg2L layer, which likely corresponds to the entire presumptive non-skeletogenic mesoderm (Croce and McClay, 2010). By the mesenchyme blastula stage, *gcm*-expressing cells are no longer present in a concentric ring

Fig. 4. Effects of altering user-defined parameters. *Lvalx1* (FISH1, green) and *Lvpks* (FISH2, red) expression in a late blastula stage embryo (prior to PMC ingression). Manual analysis of this dataset indicated 12 or 13 alx1(+) cells and 17 pks(+) cells, with no double-labeled cells. Upper panels: Increasing the threshold reduces the numbers of positive cells. Dilation and percent overlap values were held constant (at 1.0 μm and 6%, respectively) and both FISH data sets were adjusted to a threshold of 0.10, 0.15, or 0.20. (Note that for this dataset, the default thresholds were approximately 0.15.) Middle panels: Increasing the dilation value increases the number of positive cells. Threshold and percent overlap values were held constant (at 0.15 and 2%, respectively) and both FISH data sets were dilated by 0.36, 0.72, or 1.08 μm. Lower panels: Increasing the overlap value reduces the number of positive cells. For this analysis, the threshold and dilation values were held constant (at 0.15 and 1.0 μm, respectively) and both FISH data sets were analyze using overlap values of 2%, 4%, or 6%.

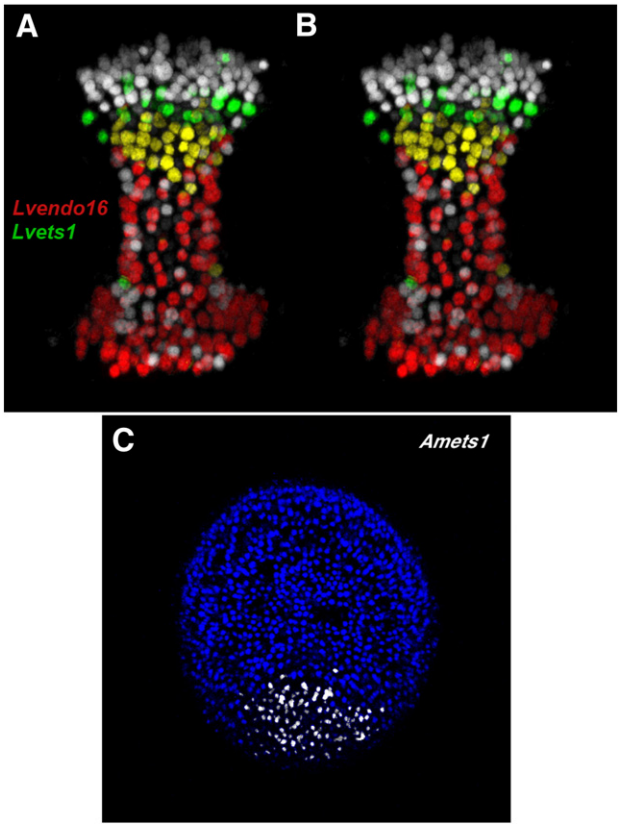


Fig. 6. (A,B) Stereo rendering of a GEM data set, produced using the Avizo 3D visualization program (Visualization Sciences Group). The archenteron of a late gastrula stage embryo, labeled with *Lvendo16* (red) and *Lvets1* (green) probes. *Lvendo16* mRNA is present throughout the archenteron and in cells that surround the blastopore, while *Lvets1* is restricted to a subset of cells near the archenteron tip. Many cells co-express both mRNAs, as shown by the yellow nuclei. (C) GEM analysis of the expression of *ets1* (yellow-gray nuclei) in a late blastula stage starfish (*Asterina miniata*) embryo.

surrounding the vegetal pole but are concentrated on the aboral side of the vegetal plate, in the subset of NSM cells that will give rise to pigment cells (Ransick et al., 2002; Ransick and Davidson, 2006). *Pks* is a terminal differentiation gene in the pigment cell GRN and encodes an enzyme complex that is required for the accumulation of echinochrome (Calestani et al., 2003).

