LUDWIG-MAXIMILIANS-UNIVERSITÄT
TECHNISCHE UNIVERSITÄT MÜNCHEN

Helmholtz Zentrum München

Masterarbeit
in Bioinformatik

Bioimage informatic analysis of the spatial impact on cell state transitions

Valerio Lupperger

Aufgabensteller:	Prof. Fabian Theis
Betreuer:	Dr. Carsten Marr, Felix Buggenthin
Abgabedatum:	15.05.2015

Ich versichere, dass ich diese Masterarbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

15.05.2015 ______________________

Valerio Lupperger

Zusammenfassung

Neuronale Stammzellen besitzen die Fähigkeit sich durch Teilung zu erneuern und in verschieden Zelltypen (Neuronen und Gliazellen) zu differenzieren. Durch diese Eigenschaften spielen sie eine wichtige Rolle in der Pathogenese und Behandlung diverser neuronaler Erkrankungen wie Multiple Sklerose oder Demenz. Dabei ist der Differenzierungsprozess dieser Zellen noch nicht vollständig verstanden. Ein Forschungsgebiet der Neurowissenschaft befasst sich daher mit dem Teilungsverhalten von neuronalen Zellen. Beeinflussen Zellen die sich teilen andere Zellen bezüglich deren Teilungswahrscheinlichkeit und wirken sich damit auf den Regenerations- bzw. Differenzierungsprozess aus? Bilder mit gefärbten Stammzellen und sich teilende Zellen können helfen räumliche Abhängigkeiten aufdecken. Die Aufgabe einzelne Zellen zu finden und deren räumiche Abhängigkeiten statistisch auszuwerten macht es notwendig rechenintensive Methoden für Quantifizierung und Analyse einzusetzen.

In dieser Arbeit präsentieren wir eine halb-automatische Pipeline um dreidimensionale Bilder von Zebrafisch Gehirnhälften zu bearbeiten und Einzelzellen auf den Gehirnhälften zu quantifizieren. Die Bilder stammen aus Experimenten des Zebrafish Neurogenetics Departments (Helmholtz Zentrum Mün-chen). Die Pipeline beinhaltet Verringerung von Bildverzerrungen und das Identifizieren von dreidimensionalen Zentroiden für jede Einzelzelle. Diese Zentroide können dazu benutzt werden um die Oberfläche der Gehirnhälfte abzuschätzen, auf der sich die Stammzellen befinden. Die Abschätzung ist nötig, da die Oberfläche uneben ist und die Gehirnhälfte eine gebogene Form hat. Wenn wir die Oberfläche mit den Zellpositionen kombinieren, können wir Zellabstände unter Berücksichtigung der Oberflächenstruktur berechnen. Wir benutzen die Abstände zwischen sich teilenden Zellen um räumliche Teilungsmuster zu identifizieren, die wiederum Aufschluss über ein zugrundeliegendes biologisches Modell geben können, d.h. inwiefern Zellenteilung benachbarte Zellen hinsichtlich Teilungswahrscheinlichkeit beeinflussen.

Mittels verschiedener statistischer Methoden können wir zeigen, dass Zellteilungen von neuronalen Vorläuferzellen höchstwahrscheinlich einem sich anziehenden Muster folgen, d.h. sich teilende Zellen befinden sich bevorzugt in der Nähe anderer sich teilender Zellen. Für sich teilende Stammzellen

hingegen können wir keine signifikante Aussage treffen, jedoch ist ein Trend zu anziehenden Mustern erkennbar. Um den biologischen Mechanismus hinter diesem Muster zu verstehen schlagen wir Folgeexperimente mit mehreren Färbungen vor.

Abstract

Neural stem cells are able to self renew via symmetric division and can give rise to differentiated cell types (neurons and glial cells). Due to these properties they play an important role in parthenogenesis and treatment of several neural diseases as multiple sclerosis or dementia. However, the differentiation process is not fully understood. A research field in neuroscience is thus the division behavior of neural cells. Do dividing cells impact on other cells in terms of division probability and therefore influence the regeneration or differentiation process? Differently stained stem cells and dividing cells of zebrafish hemispheres can reveal spatial dependencies of divisions. The complex task to identify single cell locations and statistically evaluate them makes it necessary to apply computational methods for quantification and analysis.

In this thesis we present a semi-automatic pipeline to process three-dimensional images of zebrafish hemispheres and quantify single cells that lie on the hemisphere based on experiments from the Zebrafish Neurogenetics Department (Helmholtz Zentrum München). The pipeline covers noise reduction in the image data and identifying three-dimensional centroids for every single cell. These identified centroids are used to estimate the hemisphere surface, on which the stem cells are located. This estimation is necessary since the surface is rough and the hemisphere has a bent three-dimensional structure. Concatenating the surface with cell locations we are able to calculate distances between cells on the surface. Distances between dividing cells are used to identify spatial division patterns, which then in turn can give rise to an underlying biological model how dividing cells influence neighboring cells in terms of division probability.

