

# Topological Data Analysis of pattern formation of human induced pluripotent stem cell colonies

Iryna Hartsock<sup>1</sup>, Daniel Cruz<sup>2</sup>, Eunbi Park<sup>4</sup>, Jack Toppen<sup>4</sup>, Peter Bubenik<sup>1</sup>, Elena Dimitrova<sup>3</sup>, Melissa Kemp<sup>4,5</sup>

University of Florida, <sup>1</sup>Department of Mathematics, <sup>2</sup>Department of Medicine; <sup>3</sup>California Polytechnic State University, Department of Mathematics; <sup>4</sup>Georgia Institute of Technology and <sup>5</sup>Emory University, Wallace H. Coulter Department of Biomedical Engineering

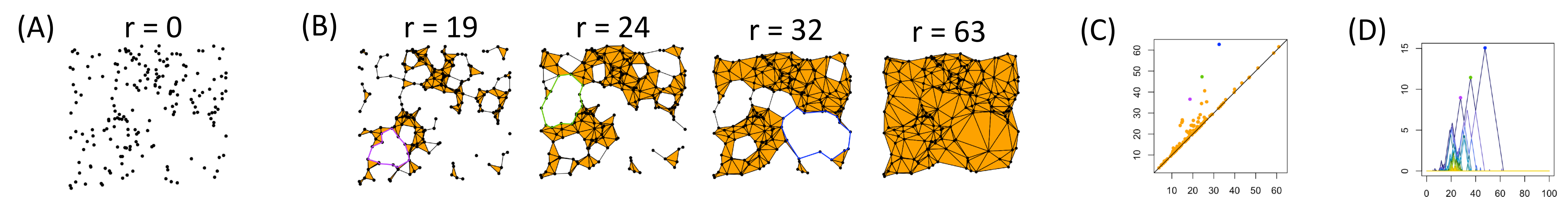
## Biology background

**Human induced pluripotent stem cells (hiPSCs)** have the potential to self-renew and the ability to differentiate into any cell of a human body. We utilize the hiPSCs line introduced in [1]. Also, we use various concentrations of a chemical called **doxycycline (Dox)** to increase the rate of cell differentiation which induces different pattern formations of stem cell colonies.

- Pluripotent stem cells we identified using a **NANOG** marker.
- Differentiated stem cells we identified using **GATA6** and **HA** markers.

## Topological data analysis

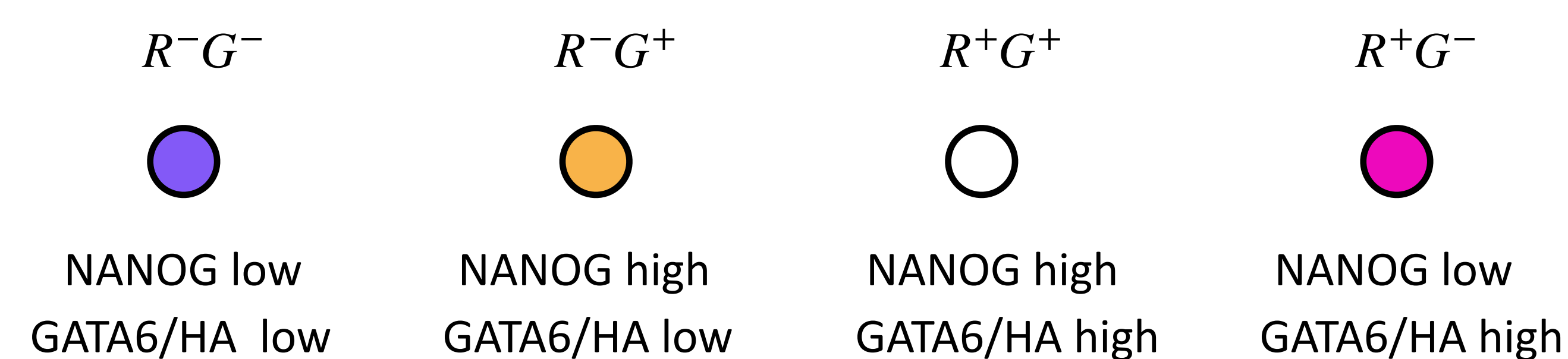
**Topological data analysis (TDA)** is a field which studies the shape of data with the complex geometric structure using techniques from topology. **Persistence homology** is a main tool in TDA. Persistence homology captures information about the numbers of holes in data and their size and then encodes it in a **persistence diagram**. We use **persistence landscapes** to transform persistence diagrams into feature vectors [2].