We used GEM to analyze embryos that had been double-labeled with *Lvgcm* and *Lvpks* probes and confirmed that these mRNAs are co-expressed by presumptive pigment cells on the aboral side of the vegetal plate at the mesenchyme blastula stage (Fig. 7A). At this stage, we reproducibly observed a small number of cells at the periphery of the labeled territory that expressed *Lvgcm* but not *Lvpks* (red nuclei in Fig. 7A). These may represent presumptive pigment cells that have not yet activated *Lvpks*, or they may represent cells that transiently express *Lvgcm* but do not give rise to pigment cells. As noted above, *gcm* exhibits a dynamic pattern of expression during early development and not all *gcm*-expressing cells activate downstream components of the pigment cell GRN.

We used the regulatory gene, *scl* (*stem cell leukemia*), as a marker for blastocoelar cells. At the mesenchyme blastula stage, this mRNA is restricted to cells in the oral region of the vegetal plate, a territory that gives rise to blastocoelar cells (Ruffins and Ettensohn, 1996). Multiplex F-WMISH analysis of mesenchyme blastula stage embryos that had been double-labeled with *Lvscl* and *Lvgcm* probes showed that *Lvscl* was expressed in a crescent-shaped region on the side the vegetal plate opposite the domain of *Lvgcm* expression (Fig. 7B). We also found that *Lvscl* co-localized with *LvgataC*, which has recently been reported to be expressed on the oral side of the vegetal plate (Duboc et al., 2010). Based on these findings, we consider *Lvscl* to be a