We can show that dividing progenitor cells follow most likely an attractive pattern i.e. dividing cells are located preferably nearby other divisions. For dividing stem cells in contrast we cannot significantly identify a spatial pattern, but they tend to follow an attractive pattern as well. To understand the underlying biological mechanism of the pattern we propose follow-up experiments using multiple stainings.

Contents

List of Figures

List of Tables

1 Introduction

Many common neurological disorders, such as stroke, Parkinson's disease and multiple sclerosis, are caused by a loss of neurons and glial cells [1], two cell types that are important for signal transduction and myelin formation in the brain [2, 3]. It is thus important to understand how populations of neurons can be maintained. One way to counter the loss of neurons is the targeted differentiation of neural stem cells (NSC) into neurons, which has been successfully performed in culture [4].
NSCs are stem cells in the nervous system that can self-renew (proliferate) and give rise to progenitor cells, which differentiate into neurons and glial cells (see Figure 1.1). In 1989 Sally Temple [5] described neural stem cells in the subventricular zone of the mouse brain and in 1992 they were isolated for the first time [6]. NSCs are predominantly in a quiescence state and need to be activated to divide and/or differentiate [7]. Commonly it is assumed that differentiated cells are generated via asymmetric cell divisions, where the mother cell divides into a stem cell and a differentiated cell [8].

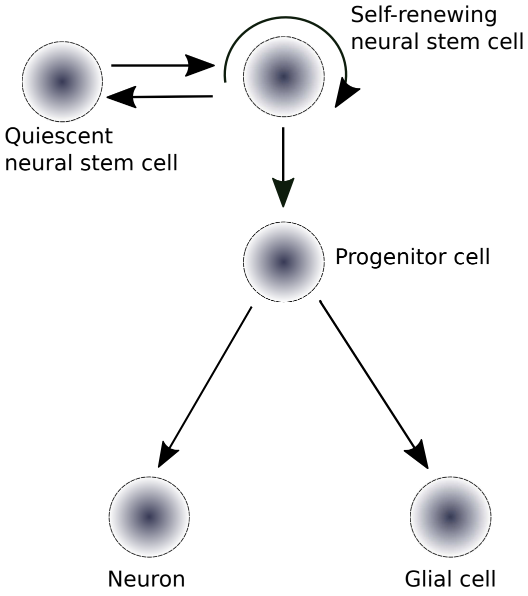

Figure 1.1: Process of neural stem cell (NSC) self-renewal and differentiation. Once they are activated from quiescence NSCs are able to retain the stem cell population via self-renewing, while they can give rise to progenitor cells, which can differentiate to neurons or glial cells.

The ability to differentiate is essential for maintaining the pool of mature, functional cells, and generating diverse cell types during development. The differentiation of NSCs to neurons (neurogenesis) or glial cells (gliogenesis) is a broad research field as the two mechanisms are complex and detailed knowledge is important for further clinical applications [9, 10].

In order to examine NSC differentiation and division patterns, brains of several model organisms like zebrasfish, mouse, C. elegans or drosophila have been imaged [11, 12, 13, 14]. The huge advantage of the zebrafish (*Danio rerio*) is its optical translucence, which allows imaging of the interior without slicing it [15]. As the brain is not flat three-dimensional (3D) imaging can reveal additional spatial information and thus increase the understanding of the imaged cells and structure. The additional spatial information received from 3D imaging allows e.g. to reconstruct lineages during zebrafish development [16] to monitor neural activity in the brain [17], or to analyze the spatial positions of neural cells on the brain hemispheres [18].

Aim of this thesis

Chapouton et al. addressed the question whether NSCs that are undergoing a division influence neighboring cells in terms of division probability [18]. Based on several local snapshots of few cells in the zebrafish brain the authors proposed a repulsive division model, in which cells near a previous division are less likely to divide. We therefore set out to evaluate the proposed model against other hypotheses by quantitatively examining spatial influences for dividing cells of brain hemispheres.

The data used in this thesis was generated at the Zebrafish Neurogenetics Department (Helmholtz Zentrum München) by P. Chapouton. The dataset comprised three zebrafish hemispheres with a GFP staining identifying stem cells. Additionally, dividing cells were labeled by adding EdU, a thymidine analogue that is incorporated in the DNA during the replication phase. For our computational analysis, we developed an analysis pipeline consisting of two main steps:

- In order to identify single cells in the 3D image stacks of zebrafish hemispheres an bioimage informatics pipeline is required. The implementation should cover noise reduction in the image data and 3D centroid identification for every single cell. The NSC centroids can then be used to estimate the locally uneven and bent hemisphere surface. The identified centroids and the hemisphere surface approximation allow proper calculation of cell to cell distances on the hemisphere.

- Using the distances between dividing cells, we will apply different statistical methods for spatial analysis to test for three hypotheses: spatially random distributed divisions, attractively distributed divisions and repulsively distributed divisions.

