Live cell imaging and tracking for studying transcription

Joshua N. Benabou (Ecole Polytechnique)

under the supervision of Thomas Gregor, Jerome Wong Ng, and Benjamin Zoller (Physics of Biological Function, Insitut Pasteur)

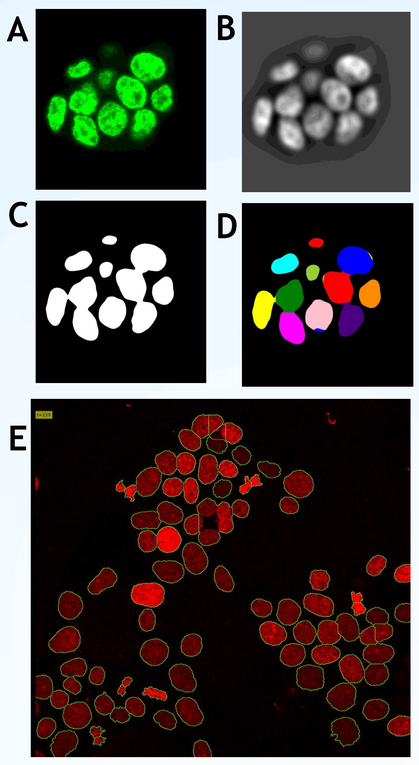
I. Summary

- **Part 1:** Image analysis to identify and track nuclei/cells from fluorescence microscopy time-lapses
- **Part 2:** Analyzing tracking data giving the intensity of transcription spots of two alleles of the *hunchback* gene in each nucleus of the developing Drosophila embryo. **Goal:** quantify the transcriptional noise and look for a possible temporal correlation in the expression of the two alleles.

II. Image Analysis : Segmentation

Segmentation method is specific to the type of image being analyzed (harder in 3D).

Fig II. A- D : 3D image (z-stack, 32 slices) of mouse embryo containing 15 cells, nuclei stained with GFP; implementation in Python using scikit-image.



Segmentation steps :

- Raw image (**A**)
- Filter via difference of Gaussians (**B**)
- Convert to black and white mask via thresholding
- Clean mask by filling holes and removing small objects (**C**)
- Apply marker-based watershed algorithm to the mask (Markers derived from distance transform of the mask) (**D**)

We then tried to automate segmentation of a time lapse of mouse ESC culture as the cells move and divide. We did not achieve an accurate segmentation at every time-point.

Fig II. E: Example of poor segmentation. Frame displaying both over and under segmentation.

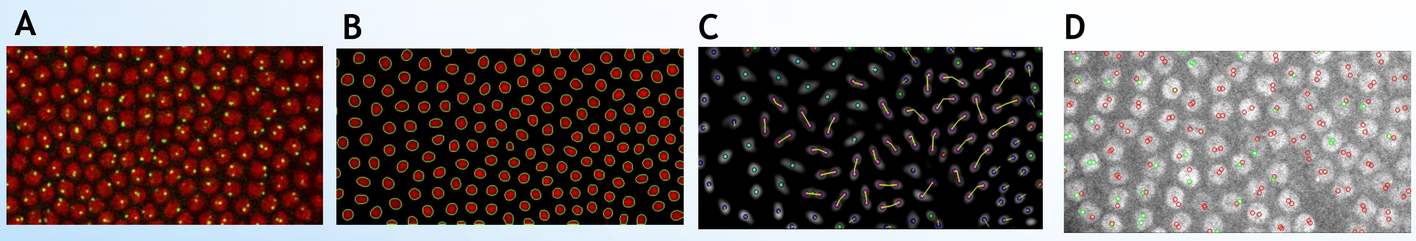
III. Image Analysis: Tracking

The lab developed a pipeline to segment+ track nuclei as they move and divide from 3D time-lapse of the developing Drosophila embryo. Accurate segmentation is possible due to good spacing between nuclei and uniformly distributed intensities.

Tracking method (simplified): For each frame:

- Calculate centroids of nuclei from segmentation
- Link nuclei from one frame to the next by solving a partial linear assignment problem allowing dividing nuclei to be assigned to its two daughters in the next frame (Hungarian algorithm)
- Cost function is distance between centroids

Fig A-D: Tracking results on embryogenesis time-lapse (see **IV. Experimental setup**). Images do not correspond to the same time-point.