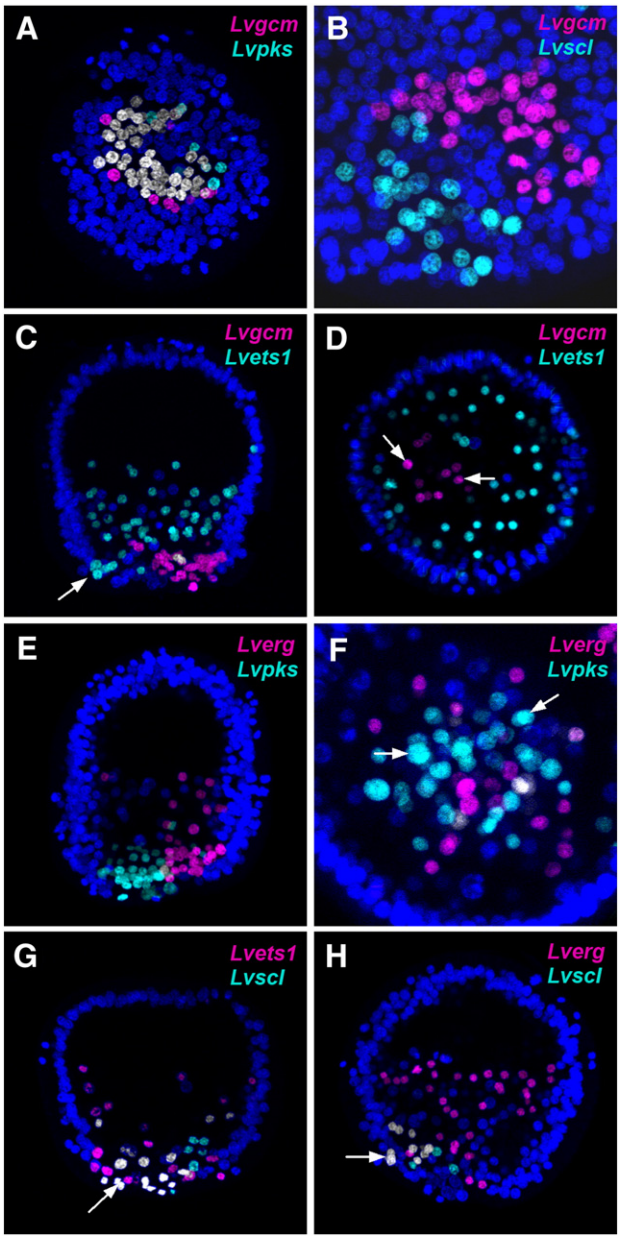


Fig. 7. Patterns of gene expression in the NSM territory of the sea urchin embryo at the late mesenchyme blastula and early gastrula stages. A) Mesenchyme blastula labeled with *Lvgcm* (red) and *Lvpks* (green) probes. The domains of expression of these two genes overlap almost completely, as indicated by the double-labeled cells (yellow-gray nuclei). B) Mesenchyme blastula labeled with *Lvgcm* (red) and *Lvscl* (green) probes. These regulatory genes are expressed by non-overlapping populations of cells in the vegetal plate. C,D) Mesenchyme blastula (C, lateral view) and early gastrula (D, vegetal view) labeled with *Lvets1* (green) and *Lvgcm* (red) probes. At the mesenchyme blastula stage (C), *Lvets1* is expressed in NSM cells in the vegetal plate (arrow) and also in PMCs that have ingressed into the blastocoel (note that not all PMCs appear labeled because the level of *Lvets1* expression in PMCs is declining by this stage). *Lvgcm* is expressed on the opposite side of the vegetal plate. At the early gastrula stage (D), pigment cells that are in the process of ingression (arrows) express *Lvgcm* but not *Lvets1*. E,F) Mesenchyme blastula (E, lateral view) and early gastrula (F, vegetal view) labeled with *Lverg* (red) and *Lvpks* (green) probes. At the mesenchyme blastula stage (E) these mRNAs are expressed in complementary regions of the vegetal plate. During early gastrulation (F), ingressing pigment cells (arrows) express *Lvpks* but not *Lverg*. G) Early gastrula (lateral view) labeled with *Lvets1* (red) and *Lvscl* (green) probes. Many cells co-express both mRNAs, as indicated by the yellow-gray nuclei (arrow). H) Early gastrula stage (lateral view) labeled with *Lverg* (red) and *Lvscl* (green) probes. Arrow indicates the nuclei of cells that co-express both mRNAs.

marker for presumptive blastocoelar cells. Expression of *Lvscl* declined during gastrulation and therefore this gene could not be used as a marker for blastocoelar cells later developmental stages.

The zygotic expression of *ets1* is initially restricted to presumptive PMCs, but this gene is activated in the NSM territory shortly before the onset of gastrulation (Kurokawa et al., 1999; Rizzo et al., 2006). The earliest stage at which we could reliably detect expression of *Lvets1* in the NSM territory was the late mesenchyme blastula stage. Analysis of mesenchyme blastula stage embryos that had been double-labeled with *Lvets1* and *Lvgcm* probes showed that the territories of expression of these two genes were non-overlapping (Fig. 7C). We also examined early gastrula stage embryos, when pigment cell EMT was underway, and confirmed that *Lvgcm*-positive cells did not express detectable levels of *Lvets1* (Fig. 7D). These findings indicate that, although regulatory inputs from *ets1* are required for PMC ingression, *ets1* is not expressed at detectable levels by presumptive pigment cells and is therefore unlikely to mediate their transformation from an epithelial to a mesenchymal state.

Erg is another member of the ETS family that is expressed in the NSM territory (Zhu et al., 2001). To test the possibility that *erg* might be expressed by presumptive pigment cells and might substitute for *ets1* in regulating pigment cell specification and EMT, we analyzed *Lverg* expression in embryos that had been co-labeled with pigment and blastocoelar cell markers. As was the case with *Lvets1*, we first detected *Lverg* expression in NSM cells at the mesenchyme blastula stage. Analysis of embryos that had been double-labeled with *Lverg* and *Lvpks* probes showed that these two mRNAs were present in distinct, non-overlapping populations of cells at this stage (Fig. 7E). At the early gastrula stage, when pigment cell ingression was underway, *Lverg* and *Lvpks* continued to be expressed in separate cellular compartments (Fig. 7F).