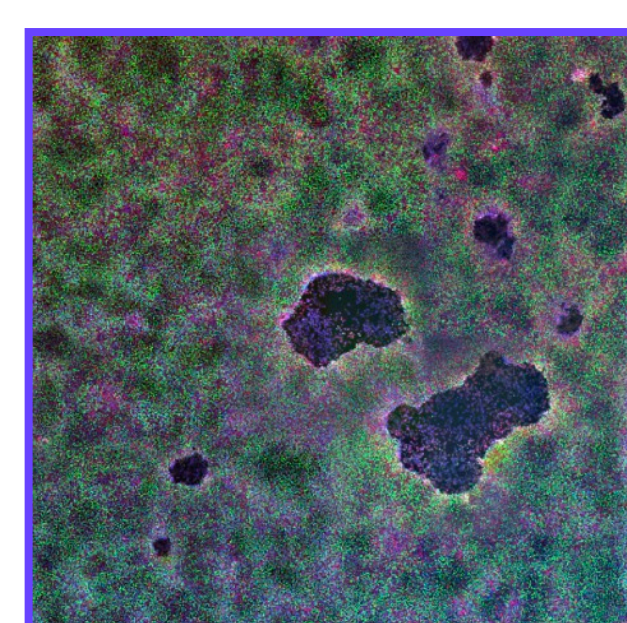
**Figure 1.** TDA pipeline: (A)  $400 \times 400$  pixels patch; (B) the filtered simplicial complex constructed on top of it (the three largest holes are identified at the scale they are “born”); (C) its corresponding persistence diagram and (D) persistence landscape.

## Data

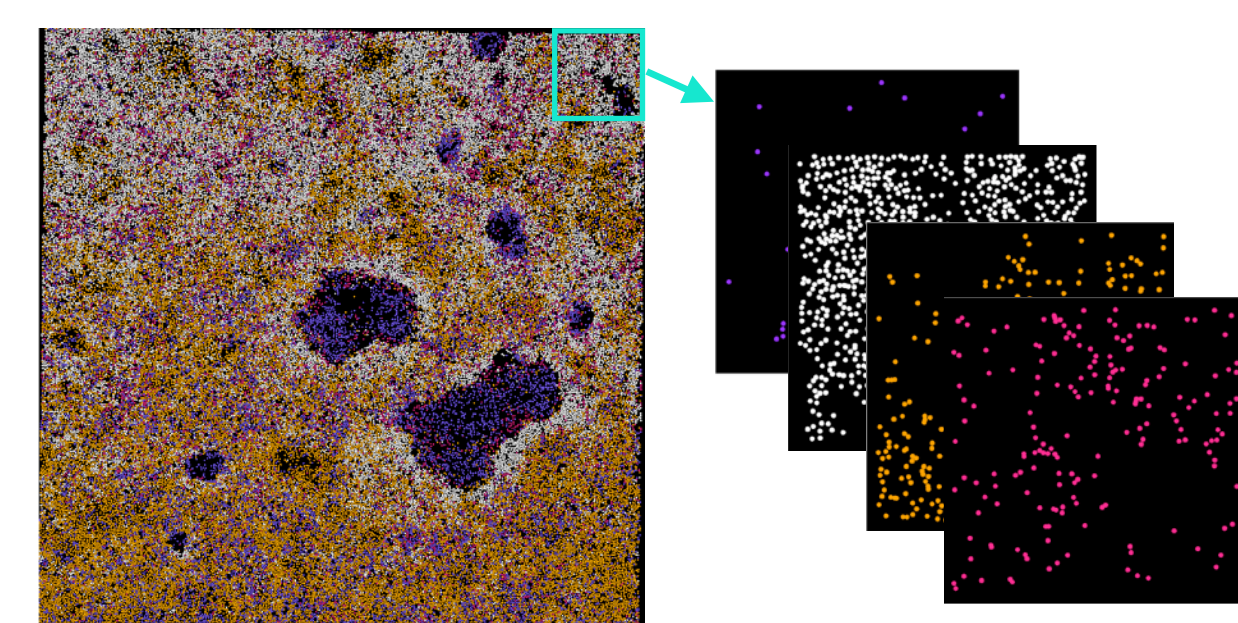
As input data for the TDA pipeline we use discretized microscopy images. After discretization cells are split into four cell types based on signal intensities.