(**A**) Raw image , nuclei (red), Hunchback-P2 transcription sites (green)
(**B**) Segmentation, nuclei boundaries in green
(**C**) Tracking information, e.g velocity vectors, pairs of daughter nuclei (pink), nuclei about to divide (green)
(**D**) Hunchback-P2 transcription spots identified as intensity peaks within each nucleus, also tracked

IV. Hunchback-P2 transcription in the Drosophila embryo

Hunchback is a maternal effect gene important for patterning of anterior parts (head and thorax) of the *Drosophila* embryo. Drosophila embryogenesis has been shown to be very precise. To understand why, we might study the noise in Hunchback expression.

Experimental setup:

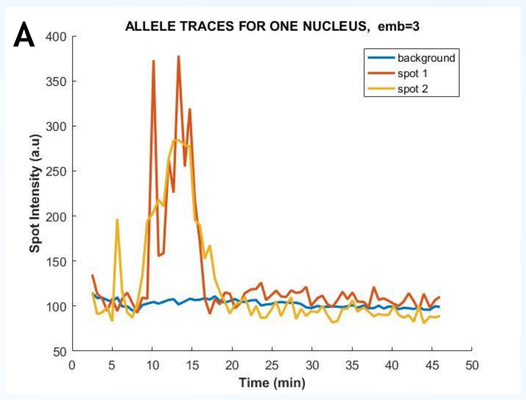
-Live image embryos during last several nuclear cycles (NC) with confocal fluorescence microscope
-Visualize nuclei with histone-RFP expression
-visualize hunchback P2 promoter transcription sites with MCP-GFP (two sites per nucleus)
-track spots through each cycle (MATLAB pipeline developed by the lab)

V. Tracking Data Analysis

- Time series data for 14 embryos extracted from tracking pipeline
- Before pooling data, spatial alignment along the anterior-posterior (AP) axis and temporal alignment (start time at NC14 mitosis)
- Mean background intensity subtracted from spot intensities

Fig A: Transcriptional activity of two spots in a single nucleus

Fig B: expression levels in a single embryo, in a given AP position bin (size 2.5% egg length). (binned mean in red) . Estimated NC14 peak is 10-12 min after mitosis.



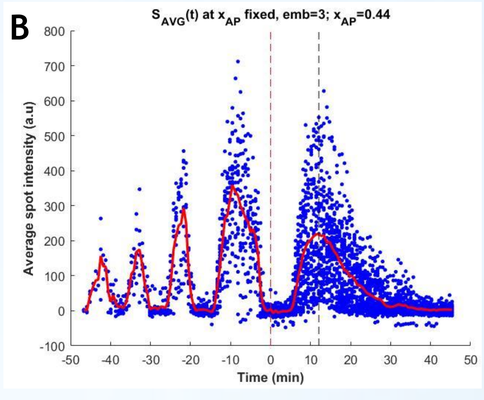


Fig C-G: Average spot intensities in nuclei at the peak of NC14 (AP position bins are 2.5% egg length)

(**C**) Raw spot intensity data as a function of position along the AP axis

(**D**) Mean expression profile for each embryo and for pooled data. We identify three regions: (I) high, (II) transition, (III) low.

(**E**) Standard deviation as a fraction of mean (roughly 40%)

(**F**) Decomposition of variance. Embryo to embryo variability represents 10 to 20 percent of total variability in region (I).