Both *Lvets1* and *Lverg* exhibited polarized expression within the NSM territory during early gastrulation and our finding that these mRNAs were not present in prospective pigment cells suggested that *Lvets1* and *Lverg* might be expressed selectively by presumptive blastocoelar cells, which are concentrated on the opposite (oral) side of the vegetal plate. GEM analysis of embryos that had been double-labeled with *Lvscl* and *Lvets1* probes confirmed that large numbers of cells co-expressed these two genes (yellow-gray nuclei in Fig. 7G). Similar observations were made of embryos that had been double-labeled with *Lvscl* and *Lverg* probes (yellow-gray nuclei in Fig. 7H). We also observed cells in the vegetal plate that expressed *Lvscl* but not *Lvets1* or *Lverg*, and *Lvets1*-positive or *Lverg*-positive cells that did not express *Lvscl* (singly-labeled nuclei in the vegetal plate territories of the embryos in Figs. 7G and H). The dynamic expression of these genes during early gastrulation, however, complicated the interpretation of such labeling patterns. As noted above, *Lvscl* expression declines during gastrulation, while the expression of *Lvets1* and *Lverg* in the NSM territory increases during this same interval. It is therefore possible that cells that were singly labeled with *Lvscl*, *Lvets1*, or *Lverg* at the mesenchyme blastula stage were in transitional phases with respect to their expression of the other regulatory genes. Despite these considerations, our analysis showed clearly that *Lvets1* and *Lverg* mRNAs were absent from pigment cell progenitors and that these regulatory factors were expressed predominantly by presumptive blastocoelar cells.

Discussion

The development of new methods for analyzing patterns of gene expression has played a central role in our understanding of cell specification and embryo patterning. We have developed a rapid pipeline for the analysis of multiplex F-WMISH data at single-cell resolution that will be valuable for many kinds of developmental studies. The application of tools such as GEM will refine our picture of the dynamic patterns of gene expression in developing tissues and will spur new experimental studies. Recently, the emergence of developmental GRNs as a conceptual and experimental tool has underscored the critical importance of gene expression analysis and, in particular, has highlighted the need to define the patterns of expression of regulatory genes at the single-cell level. Indeed, the analysis of patterns of regulatory gene expression is an essential first step in constructing a GRN.

When compared with manual methods for analyzing multiplex F-WMISH data (e.g., Croce and McClay, 2010; Peter and Davidson, 2010; Sharma and Ettensohn, 2010), GEM provides important advances in speed, reproducibility, and quantification. Manual analysis of F-WMISH data involves a qualitative assessment of the relative levels of fluorescent signal in different cells and an assignment of a cell as either positive or negative based upon a subjective evaluation that is likely to vary among users and from experiment to experiment. When applying GEM, the primary sources of variability arise from the user's choice of a) the threshold value that is applied to the F-WMISH data and b) the percent overlap value that is used to identify positive cells. To minimize variability from these sources, we have provided users with the option of using a default threshold value that is based on maximum correlation thresholding of raw data (Padmanabhan et al., 2010). The choice of the percent overlap value is inherently difficult to standardize; one option for users is to systematically explore a range of values in order to evaluate possible effects on their interpretation of the data. To further address this issue, we have provided an optional intensity map view of F-WMISH expression in GEM and a simple tool to gauge the relative expression levels of all nuclei in a data set (Fig. 2). These features help to standardize GEM analysis. In this regard, it should be noted that one critically important advantage of GEM when compared with manual approaches is that potential sources of variability in the analysis can be rapidly and systematically explored, so that their influence on the final interpretation of the data can be assessed.

For most investigators, GEM will be straightforward to employ. GEM is freely available (http://github.com/coreyflynn/GeneExpressMap) and uses the popular MATLAB platform. To use GEM, no modifications to conventional F-WMISH protocols are required other than a brief (5 min) incubation of the specimen in a nuclear dye and the acquisition of an image stack in this channel. Output from GEM can be analyzed and presented using a variety of dedicated 3D visualization programs, including the recently produced, public V3D program (Peng et al., 2010).

To test the utility of GEM, we used the program to analyze patterns of gene expression within the NSM territory of the sea urchin embryo. At present, the most thoroughly described GRN in the sea urchin is the GRN that underlies the specification and morphogenesis of the skeletogenic PMCs (reviewed by Ettensohn, 2009). Mechanisms of EMT in the sea urchin have also been studied most intensively with respect to the PMCs (reviewed by Wu et al., 2007). In the large micromere-PMC lineage, *ets1* and a second regulatory protein, *alx1*, play critically important roles in specification and provide regulatory inputs that are required for EMT (Kurokawa et al., 1999; Ettensohn et al., 2003; Röttinger et al., 2004). The mechanisms by which these regulatory factors regulate EMT are not fully understood. PMC ingression is thought to involve Snail-mediated remodeling of the cell surface; current evidence suggests that these events are dependent on *alx1*, a gene that may regulate *snail* in the PMC GRN, but are independent of *ets1* (Wu and McClay, 2007; but see also Oliveri et al., 2008).