2 Experimental data

2.1 Cell staining

Adult zebrafish (4-5 month, Figure 2.1BA) were anesthetized in 0.002% tricaine and fixated for the staining process one day before the image acquisition. To distinguish between different cell types, stem cells and dividing cells are stained differently. For this purpose the transgenic *gfap:GFP* zebrafish strain was used, where the green fluorescent protein (GFP) is tagged to a *gfap* enhancer element and inserted into the genome randomly [19]. Consequently GFP is expressed according to the endogenous *gfap*, which is only expressed in NSCs and is therefore able to mark stem cells (see Figure Figure 2.1BB). To identify dividing cells thymidine analogue 5-ethynyl-2-deoxyuridine (EdU) was used [20]. EdU is spuriously incorporated in the DNA by replacing thymidine during DNA replication phase (S-Phase). Hence even cells that did not divide at the time point of image acquisition could be marked EdU positive. But they will divide in near future, since they were already at least in S-Phase.

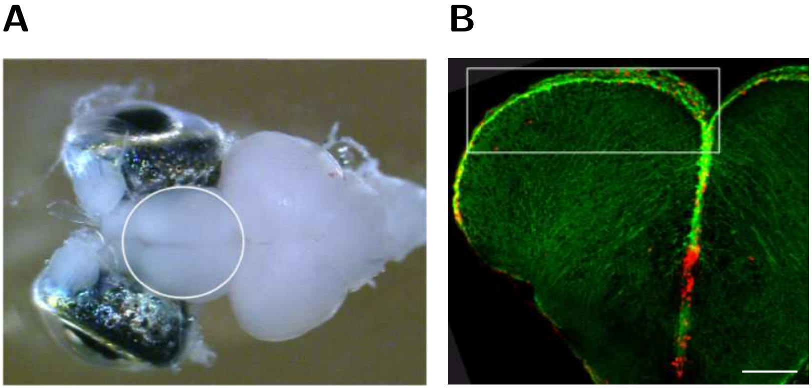

Figure 2.1: NSCs in the zebrafish brain. **(A)** Superior view of a zebrafish with the brain located directly between its eyes (white circle). **(B)** Sagittal view of the zebrafish brain. The GFP stained stem cells are located completely on top of the hemisphere. Adapted from Chapouton et al. [18]. Scale bar: 100 μm

2.2 Image acquisition

24h after adding EdU (see Figure 2.2) the zebrafish were anesthetized with 0.4% tricaine, sacrificed on ice and decapitated. For image acquisition two confocal laser scanning microscopes (Olympus FV1200, Leica sp5) using 20X glycerol-immersion and water-immersion objectives were used, respectively. The chosen resolution suffices to identify single cells, while still one hemisphere of the brain is covered nearly completely by the microscopes field of view (i.e. one image). To account for the bending of the brain hemispheres, a 3D image stack was acquired. Therefor several images were taken in 2D in small distances in z-direction (see Table 1) ending up with a stack of 2D images, which can be combined to a 3D representation of the hemisphere. A full 3D stack comprised two color channels, the first channel shows labeled stem cells (Figure 2.3A), located on the surface of the hemisphere (see Figure 2.3A). The second channel shows cells, which were at least in S-Phase or had already divided between EdU addition and imaging (Figure 2.3B).

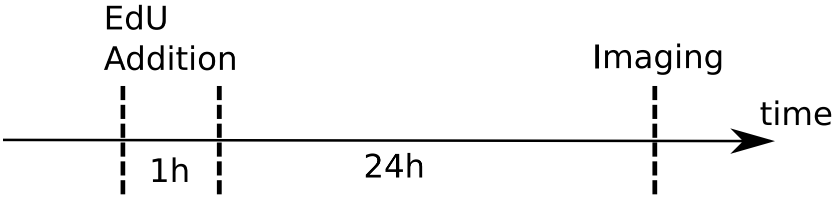

Figure 2.2: 24h after adding EdU the zebrafish is imaged.

Table 1: Experiments used for analysis. We used both hemispheres of one 4 month old fish and one hemisphere of a second 5 month old. The resolution is given in pixel in x and y directions and the number of images in the stack (s). Every data set was named according to ZFx[LR] where x was the number of the zebra fish and L/R identified the left/right hemisphere.

