

# Assessing the Cancer Stem Cell distribution in tumorspheres using data science tools.

Jerónimo Fotinós<sup>1</sup>, María Paula Marks<sup>2</sup>, Lucas Barberis<sup>1,\*,+</sup>, and Luciano Vellón<sup>2,+</sup>

<sup>1</sup>IFEG, FAMAF, CONICET, UNC, Córdoba, Argentina

<sup>2</sup>Stem Cells Lab, IBYME, CONICET, Buenos Aires, Argentina

\*lbarberis@unc.edu.ar

+these authors contributed equally to this work

## ABSTRACT

In previous theoretical research, we inferred that Cancer Stem Cells (CSCs), the cells that presumably drive tumor growth and resistance to conventional cancer treatments, are not uniformly distributed in the bulk of a tumorsphere. This knowledge is useful for mathematical modeling, which most of the time assumes uniformity as a simplification. The CSCs distribution in a 3D tumor model, such as a tumorsphere, is crucial not only for understanding and designing novel and more specific anti-CSC therapies but, also for the design and check of mathematical and computational models.

In this article, we present a method to process confocal images from tumorsphere assays and measure the amount and location of their Cancer Stem Cells. Guided by a previous computational model, we conclude that the distribution of Cancer Stem Cells in an experimental tumorsphere must be non-homogeneous.

## 1 Introduction

Cancer Stem Cells (CSCs) are defined by their capacity to replicate making copies of themselves or to differentiate giving rise to phenotypically diverse cells. They are resistant to conventional anti-cancer treatment, being thus implicated in disease recurrence and metastasis<sup>1,2</sup>. One of the main barriers to the development of CSCs-targeted therapies is their scarcity *in vivo*, limiting their availability as experimental systems for pharmaceutical development, and raising the need for production at a scale large enough to fulfill academic and industry requirements. A common solution is the anchorage-independent growth of cancer cells to generate a 3D culture of epithelial cells<sup>3</sup> enriched in cells with CSCs-like properties, since most epithelial cells are killed by anoikis<sup>4</sup>. Thus, taking into account the limitations of this assay, the resulting tumorspheres are formed by the clonal expansion of a single cell, instead of the self-aggregation of existing cells<sup>5</sup>.

Still, routinely used cell culture techniques are material- and labor-consuming tasks that generate a great amount of interculture variability and contamination risks. Moreover, traditional cell culture at a large scale is also cost-ineffective in terms of the high investment in cell culture media and growth factors. In this context, developing forefront, high-throughput screening platforms to identify cytotoxic inhibitors and/or differentiation-promoting agents targeting CSCs becomes of paramount importance. This would require the optimization of a series of bioprocesses that enable the massive culture of undifferentiated cancer cells and the analysis of high volumes of data.

One critical downstream part of such bioprocesses is to assess the cellular response in terms of viability and/or stemness markers, which requires external software for image analysis and segmentation to quantify the relative differences among treatment groups. Even though high-throughput cytometric methods have been developed<sup>6</sup>, there is still the need to know the exact identification, location, and targeting of putative CSCs. This would help the modeling of CSCs dynamics, allowing the development of cost-effective and predictive tools for the examination of tumor evolution and response to therapy. Mathematics, physics, and computational modeling offered positive prospects in cancer research, being widely used. Many biological problems, including cancer disease, demand methods and techniques requiring applied mathematics, statistics, fluid and phase transitions physics, population dynamics, and computation<sup>7-9</sup>. Indeed, analytical mathematical methods allowed us to estimate the expected fraction of CSCs for tumorspheres in different culture conditions<sup>10-12</sup> and the effect of specific therapies on their development<sup>13</sup> among other theoretical results<sup>14-17</sup>.

We have also computationally simulated the growth of a colony of cells in two dimensions using an Agent-Based Model (ABM) that mimics basic features of CSCs proliferation to form a spheroid<sup>18</sup>. The simulated spheroid grows from a single CSC and the cells can undergo mitosis at a fixed rate (the population doubling time or PDT). Depending on the intrinsic and extrinsic (micro-environment) signals, the CSC will replicate generating another CSC with a certain probability  $p_s$ , or it will differentiate generating a *differentiated cancer cell* (DCC). This simplified model allowed us to estimate the total number of

Lucas

la frase que sigue no entiendo por qué es importante. Como esencial no lo es, sencillamente describir el ensayo de esferas como un ensayo de resistencia a anoikis, es decir, a la muerte por falta de adhesión a un sustrato. (I) A mí me gusta esta aclaración, quizás pondría una coma antes de enriched y sacarla al punto antes de since.

Lucas

No sé a qué limitación se refiere. Las limitaciones son: cuidar la densidad de cultivo ya que las esferas deben ser clonales y que no detecta células madre quiescentes ya que quedarían como célula única al no dividirse, están enumeradas en la referencia, pero si es conflictivo saco esa parte. (J) Quizás convenga mencionarlas brevemente o poner la referencia al lado de limitations. (L) no, corta demasiado la frase, está bien así porque la ref-

CSCs, the fraction of CSCs situated on the periphery of the colony, and the size of the whole spheroid, showing that these traits are dependent on the replication probability ( $p_s$ ) of the CSC. Indeed, simulating with intermediate replication probabilities, we observed active CSCs at the border of the colony and detected that they form a path that links the center of the colony with its border. Furthermore, an increase in the replication probability led, as expected, to a large CSCs population that overtook the system. This last situation may describe most experimental conditions used for culturing tumorspheres<sup>3,19–22</sup> and agrees with our previous mathematical models<sup>10–12</sup>.

Inspired by these simulations, we generated tumorspheres from MCF-7 cells and analyzed them by confocal microscopy. We developed an advanced computational method to detect the expression and distribution of the stem-like cells (Sox2-positive cells), under the assumption<sup>21</sup> that this stemness factor is expressed in tumorspheres from cell lines and primary cultures. The processed images were used to study the distribution patterns of CSCs employing statistical tools that validate our main computational finding: that CSCs are heterogeneously distributed in a tumorsphere.

Summarizing, we measured a non-homogeneous distribution of CSCs in tumorspheres after a computational model predicted it. In particular, we report a protocol to analyze confocal images of tumorspheres that were stained with a stemness marker, to statistically infer the more suitable threshold to determine which cells in the culture belong to the stem phenotype. We highlight the capability of our method to extract data from the experiments allowing direct comparison with simulated cultures. With this purpose, theoretical work was useful to drive experiments and, conversely, ad hoc experiments become crucial to develop more accurate theoretical work.

## Results and discussion

According to simulations<sup>18</sup>, CSCs must be connected in tumorspheres forming “paths” that join the center of the spheroid with its border. To assess this, we performed tumorspheres assays where MCF-7 cells in suspension culture proliferate in a solution enriched in growth factors to ensure a large fraction of CSCs. After 9 days of growth, we collected the spheroids and attached them to a slide by cytocentrifugation. Due to the centripetal forces, their spheroidal shape may be lost, becoming discs smashed against the slide, however, the structure of the tumorsphere is maintained. We stained the slides to detect the location of all cellular nuclei (DAPI) and the corresponding primary and secondary antibodies needed to detect the position of the stem cells (SOX2). Finally, we took pictures using a confocal microscope, slicing the spheroids and obtaining a full 3D reconstruction. Because of the smashing of the spheroids, only one or two slices of the imaging were needed to observe all the cells present in them. These images were processed with a Data Science approach, using computer vision and statistical analysis techniques, among others. The result of the process allows us to mathematically reconstruct the discs, specifying the position of all their cells, marking those that are candidates to be CSCs and statistically deciding which ones are in fact CSCs. The experiment and the image processing procedure are fully detailed in the 1 Methods section.

### The Cancer Stem Cells distribution.

Our main results are summarized in Fig. 1, in which we can observe the distribution of the two cell phenotypes: CSCs in red and DCCs in blue. We present three representative examples belonging to the reconstruction of three different spheroids. After the filtering and reconstruction process of the confocal images, we obtained pictures of the cell’s distribution as a Voronoi tessellation (see Methods section). That is, each cell is represented, in Fig. 1, as a polygon whose centroid coincides with the centroid of the corresponding cell in the pictures. Note that this representation is just an approximation of the shape of the cell that allowed us to quantify the SOX2 content inside the cell and decide, using statistical tools, if the cell belonged to the stem or to the differentiated phenotype. After the whole process, the CSCs become represented by red polygons and the DCCs by blue polygons.

Jero: I think we should replace *spots* by *stains*. Spot means place; stain transmits the intended idea.