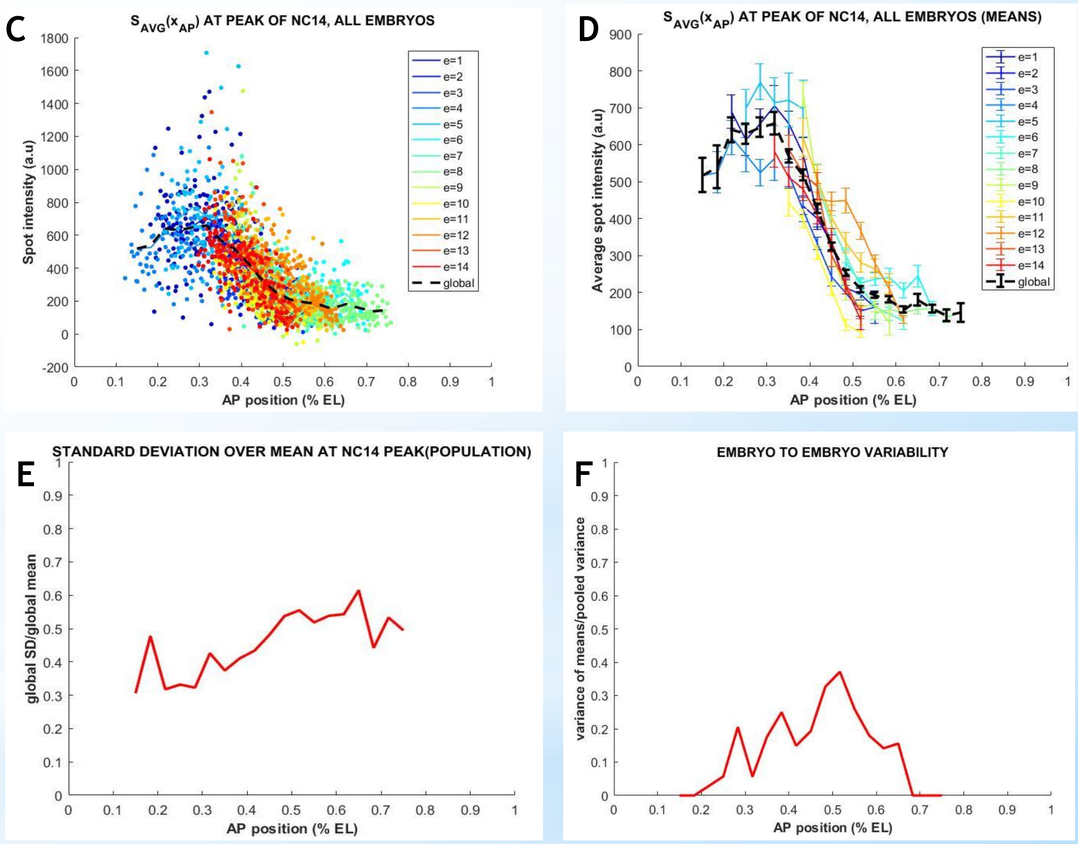
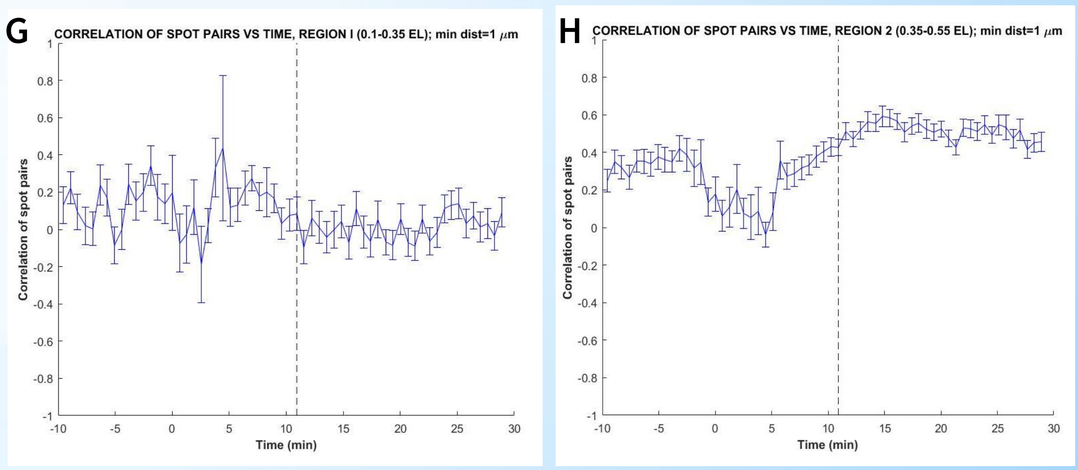


Fig G: Correlation of spot pairs in region (I) vs time, removing pairs closer than 1 micrometer (50% of pairs). Error bars = bootstrapped standard deviation.

Fig H: Correlation of spot pairs in transition region vs time. The increasing correlation could be spurious, possibly due to spatial misalignment.



Email: joshua.benabou@polytechnique.edu