Name	Resolution(x,y,z)	Voxelsize in μm	Animal age in month
ZF1L	1024px x 1024px x 158s	0.37 x 0.37 x 0.71	5
ZF2L	1600px x 1600px x 45s	0.39 x 0.39 x 2.00	4
ZF2R	1600px x 1600px x 46s	0.39 x 0.39 x 2.00	4

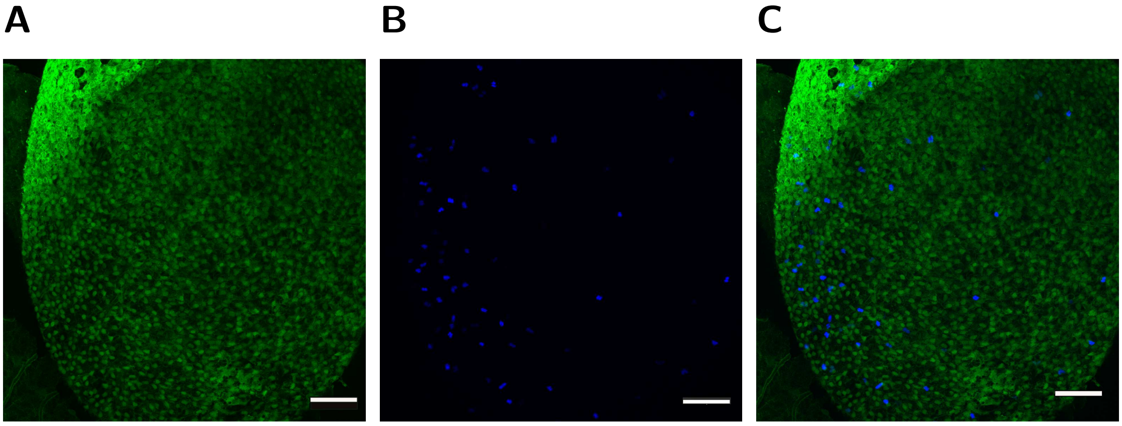

Figure 2.3: Overview of GFP and EdU staining in experiment ZF1L. **(A)** The GFP labeled stem cells on the hemisphere surface. **(B)** The EdU labeled cells where the marker was incorporated in the S-phase of the cell circle. **(C)** Combining both channels we could identify dividing stem cells observing both markers in one cell. If only the EdU was visible the cells were dividing progenitors. Non dividing stem cell gave rise to a GFP-only signal. Scale bars: 50 μm

3 Methods

3.1 Data preprocessing

3.1.1 Deconvolution

During the acquisition process the image is blurred by the light spreading from one pixel to its neighbors. Deconvolution of an image can improve its quality by correcting for the noise incorporated by light spreading.
In general the aim of deconvolution is to find the solution of a convolution equation of the form:

$$fg = h \tag{3.1}$$

where h is the observed image and f is the original image convolved via a certain function g. Many different methods are available to deconvolve an image [21]. The most common methods are blind deconvolution and several point spread function (PSF) based ones. In general a PSF can be used to describe the light spreading pattern on an image. There are several possibilities

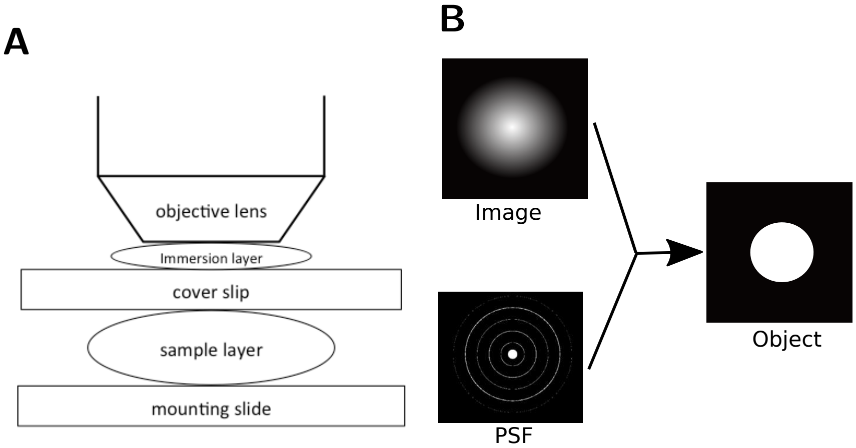

Figure 3.1: Acquisition parameters can be used to obtain a point spread function (PSF) to deblur the images taken by the respective microscope. **(A)** Schematic setting of a microscope which images the sample through an immersion layer and acover slip. The information about the microscope setting can be used for PSF generation. **(B)** During deconvolution a blurred image is combined with a PSF to restore the original objects.

to obtain a PSF, like measuring the light spreading of a point-sized bead or estimating the acquisition parameters via blind deconvolution [21]. Since acquisition parameters were known for our setup, we determined a microscope specific PSF, which is influenced by microscopic parameters like the immersion or numerical aperture (see Figure 3.1A). To restore the unblurred picture the noisy image has to be deconvolved with the PSF as shown in Figure 3.1B.

In order to determine the most appropriate deconvolution method on our data we applied 3D segmentation on the resulting images (after deconvolution). We compared the observed cell numbers to manually counted cells in three regions of the image stack (see Figure 3.2A). The cells were counted by a biological expert on the 2D maximum intensity z-projection (MIZP). The Richardson-Lucy (RL) method [22, 23] outperformed all other methods on our data (see Figure 3.2B).