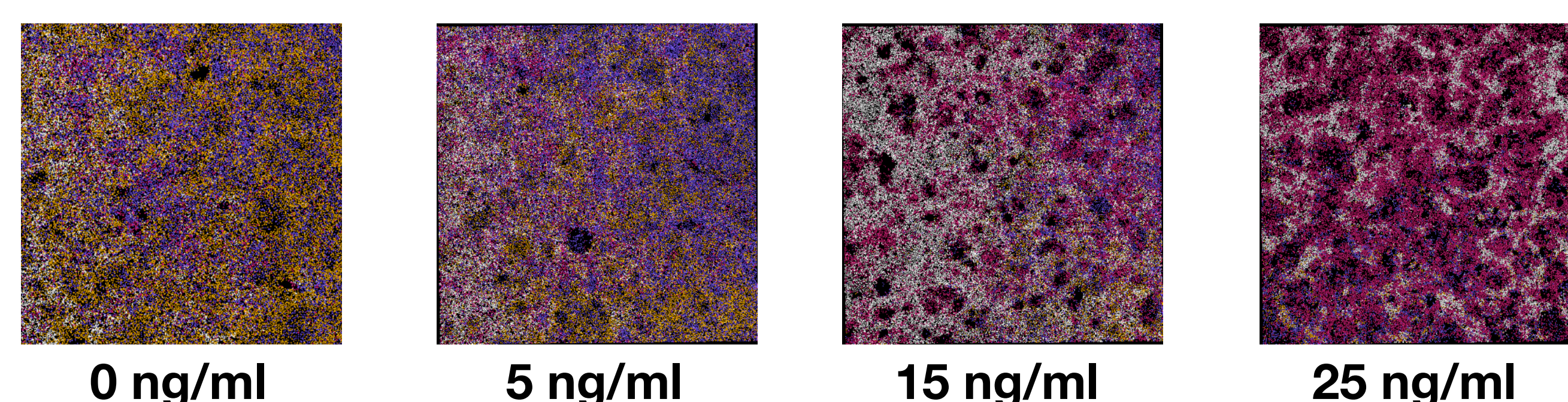
Microscopy image



Discretized image

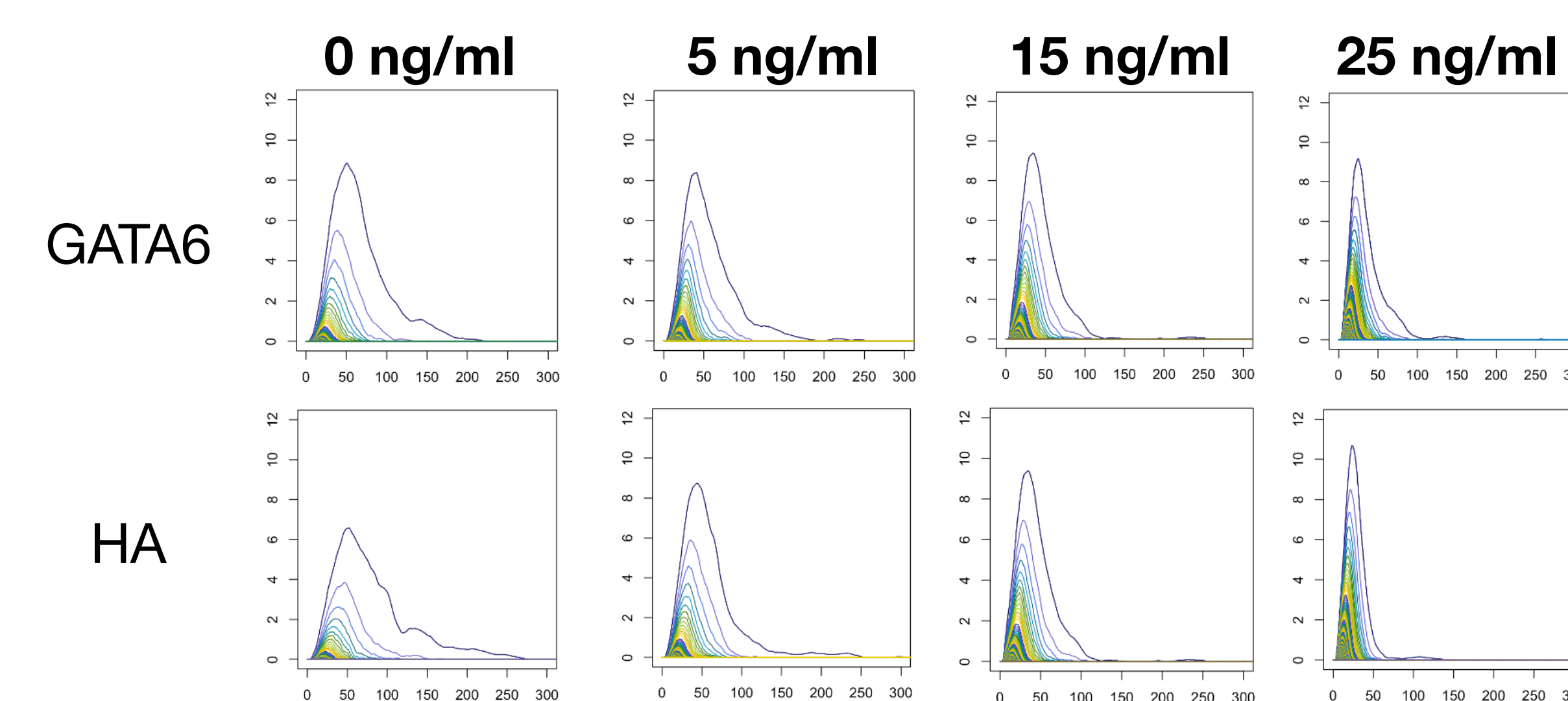


We consider four Dox concentrations: 0, 5, 15 and 25 ng/ml. For every concentration we have 15 images and every image we split into 16 patches.