### Spots in the border

Our first example, labeled Sph4, slice2 as in raw data files, a spheroid that has not grown much is depicted in Fig. 1A. Considering the different proliferative capacity of the cells in these culture conditions and the possibility of differential access to the growth factors, we expect inter-spheroid variability in both the number of cells and the CSCs fraction. In this case, the spheroid had less than a hundred cells, with half of them being CSCs, cf. Table 1. There is an evident higher concentration of the CSC in the core of the spheroids, which could be explained by the model used in our simulations: the first CSC has a given chance of generating another CSC but, when one of these CSCs differentiates the first time, its lineage will only contain DCCs that will surround its stem mother. The only exception is when a DCC cannot undergo mitosis in a nearby place, already occupied by a CSC. Indeed, this CSC will continue generating other CSCs until one of them becomes differentiated, leaving in the process a path of CSC among DCCs. This is exactly what is shown in panel (A) of Fig. 1: at an early stage, red CSCs become surrounded by blue DCCs in the lower portion of the spheroid. However, three CSCs were able to initiate paths that extend to the periphery of the spheroid. The result is that CSC will form spots in the border of the spheroid.

Jero  
Sox2 (lower case?) shouldn't be in parentheses if it's going to be referred to outside them.

Jero  
suspension cultures?

no seria centrifuga?

Jero  
this makes me think that we physically slice the spheroid

image?

Jero  
We need to redo this merged image preserving quality. Also, the caption needs a bit more explanation of the selected slices which is on top of which, and the fact that those slices allowed to reconstruct the entire corresponding spheroid.

Jero  
Maybe ‘if the cell expresses a stem of differentiated phenotype’; I don’t like *belong* in there.

Jero  
Note: the abbreviation of *confer* is either *c.f.* or *cfr.*, not ‘c.f.’.

Jero  
In Table 1’s caption, I’d replace the path hypothesis for simulations, or I’d explain what we mean by it.

A larger growth rate is associated with the case labeled Sph3, slice 3 shown in Fig. 1 (B), where the spheroid has more cells than the one shown in panel (A). Its CSC fraction is 64%, larger than the 54% of the previous case, a result almost trivial that supports our strongest modeling hypotheses: the growing rate can only be measured for the bulk of the cells, thus, the outcome of the mitosis of a CSC is given just by a probability rather than a specific growth rate for each cell phenotype<sup>12</sup>. However, the quantitative results reported in Barberis, 2021<sup>18</sup> must be significantly different when we shift from two to three dimensions, the qualitative ones remain the same: the CSCs will form paths and be heterogeneously distributed, forming “stains” at the border (surface) of the spheroids after approximately one week.

In this example, we can also appreciate the deformation produced by the smashing against the slide. Some DCCs appear in the bulk of the disk surrounded by CSCs. These DCCs initially belonged to the border of the spheroid and reached the center after centrifugation. In this way, it is easy to understand why cells closer to the center of the disk could be a mix of original center cells and some of the periphery. However, there is no way in which cells at the border of the spheroid could originally be from near the spheroid center.

### Two layers sample

As mentioned, our method for placing the sample on the slide deforms the spheroids, projecting the cells onto a disc. Depending on the forces acting on the spheroids, sometimes their cells can not push all their neighbors apart, piling up on top of them. As a consequence, we need to use the images of more than one slice to recover all the cells that constitute the spheroid.

An example of this is given by Sph1 which had two layers of cells superimposed and required two slices of confocal imaging to access all the cells. In the lower panels of Fig. 1 we depict the obtained cell distribution for the upper (C) and bottom (D) layers. As observed by direct inspection, and reported in Table 1, this spheroid has a larger amount of cells than those in the previous examples, but a much smaller CSC fraction. Thus, even though CSCs are required to drive tumor growth, their abundance is not directly related to this amount. This occurs because the first CSCs have a small chance to replicate (self-renew) becoming differentiated at early times and, as a consequence, being quickly surrounded by DCCs. This is again consistent with our description of the replication rate as a probability that models the chances of a CSC becoming undifferentiated either because of extrinsic or intrinsic causes.

In the two-layer disc, we have an example of such a situation, where it becomes difficult to follow the CSCs’ path. The CSCs seem to be less connected and not as clustered as in the previous cases, and might even seem more randomly distributed. Nevertheless, this case presents the opportunity for a better reconstruction of the spheroid due to the smaller loss of spatial information. Indeed, if we subtract the DCCs from the snapshots in panels 1 (C) and 1 (D), and overlay the remaining CSCs, we obtain the distribution shown in panel 1 (E). Note that we colored the bottom layer cells (against the plate) in dark red, and the upper layer in light red. The result is now evident, most of the CSCs are indeed connected, forming a path along the two layers as we expected from the simulations.

	Sph4, slice 2	Sph3, slice 3	Sph1, slice 3	Sph1, slice 4	Sph1, both slices
Radius ( $\mu\text{m}$ )	87.58	102.04	113.25	99.07	—
Total Cells	112	240	183	252	435
Stem Cells	59	155	55	56	111
Differentiated Cells	53	85	128	196	324
Stem Cell Fraction	53%	64%	30%	22%	26%

**Table 1.** Experimental values recovered after processing the images. For each slice, we report the radius of the spheroid, the total number of cells (as well as the number for each population), and the fraction of them that are stem cells. In the case of Sph1, we reported the measured values for each slice and the total values, summing up the data of the two slices in the last column. Note that this spheroid, being the larger one, has the smaller CSC fraction, in agreement with the path hypothesis.

Jero: Maybe we can call our model for the reproduction of the cells *path hypothesis*, since it causes the CSCs to make “paths”. It sounds like bit of marketing to me, but it might work. In any case, we should explicitly say this.

### Non-randomness of the CSC’s distribution

We measured some extra features that allow us to be sure that our findings have statistical significance, even for just a sample. The tessellation generated by this method defines the location of a cell and its neighbors, allowing to make a graph of the connections between cells. An example of this is depicted in Fig. 6A for Sph1, slice 3, where the CSCs are represented by red dots, and DCCs by blue dots. If two cells are neighbors, they are linked with a black line. With this graph, we can use network theory to measure how connected are the CSCs inside the spheroid structure.

At first glance, is easy to see that CSCs form the expected path. Moreover, we assessed if this is true by measuring how many CSCs paths are present in the graph and, most importantly, how apart one is from the other. The physics of complex

Jero  
I don't see how one thing implies the other, but more importantly, that is not a current modeling hypothesis, while a former one.

Jero  
Connector missing. Perhaps something like: Moreover, while the quantitative... (must change), the qualitative... (remains the same).

Jero  
Or they were just surrounded by CSCs, depriving them of space to reproduce. We don't want to give the impression that the model can't deal with DCCs near the center of the spheroid.

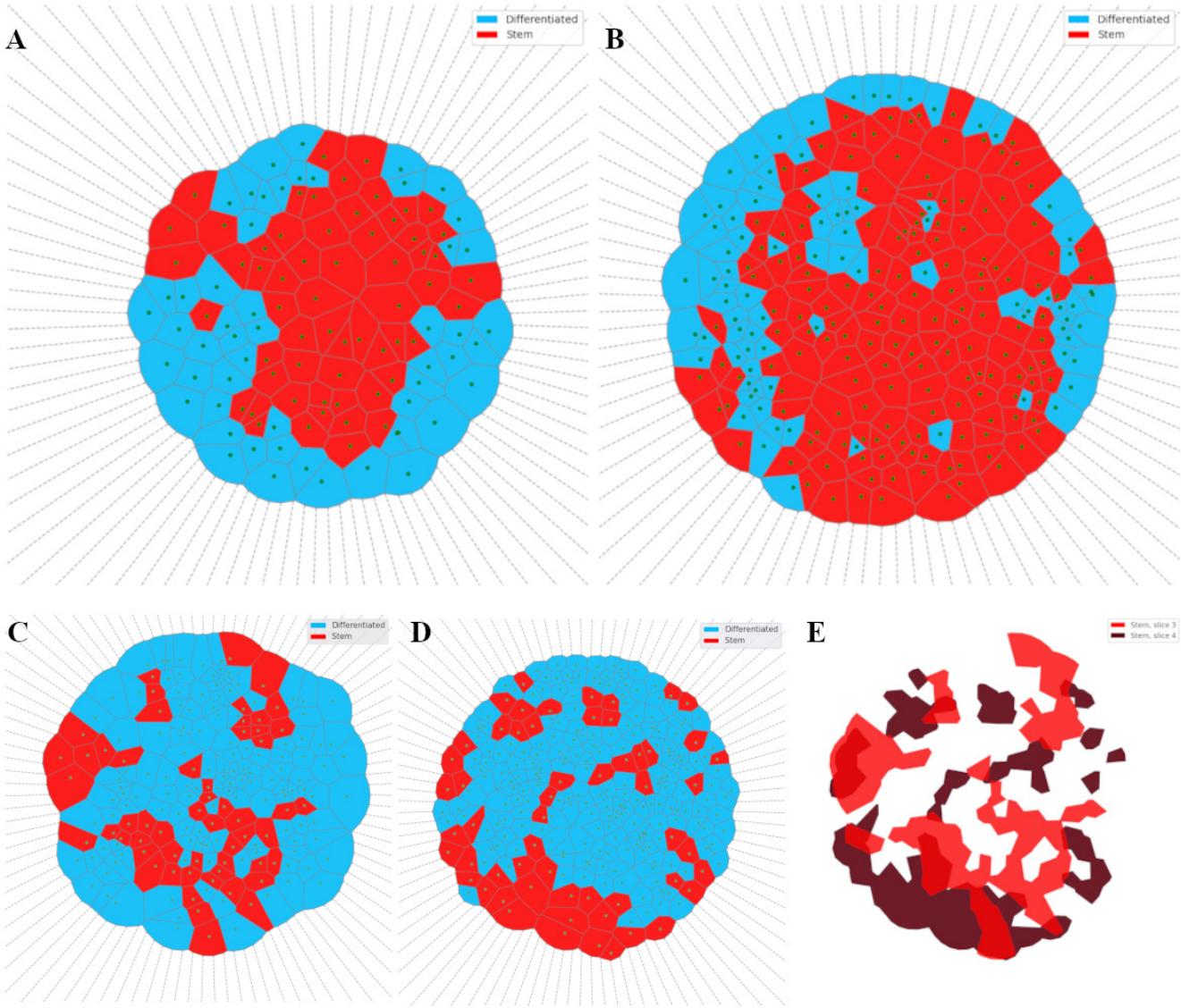
alters the shape of?