RL is a PSF based method and to obtain the PSF we applied an analytic technique called the Gibson and Lanni Optical Model, which describes the scalar-based diffraction that occurs in the microscope [24]. It accounts for the immersion, the cover-slip and the sample layers. This model defines

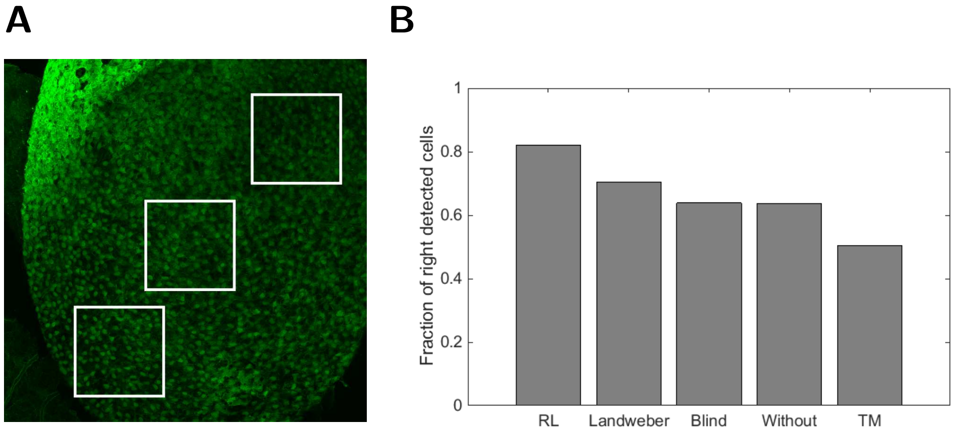

Figure 3.2: Evaluation of deconvolution methods. **(A)** The selected regions in experiment ZF1L (white rectangles) were used to evaluate different deconvolution methods. Cells that could be identified on the 2D projection were counted by an expert and were compared to identified cells after applying the particular deconvolution method. **(B)** The comparison of the deconvolution methods yields the Richardson-Lucy (RL) method as best approach followed by Landweber and Blind deconvoulution. The Tikhonov-Miller (TM) approach performed even worse than without deconvolution, making this method inappropriate for our data.

a PSF (p) by the following equation:

$$p = \left| \frac{C}{z_d} \int_0^1 J_0 \left[kNA\rho \frac{(x^2+y^2)^{0.5}}{z} \right] e^{jW(\rho)} \rho \, \mathrm{d}\rho \right|^2 \tag{3.2}$$

with

$$W(\rho) = kn_s t_s \left[1 - \left(\frac{\rho}{n_s} \right)^{0.5} \right].$$

We use the refraction index of neuronal cells ($n_s = 1.377$) as the refraction index of the sample layer [25]. See Table 2 for further parameter information

.

The resulting PSF can then be used in the Richardson-Lucy approach, which uses an iterative process to obtain the most likely pixel values (u) at every position j given v_i as the observed pixel values at position i and p_{ij} being the PSF (the fraction of light coming from true location j that is observed at position i):

$$u_j^{(t+1)} = u_j^{(t)} \sum_i \frac{d_i}{c_i} p_{ij} \tag{3.3}$$

Table 2: Parameters used to create a microscope and image specific PSF.

NA	Numerical Aperture of the objective lens.
n_s	Refractive index of the sample layer.
t_s	Axial location of the point source within the sample layer.
λ	Wavelength of the light emitted by the point source.
ρ	Normalized radius in the back focal plane.
k	Wavenumber in vacuum of the emitted light, $k = 2\pi/\lambda$.
x, y	Lateral position for evaluating the PSF at the detector plane.
zd	Axial distance between the detector and the tube lens.
W	The Gibson and Lanni phase aberration.
C	A normalizing constant.
J_0	Bessel function of the first kind.

with

$$c_j = \sum_j p_{ij} u_j^{(t)}.$$

3.1.2 Laplace of Gaussian filter

The Laplace of Gaussian (LoG) is a 2D isotropic measure of the 2nd spatial derivative of an image. The LoG of an image highlights regions of rapid intensity change and is therefore often used for edge detection. Applying the LoG on the MIZP we are able to detect possible cell edges (see equation (3.4)). First the image is convolved using a Gaussian kernel at a scale d (d is linked to the diameter of one cell) resulting in a scale space representation (SSR). Afterwards the second derivative of the SSR is calculated to reveal regions which change fast and hence show significant edges, which in turn surround blobs that can be possible cell candidates.

$$\mathrm{LoG}(x, y; d) = -\frac{1}{\pi d^4}\left[1 - \frac{x^2 + y^2}{2d^2}\right] e^{\frac{x^2+y^2}{2d^2}} \tag{3.4}$$

with x and y being coordinates of the image.