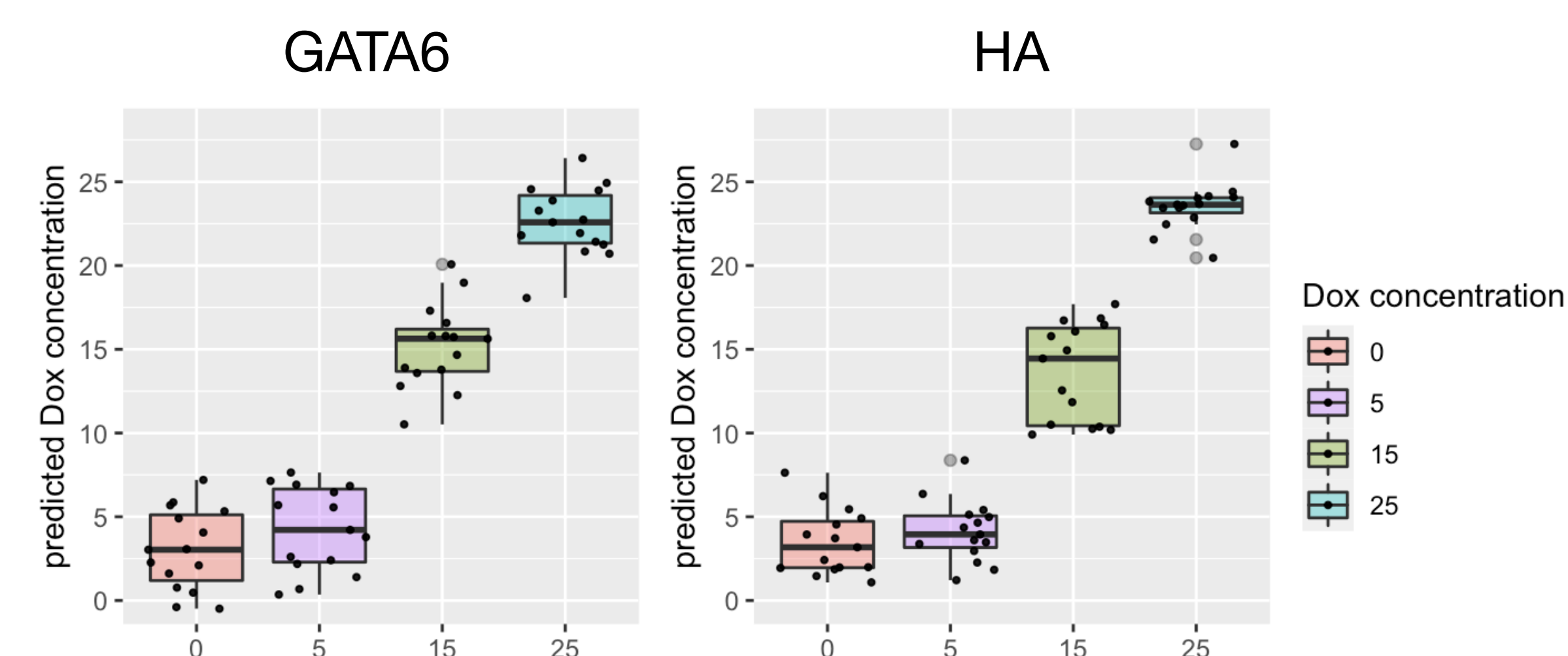


## Results - Dox concentration

Given a patch, we extract persistence landscapes for every cell type. Then we concatenate them into a vector and use it as input for machine learning computations.