Jero  
Are thought to drive?

Jero  
Amount of growth? Maybe just ...their abundance may not be directly proportional to it.

Jero  
Small compared to what? Again, I'd start this by saying: This can be understood by our modeling hypothesis...

Jero  
This figure also needs to be re-merged.



**Figure 1.** Final reconstruction of the spheroids' slices (image processing end result). For different slices (i.e., focal planes) of the experimental spheroids, cells are shown approximated by polygons. The polygons representing CSCs are colored in red, and the ones representing DCCs, in blue. It is evident (but proven later) that the CSCs are not uniformly distributed in any image, and form stains at the border of the aggregate. The panels correspond to: **A** Sph4, slice 2; **B** Sph3, slice 3; **C** Sph1, slice 3; **D** Sph1, slice 4; **E** Overlay of CSCs of slices 3 and 4 of Sph1. This last panel shows how stem cell clusters are connected through the slices.

systems has developed tools to analyze graphs and collect nontrivial information. In our case, we determine if the paths formed by CSCs lay on deterministic rules or are formed by chance. As explained in the Methods section, we test for the experimental reconstructed networks their homophily, the tendency of cells to neighbor cells of the same phenotype. This is characterized by the *assortativity coefficient* which results positive for our networks. Also, the *homophily ratio*, the fraction of neighbors of the same kind that a node possesses, gives values larger than half. Both results mean that the nodes seem to prefer to be attached to nodes of the same kind. The measured values of these quantities for the four examples shown here, are reported in Table 2 under the columns labeled 'experiment'.

For a meaningful comparison and interpretation of these values, we used a classical approach from the physics of complex systems. We maintain the connections between the nodes of the previously obtained networks but, this time we arbitrarily set which ones are the CSCs, randomly distributing them in the same amount that appears in experiments, c.f. Fig. 6B. These

random distributions are one of the most used hypotheses in mathematical modeling. This will also resemble the case of CSCs that can travel inside the spheroid. We did this random distribution 10000 times and, then, performed statistical tests with significance  $\alpha_t = 0.001$ . These tests tell us whether the quantities measured in the experimental networks significantly differed from the same quantities measured in the random networks. In Table 2 we reported the averaged results over all the realizations including their standard deviation. The assortativity coefficient is now zero meaning that, as expected, there is non-preferential attachment of cells of the same kind. The random homophily ratio is a bit larger in the two-slice spheroid but still lower than the experiential case. Thus, experimental values tell us that daughter cells indeed stay close to their parent cell.

A way to double-check this result is by looking at the stem subgraphs. As shown in Fig. 6C, this is the network formed just by the CSCs and the connections between them. We measured the degree distribution of these subgraphs, i.e. the average number of connections that have the nodes. We also calculated the number of connected components of this graph, which is the number of separated groups of CSCs. The more clustered the CSCs, the less the connected components. These values are also reported in Table 2 both for experimental and the average of random networks.

For Sph3 and Sph4 the large CSC fraction made the experimental and ensemble values statistically indistinguishable from each other. This means that, because we have more than half of CSC in the spheroid, there is no chance to get many isolated CSC, independently of the way they are distributed. On the other hand, for Sph1 the measured number of connected components was significantly smaller than the one measured for the random ensemble stressing that, despite the CSCs being fewer than the DCCs, they are indeed connected. A close inspection of Figs. 6 and 1 reveal that most of the connected components are separated from each other by just one DCC as predicted by simulations. Thus, the number of *effective* connected paths, in the sense defined in<sup>18</sup>, should be even smaller.

The mean degree of the experimental networks was a bit higher than their corresponding random averages. This implies that CSCs are significantly more connected between them than expected if located randomly. This, in turn, reinforces our previous conclusion about the tendency of neighboring cells of the same phenotype.

	Sph4, slice 2		Sph3, slice 3	
	experimental	random	experimental	random
Assortativity Coefficient	0.39	-0.010 $\pm$ 0.04	0.47	-0.01 $\pm$ 0.06
Homophily Ratio	0.74	0.50 $\pm$ 0.03	0.73	0.54 $\pm$ 0.02
Stem Connected Components	2	3.7 $\pm$ 1.5	2	1.1 $\pm$ 1
Degree of Stem Subgraph	4.5	3.0 $\pm$ 0.2	4.8	3.7 $\pm$ 0.1

	Sph1, slice 3		Sph1, slice 4	
	experimental	random	experimental	random
Assortativity coefficient	0.22	-0.01 $\pm$ 0.04	0.26	-0.01 $\pm$ 0.01
Homophily ratio	0.67	0.58 $\pm$ 0.02	0.74	0.65 $\pm$ 0.01
Stem connected components	5	17 $\pm$ 3	12	25 $\pm$ 3.3
Degree of stem subgraph	2.7	1.7 $\pm$ 0.2	2.64	1.3 $\pm$ 0.2

**Table 2.** Network properties from the experimental network in columns labeled 'experimental'. The comparison with networks with randomly distributed CSCs is reported in columns labeled 'random'. The assortativity coefficient is zero for random stressing that there is a non-preferential attachment for them contrary to what happens in the experimental cases where there are always positive values. For Sph4 and Sph3 the connected components are similar due to the large amount of CSCs.

## Conclusion

Image processing is useful to obtain information that is highly relevant for setting up simulations, formulating mathematical models, and fitting their results. Here, we measured both, the total amount of cells in each spheroid and the respective CSC fraction, information that will lead to improving the accuracy of mathematical models such as the ones summarized in<sup>11</sup>. Furthermore, we estimate that the growth rate is around  $r = 1.1$  cell/day, which allows us to establish and set the temporal scale of simulations as those in<sup>18</sup>. This value is in agreement with the one obtained by fitting experimental data with mathematical modeling<sup>12</sup>.

The method of image processing presented here is devoted to recognizing the location of the CSCs in microscopy images. Its originality lies in the fact that we are not just looking at where the SOX2 fluorescence is high enough under a subjective criterion. The key is the use of the Gaussian Mixture Model to fit the data and separate, with a statistical criterion, both cell populations. Having done this, the way of depicting the spheroid or what is done with the resulting data will depend on the questions that researchers have in mind. In our case, we presented an example that compares the experimental CSC distribution

in tumorspheres with a previous computational model. Beyond the fact that we are pleased to find that *in silico* and *in vitro* experiments give similar outcomes, the method proposed here is straightforwardly applicable to other similar experiments. Indeed, to definitively assess the CSCs distribution in a spheroid, we must be able to put into the microscope non-deformed spheroids and take pictures of several slices of them. Furthermore, according to our computational simulations, the confidence in our result will exponentially increase with the cultured time. Indeed, we are carrying out full 3D simulations of tumorsphere assays. The preliminary results still support the finding that CSCs are heterogeneously distributed inside the spheroid. However, the probability of finding CSCs on the border of the spheroid seems to be significantly enlarged. These simulations plus the help of confocal images of several slices of a non-deformed spheroid, will be the basis for studying further therapies focused on treating cancer stem cells.