3.2 Local maximum/minimum detection

The local maximum/minimum search detects extreme regions in given data. We can use the method to detect cell-sized blobs by applying it to the resulting LoG representation of the image:

$$(\hat{x}, \hat{y}; \hat{t}) = \mathrm{argmaxminlocal}_{(x,y;t)}(\mathrm{LoG}(x, y; t)). \tag{3.5}$$

3.2.1 3D segmentation

3D segmentation is used to detect objects on 3D images. We perform 3D segmentation to identify single cells on the 3D image stack. The method uses a certain pixel intensity threshold and all pixels that have an intensity below this threshold become background pixels and all pixels having a higher or same intensity foreground. Foreground pixels are then connected to pixel clusters by starting at a random pixel and moving in all directions (x, y and z). Every directly reachable foreground pixel is added to the cluster. The clustering continues until all foreground pixel are inside clusters. Two further thresholds, one lower and one upper, identify clusters as possible cells by excluding the clusters that have a lesser amount of pixels than the lower bound or more than the upper bound.

3.3 Geometric concepts

3.3.1 Distance between cells on hemisphere surface

Euclidean distances can be used to calculated distances in 3D but as the stem cells are located completely on the hemisphere surface we need a method to calculate distances with respect to the surface. The fitted surface polynomial (see Section 4.1.3) can be used to calculate cell distances on it.

The distance between two cells on a surface is calculated via a recursive procedure explained for a 2D example in Figure 3.3. First the midpoint between the two cells (a) is calculated in euclidean space and projected to the surface (b). This procedure is repeated for the two original points with the new one respectively (d). The number of points in between depends on the initial euclidean distance of the two cells. The number of points is calculated via dividing the euclidean distance by 4μm which is about one cell diameter. The final distance is then calculated by summing up all euclidean distances between neighboring points.

3.3.2 Delaunay triangulation

The Delaunay triangulation for a set of points is a triangulation such that no point in the set is inside the circumcircle of any triangle in the triangulation (Delaunay condition) [26]. The result is for example used in computer

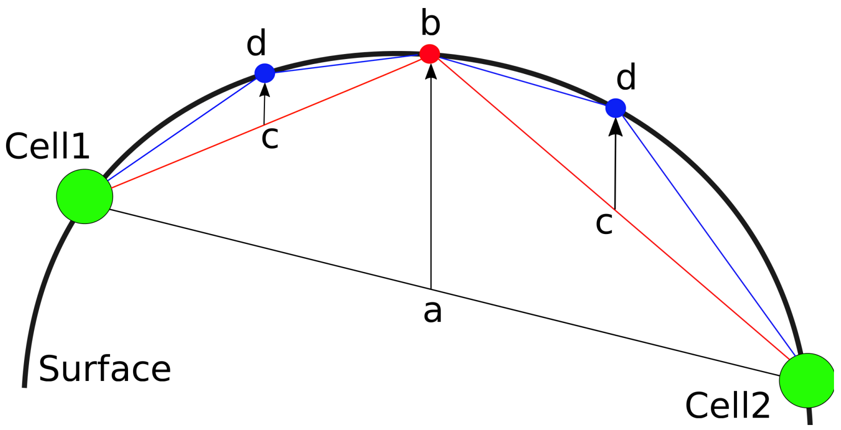

Figure 3.3: Calculating distances distances between two cells on surface (thick curve). The center of the direct link between the two cells (a) is projected to the surface (b). The resulting point is then linked to the initial cells. Afterwards each center (c) is projected to the surface (d) and linked to the end point of the previous link (b).

graphics to obtain high quality triangular meshes. We can use the method to derive the area of the hemisphere surface. We hence calculated the Delaunay triangulation of all cells. In the end we sum up all triangle areas to approximate the overall area.
The incremental approach to reach a triangulation adds all points step by step. The respective triangle, an added point lies in, is split up and the new point is connected to the triangle vertices. This can harm the Delaunay condition and it has to be checked if edges need to be "flipped" (see Figure 3.4). The new edges can in turn lead to a harming of the Delaunay condition in the neighboring triangles and consequently it could be that we need to propagate through the whole triangulation.

3.3.3 Lasso regularization

Lasso regularization can be used for model selection, in particular to prevent overfitting by penalizing models using extreme parameter values [27]:

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(z_i - a_i^T\beta\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right) \tag{3.6}$$

with N being the number of points, z_i the response values of the response vector z, a_i the combination of x- and y-values for different orders, λ the regularization parameter and β the coefficient vector.

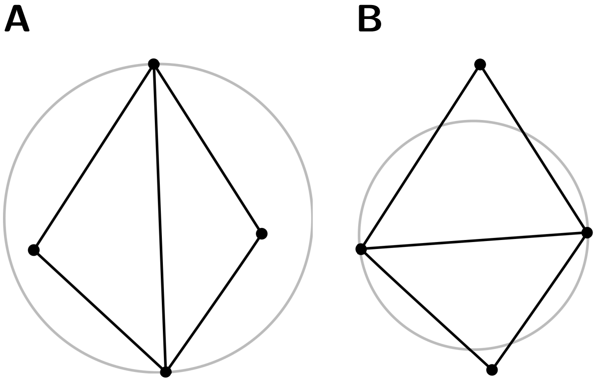

Figure 3.4: Edges harming the Delaunay condition need to be flipped. **(A)** The edge in the middle harms the Delaunay condition as its circumcircle includes the two other not adjacent points. **(B)** Accordingly the edge is flipped and the Delaunay condition is fulfilled.