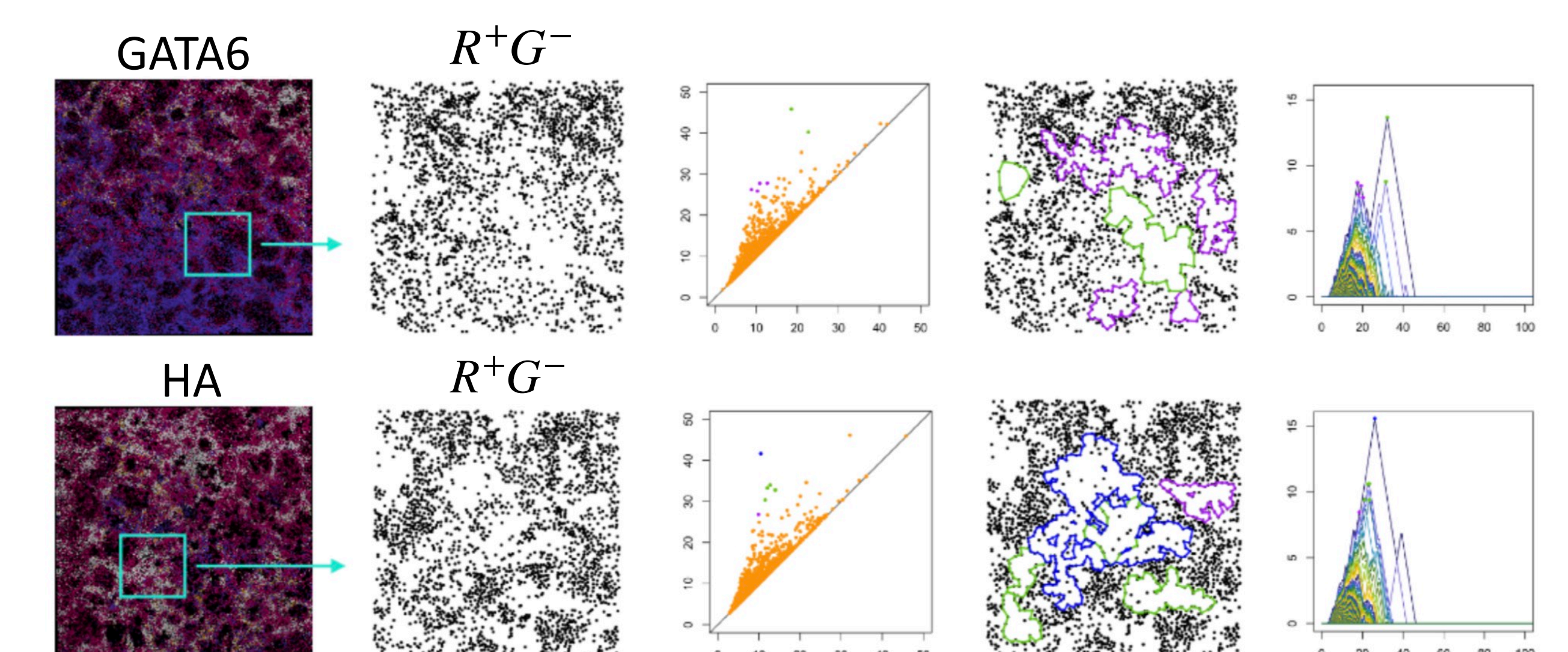
**Figure 2.** Average persistence landscapes for  $R^+G^-$  cell type.



**Figure 3.** Support vector regression. The median of predicted patch scores is assigned to every image.

## Results - GATA6 vs HA

We were able to quantify the differences in spatial organization of the stem cells based on the different markers. From the biological point of view, this can reveal insight in the strength of neighbor-to-neighbor signaling.



**Figure 4.** TDA pipeline applied to a patch of  $R^+G^-$  cell type from GATA6 and HA image with **25 ng/ml** Dox concentration.

## References

- [1] Guye, P. et al. *Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6*. Nat Commun **7**, 10243 (2016)
- [2] Bubenik, P. *Statistical topological data analysis using persistence landscapes*. J. Mach. Learn. Res., **16** (2015)