Two other populations of cells undergo EMT during gastrulation; blastocoelar cells and pigment cells. Our gene expression analysis provides strong evidence that pigment cell specification and ingression occur by *ets1*/*erg*-independent mechanisms. In this regard, it should be noted that there exists a substantial maternal pool of *ets* (but not *erg*) mRNA and that Ets1 protein accumulates transiently in the nuclei of all cells early in development (Sharma and Ettensohn, 2010). It appears unlikely that this early, ubiquitous pool of Ets1 protein plays an important role in pigment cell specification or EMT, however, as overexpression of a dominant negative form of *Lvets1* that blocks both PMC

specification and ingression has no effect on pigment cell development (Sharma and Ettensohn, 2010 and unpublished observations). Genome-wide analysis of ETS family members in the sea urchin has identified only one additional gene, *ese*, that is expressed by NSM cells during gastrulation. *Ese* mRNA is concentrated on the oral side of the vegetal plate, however (Duboc et al., 2010; Rizzo et al., 2006), suggesting that this gene, like *ets1* and *erg*, is expressed selectively by blastocoelar cells. Taken together, these observations indicate that the specification and ingression of pigment cells occur by mechanisms that are independent of any ETS-family proteins. At present, although certain features of the pigment cell GRN have been elucidated (Calestani et al., 2003; Ransick and Davidson, 2006; Sherwood and McClay, 1999; Sweet et al., 2002) the connections between this GRN circuitry and pigment cell EMT have not been established. If the cell biological events of EMT, including cell surface remodeling, are similar in pigment cells and PMCs, then such cell biological events can be triggered by very different GRN circuitry.

Our findings also support the view that the genomic regulatory states of PMCs and blastocoelar cells share certain common features. It was noted previously that several important components of the PMC GRN, including *ets1* and *erg*, are also deployed in the NSM territory (Ettensohn et al., 2007). This work refines that view by identifying *ets1*/*erg*-expressing NSM cells as prospective blastocoelar cells and adds to a growing body of evidence which suggests that the regulatory states of PMCs and blastocoelar cells are similar in many respects (Rafiq and Ettensohn, unpublished observations). Pigment cells, in contrast, appear to be distinctive with respect to the mechanisms of their specification and the genomic regulatory program that controls their morphogenesis. Further dissection of the GRNs that are deployed in the three populations of migratory mesoderm cells will clarify the extent to which similar cellular behaviors during gastrulation are driven by common gene regulatory pathways.

Supplementary materials related to this article can be found online at 10.1016/j.ydbio.2011.06.033.

Acknowledgments

The authors thank J. Rast for sharing unpublished information concerning the patterns of *scl* and *gatac* expression, J. Croce, D. McClay, and S. Ernst for providing plasmids, and B. McCauley and V. Hinman for providing F-WMISH images of starfish embryos. This work was supported by NSF Grant IOS-0745875 to C.A.E.

References

Calestani, C., Rast, J.P., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. Development 130, 4587–4596.

Chan, T.M., Longabaugh, W., Bolouri, H., Chen, H.L., Tseng, W.F., Chao, C.H., Jang, T.H., Lin, Y.I., Hung, S.C., Wang, H.D., Yuh, C.H., 2009. Developmental gene regulatory networks in the zebrafish embryo. Biochim. Biophys. Acta. 1789, 279–298.

Clay, H., Ramakrishnan, L., 2005. Multiplex fluorescent in situ hybridization in zebrafish embryos using tyramide signal amplification. Zebrafish 2, 105–111.

Croce, J.C., McClay, D.R., 2010. Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo. Development 137, 83–91.