Discutir biologicamente la ubicacion de las CSCs en los esferoides y el acceso a los farmacos

Cerramos con chamuyo de Luciano sobre por qu'e es 'util esto en biolog'ia, los biologos consideran que hay una celula madre cada tantas y que estan uniformemente distribuidas. Nosotro mostramos que esta idea es una cagada.

Bien, qué linda es la multidisciplina, al final creo que la conclusión se va a convertir en discusión y después conclusión más corta al final. Está quedando buenísimo

## Discussion2

Mathematical modeling has contributed successfully to the study of tumor growth and treatment response. One shortcoming is the risk of parameterization from multiple sources, thus mixing cancer types, experimental conditions, and even spatio-temporal scales<sup>23</sup>. Here, we attempt to overcome this issue by establishing a method of images processing that fits biological data with previously developed mathematical models for the distribution of CSCs in tumorspheres<sup>18</sup>. Taking this into consideration, we chose tumorspheres as a simplified 3D tumor model since they are cellular structures that are generated from a variety of tumors from epithelial tissues, such as breast, lung, prostate, or colorectal cancer<sup>24</sup>. Even though the optimal conditions for culturing tumorspheres may differ among tumor types, this experimental system can recapitulate the physicochemical gradients from the spheroid periphery to its core and mimic, to some extent, mechanical properties and cell-cell interactions of avascular tumor mass microregions<sup>24,25</sup>.

As expected from our simulations, CSCs are not randomly distributed in the tumorsphere, they rather form a path and tend to interact with cells from the same type. This might reflect the phenomenon of stem cell competition, briefly, stem cells reside within specific microenvironments (niches) which provide restricted maintenance signals and limited physical space, and consequently, stem cells are constantly competing with their neighbors for niche occupancy<sup>26</sup>. In the non-neutral stem cell competition, a fraction of stem cells gains a fitness advantage or disadvantage over their neighboring stem cells. Thus, the more competitive fraction overtakes the niche, sometimes disrupting it and leading to diseases such as cancer<sup>27</sup>. Indeed, mutations that affect cell fitness either in development or homeostasis can lead to a competitive growth advantage and potentially clonal expansion. Alternatively, non-neutral stem cell competition eliminates "unfit" clones, for instance when aneuploid cells are depleted during development<sup>28</sup>. However, in the case of tumorspheres, it is difficult to assume the type of stem cell competition due to the selection already imposed by the culture conditions. It is possible, however, that the hypoxic cores in the tumorsphere can trigger differential cell responses leading to drug resistance<sup>29</sup> and these CSCs, more competitive, deplete the drug-sensitive cell populations and eventually reach the surface of the spheroid through self-renewal. One example of this has been reported in tumorspheres from the breast cancer cell line BT474, in which upregulation of HER2 expression led to a hypoxia-conditioned breast CSCs population with increased resistance to trastuzumab (Rodriguez et al, 2018). Even though these effects have been observed in large sized, multicellular spheroids with significant oxygen, nutrients and metabolite gradients, the fact that small spheroids (25-50 cells) contain cells more resistant than monolayers to chemotherapeutic agents has been repeatedly observed long ago<sup>30</sup>. This suggests that other factors, possibly mechanical/geometrical restraints imposed by the shape of the tumorsphere affect CSCs location and cell-cell interactions. Therefore, mathematical models of how different subpopulations of cells interact in the context of a multicellular aggregate of spheroidal shape and validating these models with biological data may contribute to develop cost-efficient tools able to predict therapeutic responses. In this regard, image processing is useful to obtain information that is highly relevant for setting up simulations, formulating mathematical models, and fitting their results. Here, we measured both, the total amount of cells in each spheroid and the respective CSC fraction, information that will lead to improving the accuracy of mathematical models such as the ones summarized in<sup>11</sup>. Furthermore, we estimate that the growth rate is around  $r=1.1$  cell/day, which allows us to establish and set the temporal scale of simulations as those in<sup>18</sup>. Of note, this value is in agreement with the one obtained by fitting experimental data with mathematical modeling<sup>12</sup>, and encourages us to extend our simulations to 3D models.

## Conclusion2

The method of image processing presented here is devoted to recognizing the location of the CSCs in microscopy images. Its originality lies in the fact that we are not just looking at where the SOX2 fluorescence is high enough under a subjective criterion. The key is the use of the Gaussian Mixture Model to fit the data and separate, with a statistical criterion, both cell populations. Having done this, the way of depicting the spheroid or what is done with the resulting data will depend on the questions that researchers have in mind. In our case, we presented an example that compares the experimental CSC distribution in tumorspheres with a previous computational model. Beyond the fact that we are pleased to find that *in silico* and *in vitro* experiments give similar outcomes, the method proposed here is straightforwardly applicable to other similar experiments. Indeed, to definitively assess the CSCs distribution in a spheroid, we must be able to put into the microscope non-deformed spheroids and take pictures of several slices of them. Furthermore, according to our computational simulations, the confidence in our result will exponentially increase with the cultured time. Indeed, we are carrying out full 3D simulations of tumorsphere assays. The preliminary results still support the finding that CSCs are heterogeneously distributed inside the spheroid. However, the probability of finding CSCs on the border of the spheroid seems to be significantly enlarged. These simulations plus the help of confocal images of several slices of a non-deformed spheroid, will be the basis for studying further therapies focused on treating cancer stem cells.

## Methods

### Cell Culture and mammospheres assay

Human breast cancer-derived cell line MCF-7 was maintained in DMEM/F12 complete medium (DMEM F12 + FBS 10 + 1 Glutamine) at 37 °C in a humified 5 CO<sub>2</sub> atmosphere. The mammosphere assay was performed according to modifications<sup>21</sup> from the original protocol by Dontu *et al.*<sup>3</sup>. Briefly, 6-well plates were treated with poly(2-hydroxyethyl methacrylate) to prevent cell adhesion. MCF-7 cells were seeded at 3000 cells/ml in complete mammospheres medium (DMEM/F12 + B27 2 + glutamine 1 +20 ng/ml EGF + 20 ng/ml bFGF). Cells were incubated at 37 °C in a humified 5 CO<sub>2</sub> atmosphere for 7-9 days, adding 0.5 ml of medium with growth factors every 48 hs. The resulting mammospheres were collected and attached by cyt centrifugation onto slides for immunofluorescent detection of SOX2.

Lucas  
SOX no va con todo mayuscula?  
La convención dice eso para proteinas, si.  
igual como que ya cada uno hace lo que se le canta

### Immunofluorescence

Immunofluorescent detection of SOX2 in mammospheres was performed as described in Leis *et al.*<sup>21</sup>: slides were fixed in methanol and then washed 3 times in washed solution (PBS-BSA 0.1) for 5 minutes and permeabilized with PBS-BSA 0.1 + 0.3 triton X100 for 30 min at RT. Following permeabilization, the slides were incubated in blocking solution (PBS-BSA 0.1 + Normal Goat Serum) and the corresponding primary antibody (anti-SOX2, cat # PA1-16968, Thermo) ON at 4C. Next, the slides were washed 3 times in washing solution for 5 minutes and incubated with the corresponding secondary antibody (anti-rabbit Alexa Fluor (TM) 555 cat # A31572, Life Technologies).

