

Towards a digital model of zebrafish embryogenesis. Integration of cell tracking and gene expression quantification

Carlos Castro-González*, Miguel Ángel Luengo-Oroz*, Louise Douloquin[‡], Thierry Savy[†], Camilo Melani[†], Sophie Desnoullez[‡], María Jesús Ledesma-Carbayo*, Paul Bourguine[†], Nadine Peyriéras[‡], Andrés Santos*

Abstract—We elaborate on a general framework composed of a set of computational tools to accurately quantificate cellular position and gene expression levels throughout early zebrafish embryogenesis captured over a time-lapse series of *in vivo* 3D images. Our modeling strategy involves nuclei detection, cell geometries extraction, automatic gene levels quantification and cell tracking to reconstruct cell trajectories and lineage tree which describe the animal development. Each cell in the embryo is then precisely described at each given time t by a vector composed of the cell 3D spatial coordinates (x, y, z) along with its gene expression level g . This comprehensive description of the embryo development is used to assess the general connection between genetic expression and cell movement. We also investigate genetic expression propagation between a cell and its progeny in the lineage tree. More to the point, this paper focuses on the evolution of the expression pattern of transcriptional factor goosecoid (*gsc*) through the gastrulation process between 6 and 9 hours post fertilization (hpf).

I. INTRODUCTION

Understanding the role of cell dynamics and morphogenetic processes during the embryogenesis of living animals is a major question in bio-medical research. Capturing quantitative data of gene expression in time and space at the single cell resolution becomes then a crucial step to fulfill this goal. This topic has been approached by a few big projects [7] [3] [6] concerning different animal models: the *C. elegans*, the *Drosophila* and the mouse brain. Our focus is on the zebrafish (*Danio rerio*) embryo which, due to its embryonic optical transparency, quick growth and similarities to human differentiated cell types, provides an interesting system to study vertebrate development.

The construction of a spatiotemporal 4D (3D+T) map of gene expression in zebrafish development has challenging characteristics: there are no anatomical references as there are for the mouse brain, development is not stereotyped at cellular level as it is in *C. elegans*, and its developmental complexity (cell morphology, morphogenetic movements, etc.) is much higher than in the *Drosophila* case.

Compared to previous works on zebrafish, Castro et al. [1] presented a strategy to create a 3D map of gene expression

levels while Keller et al. [4] proposed an in toto global description of cell dynamics. The protocol introduced in this paper includes both an accurate quantification of the expression of gene products and their cellular location, together with the possibility to perform a 4D (3D+T) [9] cell tracking. This strategy can then offer new insights into the relations between gene expression and cell migrations. As a result, we get a complete (x, y, z, t, g) description of each embryo cell. With simple modifications, this systematic analysis could be easily extended to further developing stages and genetic products.

In this context, we applied time-lapse, confocal, bi-photon laser scanning microscopy with the aim of observing the spatial location and *in vivo* evolution of gene expression data and zebrafish nuclei during the gastrulation period [5] between 6 and 9 hours post fertilization (hpf). During this period, the transcriptional factor we focused on -named goosecoid (*gsc*)- starts its expression at the dorsal side to progressively move up to the animal pole leading to the dorsal axis formation (Fig. 2).

The paper is organized as follows: Section II shows the general framework and describes the computational methods involved at cell detection, tracking and gene quantification. Section III offers details on the employed data set and shows the results achieved on global cell migration and the particular behavior of 10 representative cells. Finally we discuss the results and bibliography.

II. METHODOLOGY

Our general framework (Fig. 1), which aims at describing each cell by its position at the (x, y, z, t, g) space, includes two main processes: 1) An automatic genetic quantification scheme that provides us information about the location in time and space and levels of expression of one gene, *gsc*. 2) An automatic lineage reconstruction scheme that provides each embryo cell progeny and trajectory which can be optionally visually validated.