With rising λ more parameters are set to 0 while keeping the mean square deviation to the response vector as low as possible. We used the regularization to find an upper bound for the order of our surface fitting (Section 4.1.3) by hopefully excluding x,y combinations with huge order. The input in our special case were all stem cell coordinates as target array and the matrix of all possible combinations of x and y up to a given order (see Table 3).

Table 3: Matrix used for Lasso input. The z-values for every cell are the response vector and all columns represent different combinations of x and y coordinates.

z	x	y	xy	x^2y	xy^2	x^2y^2	x^2	...
z_1	x_1	y_1	x_1y_1	$x_1^2y_1$	$x_1y_1^2$	$x_1^2y_1^2$	x_1^2	...
z_2	x_2	y_2	x_2y_2	$x_2^2y_2$	$x_2y_2^2$	$x_2^2y_2^2$	x_2^2	...
...	...	...	...	...	...	...	...	...

3.4 Spatial Statistics

3.4.1 Spatial patterns

The spatial organization of points or cells can be described by different spatial pattern. We focus on complete spatial randomness (CSR), attractiveness and repulsiveness. CSR describes points that are distributed over an area as if they are drawn randomly out of all possible positions (see Figure 3.5A).

Points following an attractive pattern in contrast are forming clusters (see Figure 3.5B), while points of a repulsive pattern repel each other and points are predominantly separated (see Figure 3.5C).

3.4.2 Nearest neighbor distance statistics

Nearest neighbors methods are used to find the nearest point of a set to a given point under a certain distance measure. We use a nearest neighbor method in terms of distances on the hemisphere surface (see Section 4.1.3) by looking for the nearest division of every dividing cell (c_i):

$$d_i = \min_{i \neq j} dist(c_i, c_j)$$

3.4.3 F-Function by Andrey et al.

A spatial method introduced by Andrey et al. [28] used a so called F-function to detect spatial properties like CSR, attractive or repulsive patterns with the help of evaluation points. Evaluation points are typical positions of the considered system the hemisphere in our case). The F-function can be described as a cumulative distribution function (cdf) of the distance d' between evaluation points on the hemisphere surface and its closest point in the observed pattern:

$$F(d) = P(d' < d)$$

As evaluation points we used all non-pattern cells on the hemisphere surface, since they represent typical cell locations. For each evaluation point we determine the distance to the nearest point of the observed pattern (Figure 3.6A). Using this distances the cdf is obtained by the proportion of evaluation points with distance smaller than distance d.

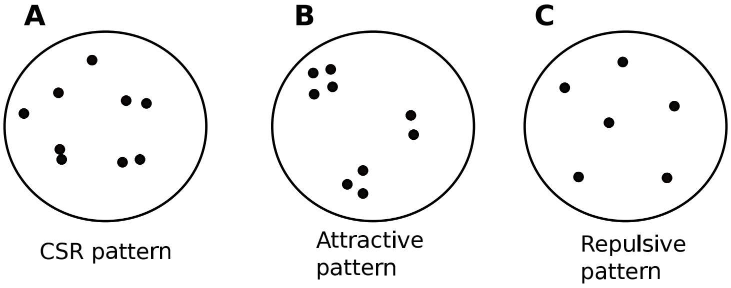

Figure 3.5: Different spatial patterns. **(A)** Completely spatially random (CSR) pattern distributes points randomly. **(B)** In an attractive pattern points cluster together. **(C)** Points are separated if they form a repulsive pattern.

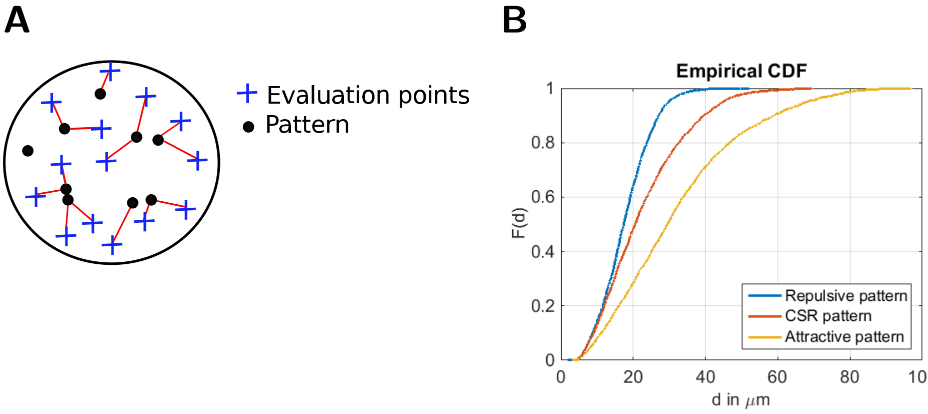

Figure 3.6: The F-function can categorize the data in different patterns. **(A)** To calculate the F-function evaluation points (donated by +) are distributed over the whole area. For every evaluation point the nearest neighbor distance to the points of the pattern is calculated. The distances are predominantly small for regular patterns and large for aggregated patterns. **(B)** The simulation (Section 4.4.1) of repulsive, attractive and CSR patterns visualizes how the cdf of the F-function represents the different pattern.