### Confocal Imaging

Luciano, describí como se tomaron las fotos en los tres canales

Images were acquired Zeiss LSM 880 Laser Scanning Confocal Microscope with a 20X objective. The corresponding lasers and photomultipliers for excitation and detection of Alexa Fluor 555 and DAPI were used. Confocal images were taken every X micrometer (tengo que consultar cada cuánto se tomaron estos stacks, no me acuerdo)

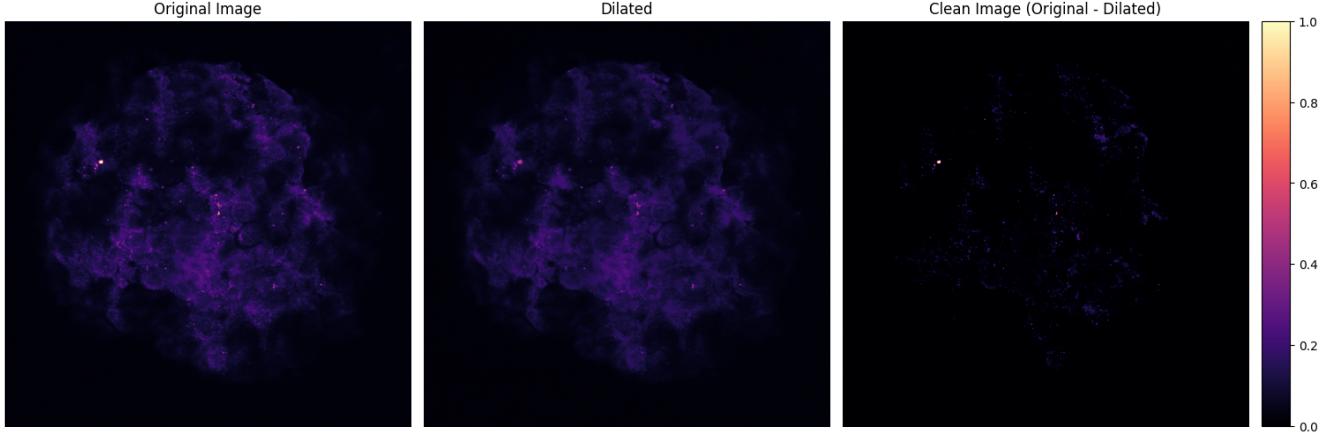
### Image Analysis

In the files obtained by digital imaging of the confocal microscopy, we recognize the number, position, and extent of cells in the spheroids and tag CSCs and DCCs according to the amount of SOX2 fluorescence using a statistical criterion. In the following, we describe the whole process and show the partial results on spheroid number 1, Sph1 hereafter, as an example. We began with the corresponding .czi file that contains data on three channels labeled *optic*, *DAPI* with the fluorescence of the nuclei, and *SOX2* with the SOX2 fluorescence. The data are files containing the position of the pixels in three spatial dimensions (*x*, *y*, *z*) with 2292 pixels representing 0.12 μm for each side in the horizontal *x* – *y* plane, and a thickness of 9 pixels representing 2 μm., in the *z* direction. For each pixel, the corresponding value for each channel is included. We start analyzing each channel separately and then, we merge their information.

### The SOX2 Channel

The technique reveals the spots where the marker was attached to the cells, getting rid of the diffused SOX2 in the bulk. To achieve this, we perform a reconstruction by dilation, a morphological image processing technique that provides a denoised version of the image by, subtracting from it, a blurred version of the original one. Fig. 2 shows the original image in the left

panel, the dilated image at the center, and the resulting cleaned image in the right panel. Specifically, we used a Scikit-image’s implementation of the morphology algorithms. For further information, see [scikit-image’s documentation](#)).



**Figure 2.** Cleaning of the SOX2 channel. From left to right: the original image, a dilated version of it (a blurred version or background noise), and the cleaned image as the difference between them. The original image corresponds to the SOX2 channel of Sph1, slice3. Note that the dilated image allows us to remove a large amount of non-informative pixels.

### **The Nuclei Channel**

For this channel, we use the nuclei to individualize and identify the cells. Specifically, we performed instance segmentation on the nuclei, extracting geometrical features such as the area and the coordinates of the center of the segmented objects by the following the next steps:

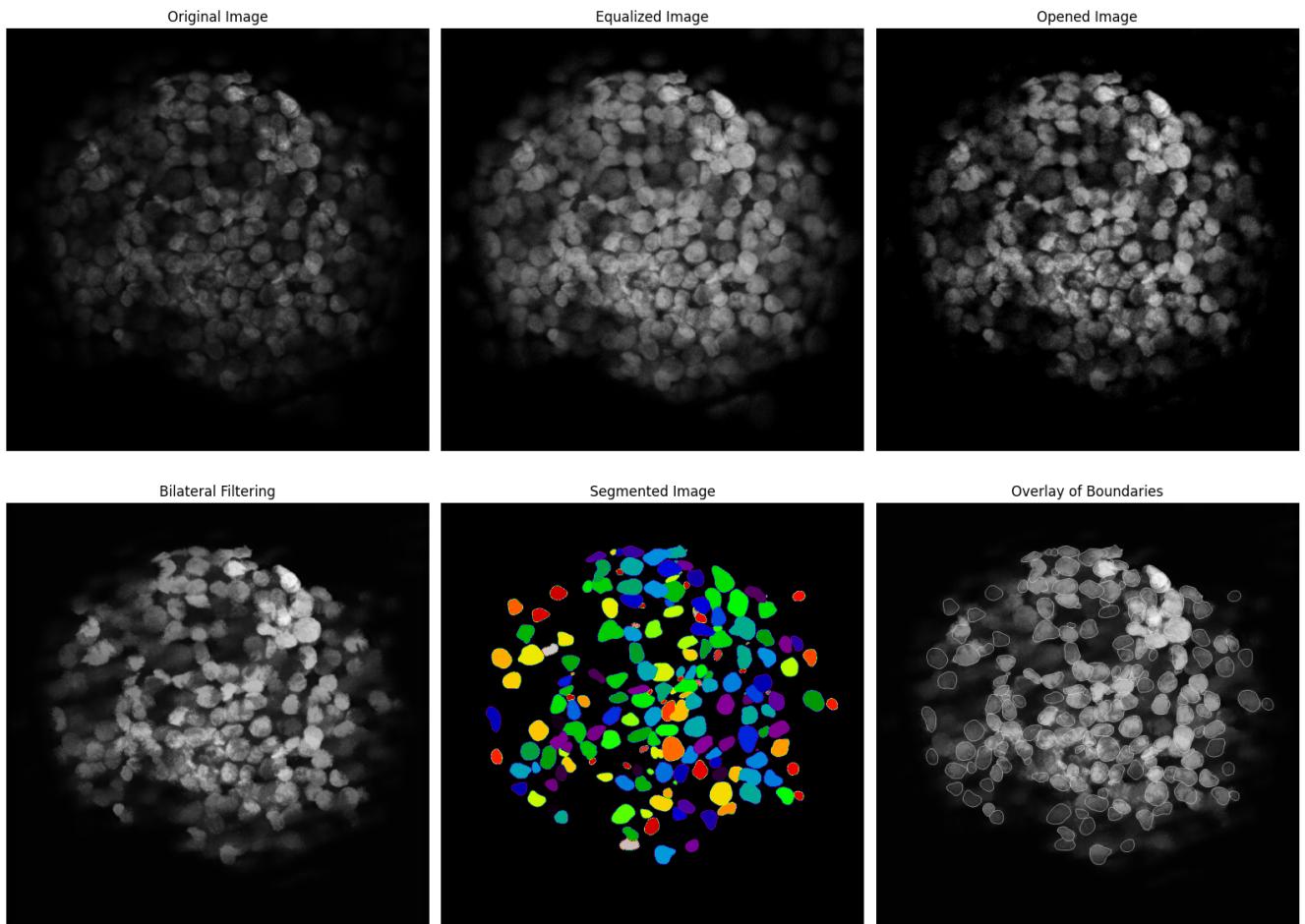
1. Contrast Limited Adaptive Histogram Equalization (CLAHE): Used for local contrast enhancement (see [documentation](#)) makes the edge recognition easier for the segmentation algorithm.
2. Morphological processing: To further improve the edge recognition, we performed a morphological opening (erosion followed by dilation), followed by an area closing (similar to a morphological closing, dilation followed by erosion, but using a deformable rather than a fixed footprint). The opening helps separate objects that may be in contact. The area closing removes small dark structures to avoid single nuclei being segmented as many objects due to dark spots within them (see [documentation](#)).
3. Bilateral Denoising: An edge-preserving bilateral filter denoises the image, averaging pixels based on their spatial closeness and radiometric similarity. This filter works both to reduce the noise introduced by the morphology operations and by out-of-focus objects (see [documentation](#)).
4. Instance Segmentation: For identifying the nuclei, we used the `2D_versatile_fluo` model (see [documentation](#)). This is a Stardist<sup>31</sup> convolutional neural network with a U-Net architecture, trained on fluorescence microscopy images similar to the ones of our experiment.

We chose this model instead of instance segmentation with bounding boxes because, in this way, we do not need a subsequent shape refinement. Furthermore, semantic (per-pixel) cell segmentation requires a subsequent pixel grouping that can result in segmentation errors such as falsely merging bordering cells. Since the images portray situations of very crowded cells, such errors would be very likely to happen in our case. Star-convex polygons provide a much better shape representation that overcomes these difficulties.

The complete process on DAPI channel is depicted in Fig. 3. The original image is in the upper left corner. The lower right image is an enhanced representation highlighting the reconstructed cell boundaries.