A. Nuclei Detection and Tracking.

We performed a cell nuclei detection based on the numerical solution of a 3D nonlinear advection-diffusion equation proposed in [2]. Given the set of identified nuclei, in [10] we implemented an iterative greedy algorithm that builds the cell lineage tree as follows: The vector field of every image I_t is computed by registering it into next image I_{t+1} . Based on this vector field, it is possible to predict the position of each

This work is supported by the Spanish Ministry of Education through its FPU grant program.

*C. Castro-González, M.A. Luengo-Oroz, M.J. Ledesma and A. Santos are with Biomedical Image Technologies, ETSIT, Universidad Politécnica de Madrid, 28040, Spain. {ccastro, maluengo, mledesma, andres}@die.upm.es

[‡]L. Douloquin, S. Desnoullez and N. Peyriéras are with DEPSN, CNRS, Institut de Neurobiologie Alfred Fessard, Gif-sur-Yvette 91190, France. {desnoullez, louise.duloquin, nadine.peyrieras}@inaf.cnrs-gif.fr

[†]T. Savy, C. Melani and P. Bourguine are with CREA-École Polytechnique, Paris 75015, France. {paul.bourguine}@polytechnique.edu

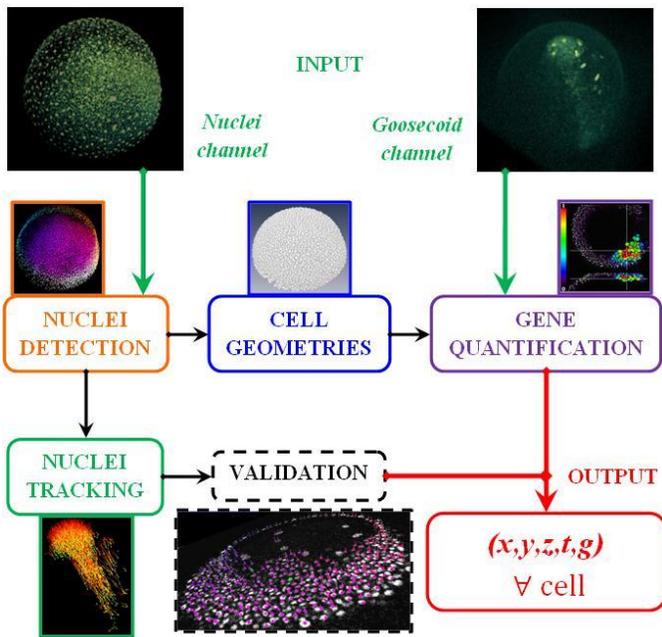


Fig. 1. Block diagram depicting the framework.

detected nuclei at I_t in I_{t+1} , then assign the closest detected nuclei in $t+1$ to be the continuation of its trajectory. Mitosis are identified by the Hough transform, which detects dividing nuclei -which are not longer spherical-, and by the presence at I_{t+1} of two nearest-neighbors. After visual inspection, the experts assessed the classification to have at least a 95% of correctly tracked cells per time step.

B. Cell Geometries Extraction.

Given that our data set did not dispose of membrane images, we used the detected nuclei centers as seeds to generate their corresponding Voronoi regions within the embryo volume. The Voronoi region associated to each nuclei n_i can be described as the set of points which are closer to n_i than to any other nuclei, (Fig. 1). In [8] we showed that these Voronoi regions, compared against a watershed-based segmentation, can provide an approximate model of the true cell geometries.