Davidson, E.H., 2009. Network design principles from the sea urchin embryo. Curr. Opin. Genet. Dev. 19, 535–540.

Duboc, V., Lapraz, F., Saudemont, A., Bessodes, N., Mekpoh, F., Haillot, E., Quirin, M., Lepage, T., 2010. Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. Development 137, 223–235.

Ettensohn, C.A., 2009. Lessons from a gene regulatory network: echinoderm skeletogenesis provides insights into evolution, plasticity and morphogenesis. Development 136, 11–21.

Ettensohn, C.A., Illies, M.R., Oliveri, P., De Jong, D.L., 2003. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. Development 130, 2917–2928.

Ettensohn, C.A., Kitazawa, C., Cheers, M.S., Leonard, J.D., Sharma, T., 2007. Gene regulatory networks and developmental plasticity in the early sea urchin embryo: alternative deployment of the skeletogenic gene regulatory network. Development 134, 3077–3087.

Hinman, V.F., Yankura, K.A., McCauley, B.S., 2009. Evolution of gene regulatory network architectures: examples of subcircuit conservation and plasticity between classes of echinoderms. Biochim. Biophys. Acta. 1789, 326–332.

Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., Bier, E., 2004. Multiplex detection of RNA expression in *Drosophila* embryos. Science 305, 846.

Kurokawa, D., Kitajima, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Shimada, H., Akasaka, K., 1999. HpEts, an ets-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. Mech Dev. 80, 41–52.

Nikitina, N., Sauka-Spengler, T., Bronner-Fraser, M., 2009. Gene regulatory networks in neural crest development and evolution. Curr. Top. Dev. Biol. 86, 1–14.

Oliveri, P., Tu, Q., Davidson, E.H., 2008. Global regulatory logic for specification of an embryonic cell lineage. Proc Natl Acad Sci U S A 105, 5955–5962.

Padmanabhan, K., Eddy, W.F., Crowley, J.C., 2010. A novel algorithm for optimal image thresholding of biological data. J. Neurosci Methods. (Epub ahead of print).

Peng, H., Ruan, Z., Long, F., Simpson, J.H., Myers, E.W., 2010. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. Nat. Biotechnol. 28, 348–353.

Peter, I.S., Davidson, E.H., 2010. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. Dev. Biol. 340, 188–199.

Ransick, A., Davidson, E.H., 2006. cis-Regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. Dev Biol. 297, 587–602.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev Biol. 246, 132–147.

Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A., Arnone, M.I., 2006. Identification and developmental expression of the *ets* gene family in the sea urchin (*Strongylocentrotus purpuratus*). Dev Biol. 300, 35–48.

Röttinger, E., Besnardeau, L., Lepage, T., 2004. A Raf/MEK/ERK signaling pathway is required for development of the sea urchin embryo micromere lineage through phosphorylation of the transcription factor Ets. Development 131, 1075–1087.

Ruffins, S.W., Ettensohn, C.A., 1996. A fate map of the vegetal plate of the sea urchin (*Lytechinus variegatus*) mesenchyme blastula. Development 122, 253–263.

Sharma, T., Ettensohn, C.A., 2010. Activation of the skeletogenic gene regulatory network in the early sea urchin embryo. Development 137, 1149–1157.

Sherwood, D.R., McClay, D.R., 1999. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. Development 126, 1703–1713.

Stathopoulos, A., Levine, M., 2005. Genomic regulatory networks and animal development. Dev. Cell 9, 449–462.

Sweet, H.C., Gehring, M., Ettensohn, C.A., 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. Development 129, 1945–1955.

Wu, S.Y., McClay, D.R., 2007. The Snail repressor is required for PMC ingression in the sea urchin embryo. Development. 134, 1061–1070.

Wu, S.Y., Ferkowicz, M., McClay, D.R., 2007. Ingression of primary mesenchyme cells of the sea urchin embryo: a precisely timed epithelial mesenchymal transition. Birth Defects Res. C Embryo Today 81, 241–252.

Zhu, X., Mahairas, G., Illies, M., Cameron, R.A., Davidson, E.H., Ettensohn, C.A., 2001. A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. Development 128, 2615–2627.