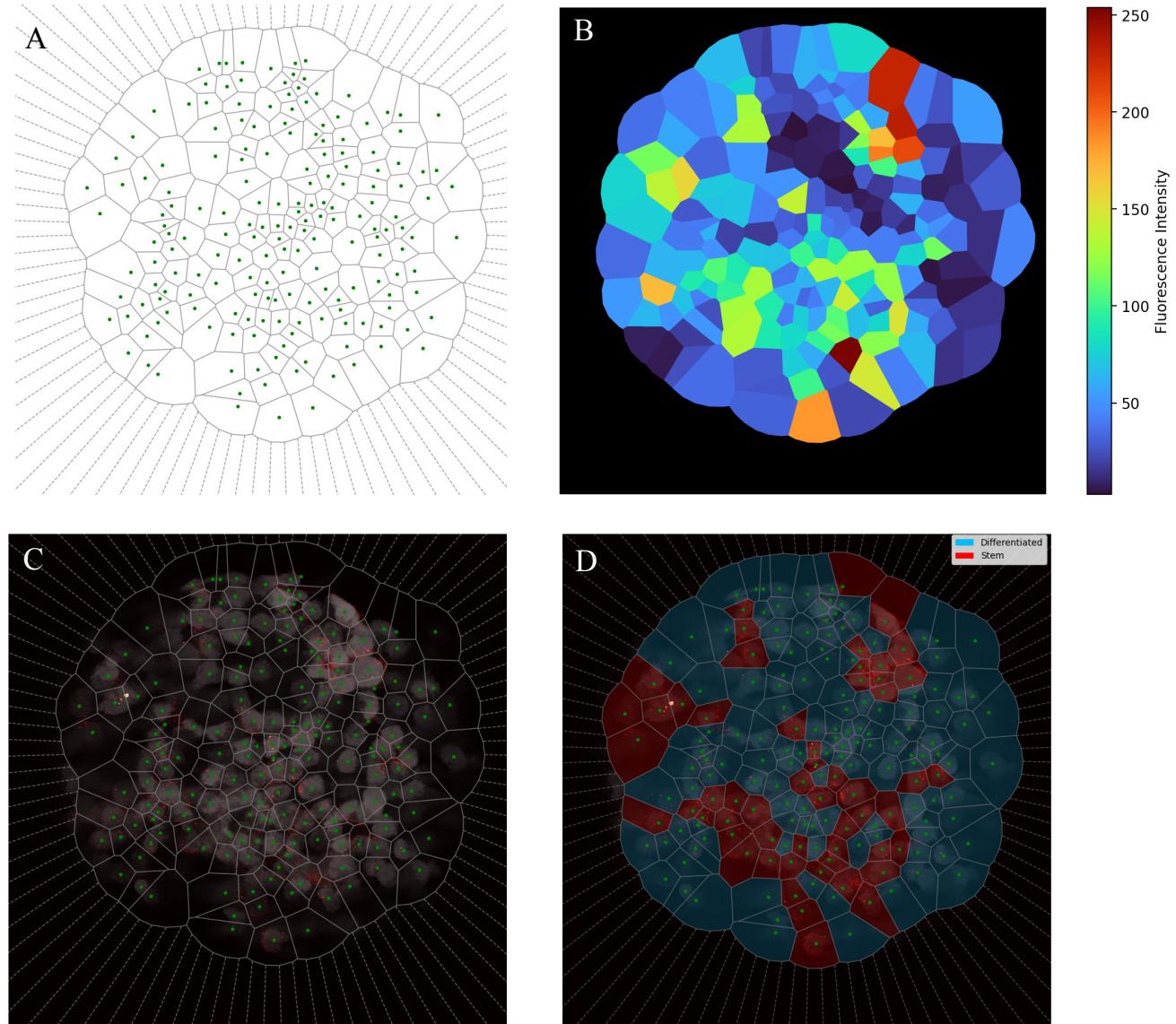
### **Tessellation for cell region identification**

Since the SOX2 marker is not necessarily bound to the nucleus of the cell, but rather to its cytoplasm, we need a way of assigning each point in space (i.e., each pixel) to a single cell to further associate the fluorescence of the marker with a given cell. To do this, we approximated the division of the space associated with each cell by a Voronoi tessellation. We used the



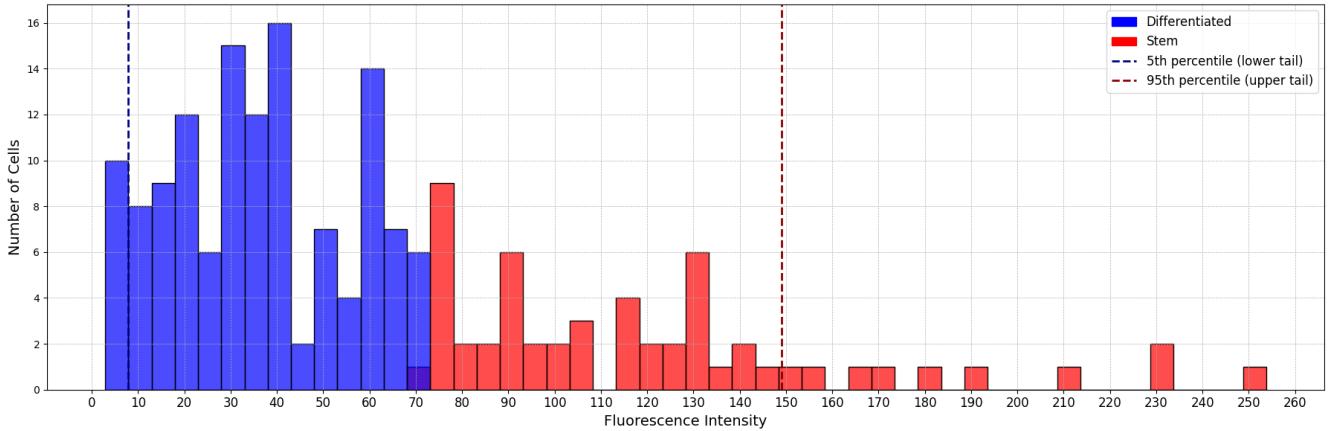
**Figure 3.** Image processing of the nuclei fluorescence channel (DAPI). From left to right and top to bottom, the different steps in the processing of an image are followed. In the bottom right corner are shown the borders of the identified structures over the original image. The darker, unrecognized nuclei correspond to cells that are outside, in this case behind, the focal plane.

centers of the segmented objects filtering out the ones corresponding to cells that do not belong to the spheroid when needed. The result is shown in Fig. 4A. It is worth mentioning that we used artificial points, uniformly distributed in a circumference enclosing the spheroid, to limit the area of the cells in its border.



**Figure 4.** From tessellation to CSC recognition. **A** Voronoi tessellation. The centroids of the cells, green dots, were used to estimate the cell area. **B** Regions colored according to SOX2 fluorescence intensity. For each region of the tessellation, the sum of its SOX2 fluorescence intensity is computed. **C** Tessellation overlaid with nuclei (gray scale) and SOX2 (red palette) channels. The boundaries between Voronoi regions are plotted with gray lines. **D** Tessellation regions are colored in red for Cancer Stem Cells and blue for Differentiated Cancer Cells, according to their SOX2 fluorescence content. If the fluorescence surpasses the threshold given by the GMM, the cell is considered a stem one.

We thus associated to each cell, the sum of the fluorescence intensity of each pixel of the SOX2 channel contained inside its corresponding Voronoi region. The result of this association is shown in Fig. 4 as a color scale depicting the amount of SOX2



**Figure 5.** Histogram of SOX2 fluorescence intensities, colored by the GMM clustering. Intensity values were clustered into differentiated and stem cells by fitting a GMM. Regions with intensities falling in the 5th and 95th percentile were automatically considered as differentiated and stem respectively and were not taken into account when fitting the GMM.

in each Voronoi region. To better visualize the tessellation, we overlayed it on the nuclei fluorescence and the cleaned SOX2 channel in Fig. 4C.

#### Stemness threshold in SOX2 fluorescence intensity

To identify the CSCs, is needed to define a threshold in the intensity of SOX2 fluorescence that divides non-stem from stem cells. We implemented a clustering algorithm to group the intensities of the cells and an “elbow” plot using k-means clustering. Its result, 2, confirmed the suitability of our model with two cell phenotypes. Thus, we considered a fixed number of two groups were data points were assigned according to a Gaussian Mixture Model (GMM) fitted to the data. In short, we assumed that the points were generated by two normal distributions, fitting their means and standard deviations using the maximum likelihood criterion. Additional details can be found in the [documentation](#).