C. Automatic Gene Quantification.

We assumed the gene expression level to be proportional to the fluorescence brightness within the embryo. Therefore, our approach was to use the mean intensity value of the *gsc* channel within each segmented cell in order to assign them a genetic activity score ranging from 0 to 1. However, there are many factors that can significantly bias these scores. In consequence, we performed two kinds of normalization: first, in order to compensate for differences on the depth extinction and object transparency, we weighted our cell scores by the luminosity of their corresponding nuclei (which is supposed to keep constant at all cases). Then, to enable comparison across acquisitions with different gain, saturation and offset conditions, we applied an intra-step normalization, meaning that values assigned to each cell were not absolute

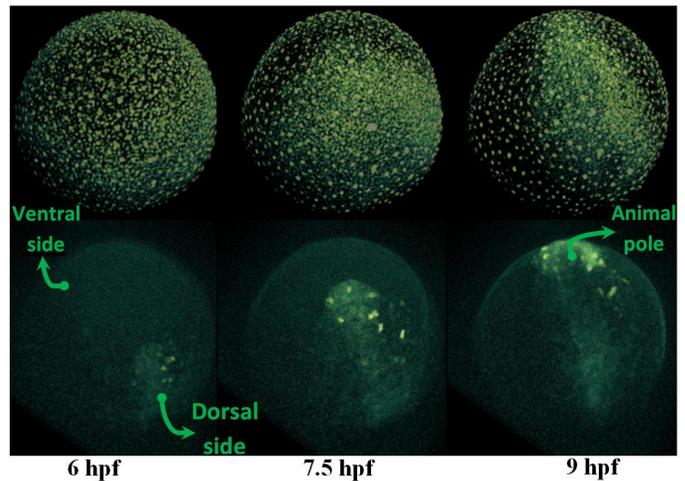


Fig. 2. Nuclei (top) and *gsc* (bottom) channels at 6, 7.5 and 9 hpf.

but dependent on how they rated against the rest of the cells present at their same time stage.

D. Validation.

We used the visualization tool Movit [11] which was specifically designed to visualize and validate the described data. Its validation unit includes the possibility of correcting false positives (removing nuclei), false negatives (adding nuclei) and adjusting nuclei positions. It is also possible to verify cells lineage by validating the links between their temporal trajectories, removing false mitosis, creating new links to manually added nuclei, etc. This utility was employed to label and verify the trajectories of 10 clones (10 cells together with all their progeny) which were used to study in detail the evolution of *gsc* expression and cell speed throughout the cell lineage (Section III-C).

III. EXPERIMENTS

A. Image Acquisition

The image data set used in this paper was acquired by a bi-photon laser scanning microscope Leica SP5 with a 10X objective and 1030nm and 980nm laser light excitation which resulted in two differentiated channels: one holding the *gsc* expression in the zebrafish, the other holding the embryo nuclei. We developed a new GFP transgenic line to label the *gsc* expression, while the nuclei were labeled with a H2bmCherry mRNA injection. The resulting 3D stacks are composed of a set of 2D images, homogeneously spaced in depth, acquired in the XY plane from the animal pole. Image size was 512x512x245 voxels with a voxel resolution of 1.51x1.51x1.51 μm . The repetition of this imaging process once every 2m44secs generated a sequence in time depicting the embryo *in vivo* development between 6 and 9 hpf (Fig. 2).

B. Results: Statistical Global Patterns

Following the scheme depicted in Fig. 1 our gene quantification strategy was employed to obtain measures of the *gsc* levels of expression at every cell detected between 6 and

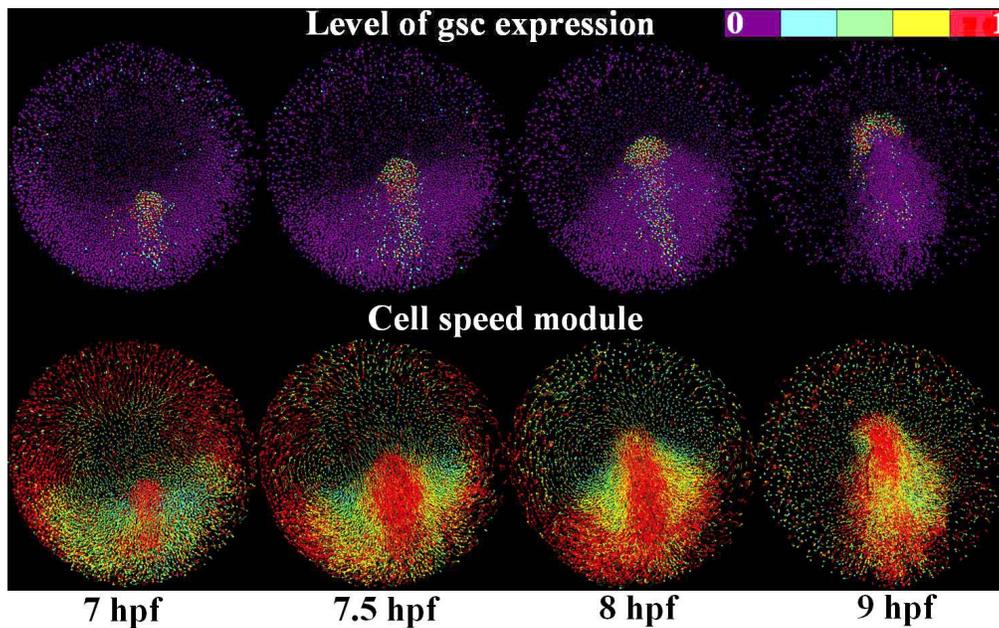


Fig. 3. *Top row*: *Gsc* levels of expression, each colored dot stands up for one cell. *Bottom row*: Speed modules, each colored dot stands up for one current cell whereas the colored lines represent their future positions. Views are taken from underneath the animal pole.

9 hpf. Likewise, our cell tracking strategy let us compute the current speed and direction of each cell during that period (Fig. 3). After comparing both *gsc* expression and cell speed, we can observe (Fig. 3) that movements towards the dorsal side -the site where *gsc* starts its expression- start with cells sited at the opposite ventral side which do not express *gsc* at first. However, as the time goes by, we can appreciate that there is a crescent connection between those cells expressing *gsc* and those moving quickly together to the animal pole.

Studying cells speed vectors altogether with their levels of *gsc* expression pointed out that cells expressing *gsc* tend to move fast and they do it together in the same direction. Their speed vectors steered up coordinate and progressively from bottom to top as the *gsc* expression evolved from the dorsal side to the animal pole.

C. Results: Linking Lineage and Cell Gene Expression

Ten clones, chosen to representatively sample the cell migration, were manually validated and labeled, including the six children derived from their mitosis, in order to accurately illustrate the evolution of *gsc* expression and cell speed across the lineage tree. Fig. 4 shows the complete cell trajectories of the corrected cells indicating their respective starting and end points. Some representative results, shown in Fig. 5, indicate for instance the fact that the gene level of expression is not constant throughout the cell life. This suggests that *gsc* expression does not follow a clonal behavior but it is rather dynamical since one cell can turn on and off the gene along its life. As a consequence, the gene evolution in time can not be adequately conserved by just propagating its expression through the cell tracking. In other words, *gsc* level of expression is neither constant nor restricted to the same cell population during the embryo development and do

not strictly follow a clonal descentance behavior.

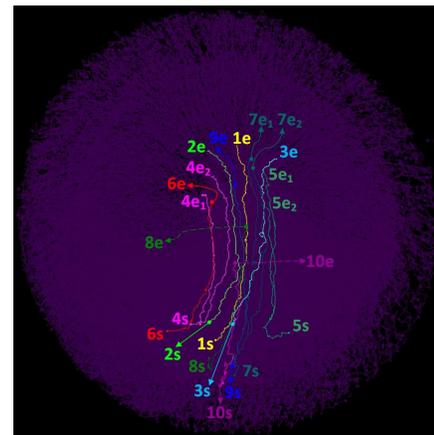


Fig. 4. Complete trajectories of the 10 validated cells from their start (s) at 6hpf to their end (e) at 9hpf. View taken from underneath the animal pole.