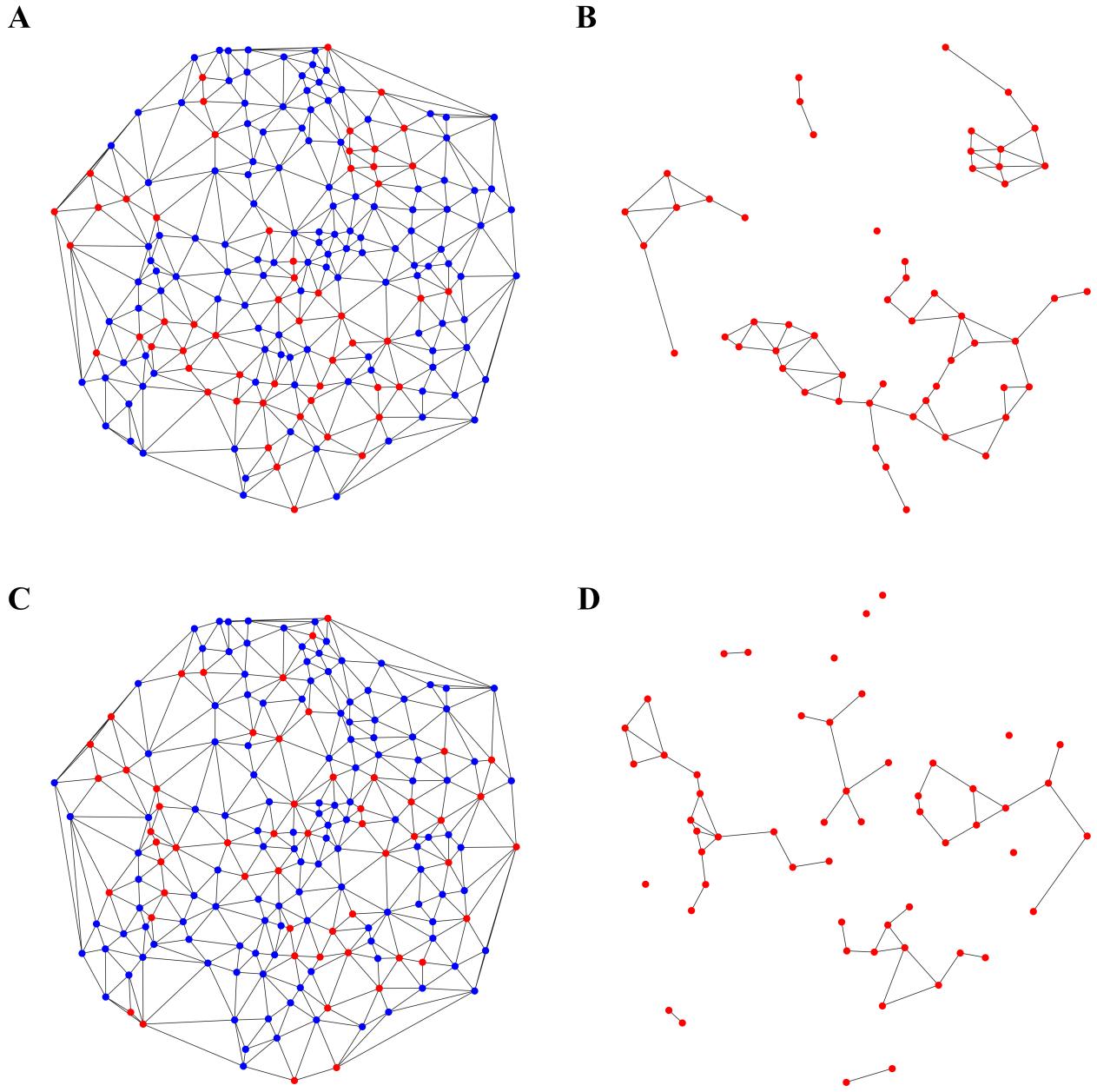
Since the fitting of the distributions is sensitive to extreme values, the outliers, below 5th and above 95th percentiles, were directly assigned to their corresponding category, differentiated for the lowest and stem for the highest values, and not were taken into account when fitting the GMM. An example of the histogram of SOX2 fluorescence intensities for Sph1,slice 3 is shown in Fig. 5. The threshold value  $V = 71.4$  implies that from a total of  $N = 183$  cells present in the spheroid,  $S_V = 55$  are stem cells, a fraction of  $f_V \simeq 0.3$  of the total. For every case, we checked that the random seed of the clustering algorithm had not modified the threshold significantly. Usually, there are a couple of values to which the threshold converges. Checking for robustness of the clustering means that these values are close to each other, regardless of the random state used and the clustering algorithm. If these values vary a lot or are not close to each other, the clustering becomes unreliable. For the case of our example, both GMM and K-means yield a threshold of 71.4, for a large number of random seeds.

A more intuitive way of visualizing this result is taking Fig. 4C and coloring each region according to its clustering. This is shown in Fig. 4D where, superimposed to the confocal image, the regions corresponding to stem cells are colored in red, and the regions corresponding to differentiated cells are colored in blue.

The final result of the image processing is shown in Fig. 1. There just the Voronoi regions are depicted as being colored according to their category. The image depicted in Fig. 1C corresponds to the one in ig. 4D.

#### CSC connectedness and path forming

The location of the CSCs seems to be far from random by visual inspection of Fig. 1, since the paths that they form, break the homogeneity of their distribution. This intuition is confirmed by the graph of a *Delaunay triangulation*, the dual lattice of the Voronoi tessellation. This is nothing else than the graph given by using the centroids of the Voronoi as nodes, and adding a link between two nodes if the corresponding cells in the Voronoi diagram are neighboring cells. For instance, in Fig. 6A we depict the network corresponding to Sph1, slice3, the red and blue dots correspond to CSCs and DCCs respectively. Also we studied the subgraph with just the CSCs shown in Fig. 6B. This graph is obtained by removing all the blue nodes and all the links connected to them. We used these isolated paths, to assess if two CSCs are neighboring by chance. Particularly, we look if the nodes of the network have preferential attachment for nodes of the same type. We calculated the *homophily* of the graph, i.e. how much more likely are CSCs to be connected between them, than to DCCs. This was achieved by measuring, on the graphs, the following parameters:



**Figure 6.** **A** The reconstructed network for Sph1, slice3 with its Cancer Stem Cells plotted as red dots while its Differentiated Cancer Cells plotted as blue dots. Neighbor cells are connected with gray links. **B** The CSCs subgraph corresponding to the network in **A** is obtained by deletion of the DCCs (blue) nodes. The number of stem/connected components is 5. **C** The random graphs are constructed by randomly redistributing the red and blue dots on the same network. **D** The CSCs subgraph corresponding to the randomized network in **C** has a different number of connected components, 13 in this case. In most random cases, this number is larger than in the experimental network.

- *Assortativity coefficient:* The measurement of preferential attachment that is used in network science. It is calculated as the Pearson correlation coefficient of the attribute indicating the type of cell, between pairs of linked nodes. The coefficient equals 1 for perfect assortativity (perfect preferential attachment), 0 for non-assortative networks (i.e., no preferential attachment), and -1 for perfectly disassortative networks (perfect heterophily).
- *Homophily ratio:* Is the fraction of the edges of the graph that link nodes of the same kind. The calculation is direct,

dividing the number of edges connecting cells of the same type by the total number of edges.

- *Number of stem-connected components*: At the subgraph formed by considering only the stem cells of the original graph and the links between them (stem subgraph), is the number of separated clusters. If stem cells form clusters, the number of connected components will be smaller than expected for a random distribution.
- *Degree distribution of the stem subgraph*: In the stem subgraph is the average number of stem cells that are connected to each stem cell. For instance, if the mean degree is higher than expected for randomly distributed stem cells, it means that stem cells are more connected between them than what would be expected from the random case.

A NetworkX's<sup>32</sup> implementation was used for measuring all parameters, except for the homophily ratio. For more details consult the package's [documentation](#).

We also look for the possibility that CSCs were placed at random positions. On the same networks obtained by Delaunay triangulation on the pictures, and leaving the connections untouched, we randomly redistributed the CSC in the graph. An example is shown in Fig. 6C where the red and blue dots are now randomly placed in other nodes, as seen when compared with the panel A. In panel D we depicted the subgraph corresponding to the randomized network of panel C. Note that the number of connected components has increased from 5 in panel B to 13 for this case, a situation that occurs for most randomized networks. We performed this relocation of the whole set of CSCs 10000 times and measured the mentioned parameters joined with their corresponding standard deviations. The results comparing both, experimental and randomized networks, are reported in Table 2. We also performed statistical tests (z-test via `statsmodels` implementation<sup>33</sup>) to decide whether a given value could be obtained from the distribution of the sample, i.e. if the parameter is significantly different from the random case. In every case, we used a significance  $\alpha_t = 0.001$ , but results are robust against even lower values of  $\alpha_t$ .