IV. CONCLUSIONS AND FUTURE WORK

We have proposed a general framework that uses a set of computational tools to automatically extract all cells trajectories and their progeny as well as their quantitative levels of expression of *gsc* in a zebrafish embryo undergoing gastrulation between 6 and 9 hpf. The resulting knowledge of the (x, y, z, t, g) space let us establish implications about how genetic expression relates to the cell movements and descendants. For instance, the collective nature of cell migration was established as well as the fact that gene expression is not deterministically attached to progeny links. We detected cells flocking into the area where *gsc* was expressing even though they were not yet expressing *gsc*. It appears, thus, that

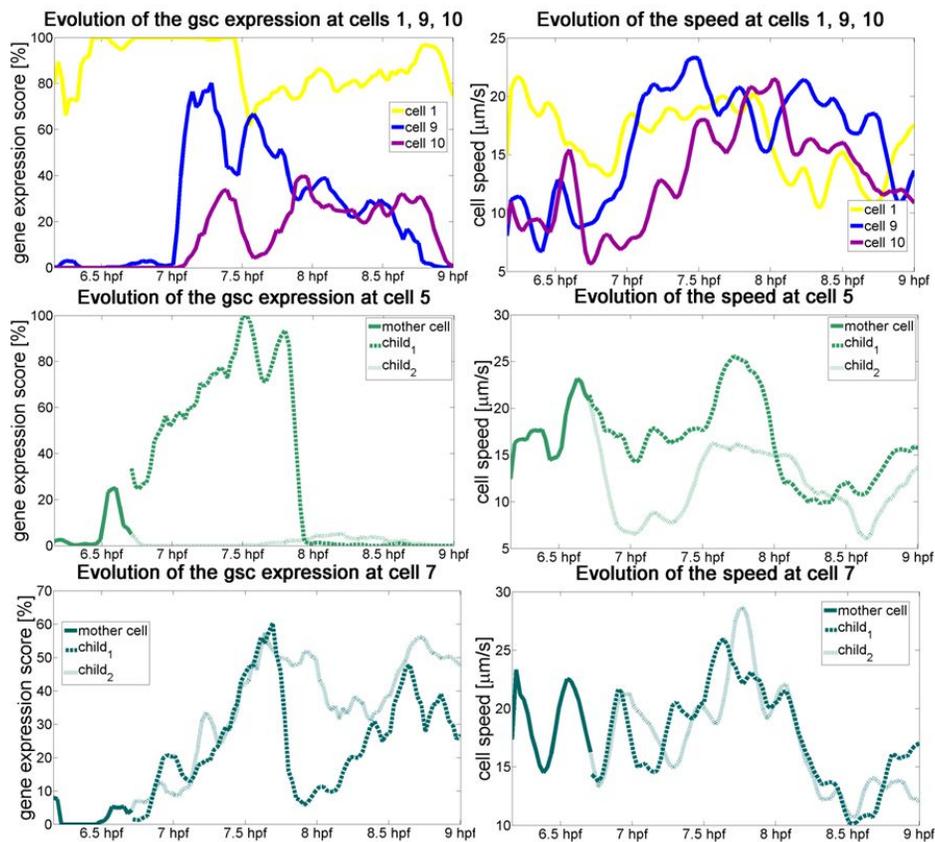


Fig. 5. Relationship between the levels of *gsc* expression and speed. The top row shows 3 cell trajectories with no mitosis: Cell 1, placed at the head of the migrating population shows a strong *gsc* expression from the beginning at 6 hpf to later on decay when the gene has already reached the animal pole at 9 hpf. Cell 9, placed in the middle of the migrating population, only shows *gsc* transcriptional activity during a short period, centered around 7.5 hpf, to rapidly decay afterwards when the *gsc* pattern moves ahead into the animal pole. Cell 10, placed at the tail of the migrating population only starts to timidly show *gsc* activity around 8 hpf when it flocks into the rear part of the *gsc* domain. In the three cases we can see a connection between their levels of *gsc* expression and their speed modules. The middle row shows cell 5 undergoing mitosis at 6.7 hpf. Being placed on the margin of the *gsc* domain, one of the children cells starts expressing *gsc* and completes its migration to the animal pole whereas the other do not show any expression and ends up moving slower and finishing its migration earlier. The bottom row shows cell 7 undergoing mitosis at 6.7 hpf as well. Being placed on the central area of the *gsc* domain, both children show similar levels of *gsc* expression and follow resembling speeds and trajectories to the animal pole in this case.

gene expression does not exclusively follow cell movements: *gsc* transcription can be active at one cell trajectory for a while, then vanish. Therefore, just propagating gene expression through the tracked itineraries of detected cells do not appear to be an accurate way to model its domain evolution. Finally, the proposed scheme can be easily extended to further gene products and developmental stages. Future work includes adding multiple gene expressions into the same *in vivo* embryo to study further relationships among them and creating fate maps for cells in later developmental stages.

REFERENCES

- [1] C. Castro, M. Luengo-Oroz, S. Desnoullez, L. Duloquin, L. Fernández-de Manuel, S. Montagna, M. Ledesma-Carbayo, P. Bourguine, N. Peyriéras, and A. Santos, "An automatic quantification and registration strategy to create a gene expression atlas of zebrafish embryogenesis," *Proc. IEEE EMBS Conference*, vol. 1, pp. 1469–1472, 2009.
- [2] O. Drblikova, M. Komornikova, M. Remesikova, P. Bourguine, K. Mikula, N. Peyriéras, and A. Sarte, "Estimate of the cell number growth rate using PDE methods of image processing and time series analysis," *J. Electr. Eng.*, vol. 58, no. 7, pp. 86–92, 2007.
- [3] C. Fowlkes, C. Hendriks, S. Keranen, G. Weber, O. Rubel, M. Huang, S. Chatoor, A. DePace, L. Simirenko, C. Henriquez, *et al.*, "A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm," *Cell*, vol. 133, no. 2, pp. 364–374, 2008.
- [4] P. Keller, A. Schmidt, J. Wittbrodt, and E. Stelzer, "Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy," *Science*, vol. 322, no. 5904, pp. 1065–1069, 2008.
- [5] C. Kimmel, W. Ballard, S. Kimmel, B. Ullmann, and T. Schilling, "Stages of embryonic development of the zebrafish," *Developmental Dynamics*, vol. 203, no. 3, pp. 253–310, 1995.
- [6] E. Lein, M. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A. Boe, M. Boguski, K. Brockway, E. Byrnes, *et al.*, "Genome-wide atlas of gene expression in the adult mouse brain," *Nature*, vol. 445, no. 7124, pp. 168–176, 2007.
- [7] F. Long, H. Peng, X. Liu, S. Kim, and E. Myers, "A 3D digital atlas of *C. elegans* and its application to single-cell analyses," *Nature Methods*, vol. 6, pp. 667–672, 2009.
- [8] M. Luengo-Oroz, L. Duloquin, C. Castro, T. Savy, E. Faure, B. Lombardot, P. Bourguine, N. Peyriéras, and A. Santos, "Can voronoi diagram model cell geometries in early sea-urchin embryogenesis?" *IEEE ISBI 2008*, pp. 504–507, 2008.
- [9] S. Megason and S. Fraser, "Imaging in Systems Biology," *Cell*, vol. 130, no. 5, pp. 784–795, 2007.
- [10] C. Melani, M. Campana, B. Lombardot, B. Rizzi, F. Veronesi, C. Zanella, P. Bourguine, K. Mikula, N. Peyriéras, and A. Sarti, "Cells tracking in a live zebrafish embryo," in *Proc. IEEE EMBS Conference*, 2007, pp. 1631–1634.
- [11] T. Savy and EMBRYOMICS, "MOVIT: Morphogenesis Visualization Tool," unpublished, 2010.