## References

1. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc.Nat.Ac.Sc.* **100**, 3983–3988 (2003).
2. Kakarala, M. & Wicha, M. Implications of the cancer stem-cell hypothesis for breast cancer prevention and therapy. *J. Clin. Oncol.* **26**, 2813–2820, DOI: [10.1200/JCO.2008.16.3931](https://doi.org/10.1200/JCO.2008.16.3931) (2008).
3. Dontu, G. *et al.* In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270, DOI: [10.1101/gad.1061803](https://doi.org/10.1101/gad.1061803) (2003).
4. Ehmsen, S. *et al.* Increased cholesterol biosynthesis is a key characteristic of breast cancer stem cells influencing patient outcome. *Cell Rep.* **27**, 3927–3938 (2019).
5. Pastrana E, D. F., Silva-Vargas V. Eyes wide open: a critical review of sphere-formation as an assay for stem cells. *Cell Stem Cell* **8**, 486–498, DOI: [10.1016/j.stem.2011.04.007](https://doi.org/10.1016/j.stem.2011.04.007). (2011).
6. Kessel SL, C. L. A high-throughput image cytometry method for the formation, morphometric, and viability analysis of drug-treated mammospheres. *SLAS Discov.* **25**, 723–733, DOI: [10.1177/2472555220922817](https://doi.org/10.1177/2472555220922817) (2020).
7. La Porta, C. A. & Zapperi, S. *The physics of cancer* (Cambridge University Press, 2017).
8. Bull, J. A. & Byrne, H. M. The hallmarks of mathematical oncology. *Proc. IEEE* **110**, 523–540 (2022).
9. Vieira LC, V. D., Costa RS. An overview of mathematical modelling in cancer research. *Fractal Fract.* **7**, 595, DOI: [10.3390/fractfract7080595](https://doi.org/10.3390/fractfract7080595) (2023).
10. Benítez, L., Barberis, L. & Condat, C. A. Modeling tumorspheres reveals cancer stem cell niche building and plasticity. *Phys. A* **533**, 121906, DOI: [10.1016/j.physa.2019.121906](https://doi.org/10.1016/j.physa.2019.121906) (2019).
11. Barberis, L. M., Benitez, L. & Condat, C. Elucidating the Role Played by Cancer Stem Cells in Cancer Growth. *MMSB* **1**, 48–54 (2021).
12. Benítez, L., Barberis, L., Vellón, L. & Condat, C. A. Understanding the influence of substrate when growing tumorspheres. *BMC Cancer* **21**, 276, DOI: [10.1186/s12885-021-07918-1](https://doi.org/10.1186/s12885-021-07918-1) (2021).
13. Fotinós, J., Barberis, L. & Condat, C. Effects of a differentiating therapy on cancer-stem-cell-driven tumors. *J. Theor. Biol.* **572**, 111563, DOI: [10.1016/j.jtbi.2023.111563](https://doi.org/10.1016/j.jtbi.2023.111563) (2023).
14. Condat, C. A. & Menchon, S. A. Ontogenetic growth of multicellular tumor spheroids. *Phys. A* **371**, 76–79, DOI: [10.1016/j.physa.2006.04.082](https://doi.org/10.1016/j.physa.2006.04.082) (2006).
15. Delsanto, P. P., Condat, C. A., Pugno, N., Gliozzi, A. S. & Griffa, M. A multilevel approach to cancer growth modeling. *J. Theor. Biol.* **250**, 16–24, DOI: [10.1016/j.jtbi.2007.09.023](https://doi.org/10.1016/j.jtbi.2007.09.023) (2008).

16. Menchon, S. A. & Condat, C. A. Quiescent cells: A natural way to resist chemotherapy. *Phys. A* **390**, 3354–3361., DOI: [10.1016/j.physa.2011.05.009](https://doi.org/10.1016/j.physa.2011.05.009) (2011).
17. Barberis, L., Pasquale, M. A. & Condat, C. A. Joint fitting reveals hidden interactions in tumor growth. *J. Theor. Biol.* **365C**, 420–432, DOI: [10.1016/j.jtbi.2014.10.038](https://doi.org/10.1016/j.jtbi.2014.10.038) (2015).
18. Barberis, L. Radial percolation reveals that Cancer Stem Cells are trapped in the core of colonies. *Pap. Phys.* **13**, 130002–130002, DOI: [10.4279/pip.130002](https://doi.org/10.4279/pip.130002) (2021).
19. Chen, Y. C. *et al.* High-throughput single-cell derived sphere formation for cancer stem-like cell identification and analysis. *Sci. Reports* **6**, 1–12, DOI: [10.1038/srep27301](https://doi.org/10.1038/srep27301) (2016).
20. Wang, J. *et al.* A novel method to limit breast cancer stem cells in states of quiescence, proliferation or differentiation: Use of gel stress in combination with stem cell growth factors. *Oncol. Lett.* **12**, 1355–1360, DOI: [10.3892/ol.2016.4757](https://doi.org/10.3892/ol.2016.4757) (2016).
21. Leis, O. *et al.* Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene* **31**, 1354–1365 (2012).
22. Marks, M. P. *et al.* Role of hydroxymethylglutharyl-coenzyme a reductase in the induction of stem-like states in breast cancer. *J. Cancer Res. Clin. Oncol.* (2024).
23. Brady, R. & Enderling, H. Mathematical models of cancer: When to predict novel therapies, and when not to. *Bull Math Biol* 3722–3731, DOI: [10.1007/s11538-019-00640-x](https://doi.org/10.1007/s11538-019-00640-x) (2019).
24. Zanoni, M. *et al.* Modeling neoplastic disease with spheroids and organoids. *J Hematol Oncol* DOI: [10.1186/s13045-020-00931-0](https://doi.org/10.1186/s13045-020-00931-0) (2020).
25. Rolver, M. G., Elingaard-Larsen, L. O. & Pedersen, S. F. Assessing cell viability and death in 3d spheroid cultures of cancer cells. *J Vis Exp* DOI: [10.3791/59714](https://doi.org/10.3791/59714) (2019).
26. Stine, R. R. & Matunis, E. L. Stem cell competition: finding balance in the niche. *Trends Cell Biol* 357–364, DOI: [10.1016/j.tcb.2013.03.001](https://doi.org/10.1016/j.tcb.2013.03.001) (2013).
27. Johnston, L. A. Competitive interactions between cells: death, growth, and geography. *Science* 1679–1682, DOI: [10.1126/science.1163862](https://doi.org/10.1126/science.1163862) (2009).
28. Derkx, L. L. M. & van Boxtel, R. Stem cell mutations, associated cancer risk, and consequences for regenerative medicine. *Cell Stem Cell* 1421–1433, DOI: [10.1016/j.stem.2023.09.008](https://doi.org/10.1016/j.stem.2023.09.008) (2023).
29. Fisher, M. F. & Rao, S. S. Three-dimensional culture models to study drug resistance in breast cancer. *Biotechnol Bioeng* 2262–2278, DOI: [10.1002/bit.27356](https://doi.org/10.1002/bit.27356) (2020).
30. Olive, P. L. & Durand, R. E. Drug and radiation resistance in spheroids: cell contact and kinetics. *Cancer Metastasis Rev* 121–138, DOI: [10.1007/BF00689632](https://doi.org/10.1007/BF00689632) (1994).
31. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II* 11, 265–273 (Springer, 2018).
32. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T. & Millman, J. (eds.) *Proceedings of the 7th Python in Science Conference*, 11 – 15 (Pasadena, CA USA, 2008).
33. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference* (2010).

## Acknowledgements

The authors thank Dr. C. A. Condat for insightful discussions and comments. The confocal imaging was performed by .

Luciano: poner  
CPA del MI-  
CROSCOPIO

## Funding

The theoretical work was supported by SECyT-UNC (Project 113/17) and CONICET (PIP 11220110100794), Argentina. Experiments were supported by .

FINANCIAMIEnto  
LUCIANO

## Author contributions statement

JF developed and implemented the image processing pipeline, LB conceived the general idea, LV carried out the experiment. All authors analyzed the results and wrote the manuscript.

## **Additional information**

**Competing interests:** The authors declare no competing interests.

**Accession codes:** All code and data are accessible through the GitHub repository:  [experimental\\_image\\_analysis](#)