The resulting cdf is compared to a CSR distribution conditioned on the observed pattern size n_p and the overall cell number n_c. To achieve comparable completely random pattern we draw $n_s = 10000$ samples out of all cells with sample size equal to n_p. We determine the mean of every quantile and compare the cdf to the observed one. Figure 3.6B shows the cdfs of attractive and repulsive pattern compared to CSR pattern. Repulsive pattern are found on the left side of the random sampled mean cdf as every non-pattern cell position finds a relatively near cell of the repulsively distributed cells of the pattern. In contrast attractively sampled cdfs are found on the right of the sampled mean cdf as non-pattern cells often lack of a near cell of the pattern as they are only found in a few clusters.
To evaluate the difference of CSR and the observed pattern we calculated the maximal deviation of the two cdf curves. Low values correspond to light developed pattern while high values indicate strong pattern.

3.4.4 Ripley's K

Ripley's K function is a spatial analysis method, which can describe point patterns occurring over a given area (A) of interest [29]. It has been used to to summarize a point pattern, test hypotheses about the pattern, estimate parameters and fit models. In our case the point patterns are dividing cells and the area is the hemisphere surface. It is important to mention that we evaluate the K function on a surface with non-uniform cell appearance probabilities as divisions can only appear at cell locations. Ripleys K function is calculated at multiple distances to see how point pattern distributions change with scale:

$$\begin{aligned} K(d) &= \lambda^{-1}[\text{number of extra events within} \\ &\quad \text{distance d of an event}] \end{aligned} \tag{3.7}$$

with λ being the density of events corresponding to the ratio of the number of all occurring divisions and the area on which the divisions could appear (n_p/A). In our case the area is approximated by Delaunay triangulation (see Section 3.3.2) and the overall number of events is counted. According to equation (3.7), we can estimate the K function for our data with

$$\hat{K}(d) = N^{-1} \sum_i \sum_{j \neq i} I(d'_{ij} < d) \tag{3.8}$$

where d'_{ij} is the distance between the ith and jth cells, and $I(x)$ is the indicator function with value 1 if x is true and 0 otherwise. The estimated $\hat{K}(d)$ function can be compared to $\hat{K}(d)$ functions of simulated patterns. With our simulation tool (see Section 4.4.1) we check whether our data fits to patterns for (CSR), repulsiveness or attraction (see results in Section 4.4.4).
Increasing d can lead to the inclusion of regions where no events can occur. To take this case into account we use edge correction, which corrects for the proportion of the circle with radius d outside of the area where events can occur. The edge effect is important for large d because it is more likely that large circles reach the outside of the surface. We used a self implemented approach to calculate the cut of the circle with the hemisphere surface. To approximate the area of the hemisphere surface we used the previous calculated Delaunay triangulation (see Section 3.3.2) to first identify the cells defining the border (we picked all edges that just appear once). By traversing the border edges we identified the regions where the circle around the current event cuts the border defined for the hemisphere (J_1, J_2 in Figure 3.7). The intersections and border points inside the circle could approximately determine the slice plane of surface and circle by adding up all triangles defined

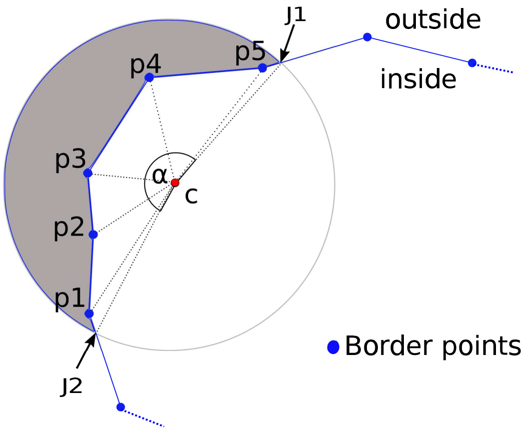

Figure 3.7: Edge correction Calculating the proportion of the circle around an event (red dot), which is not part of the surface. Using the polygonal line of cells defining the border of the hemisphere we obtain the two intersections (I_1, I_2) where the circle cuts the border. To obtain the grey area outside the hemisphere we sum up the area of the triangles defined by the border points inside the circle and the intersection points with the center (dashed edges). Subtracting the resulting area from the pitch circle with angle α we receive the desired area with which we calculate the proportion of the circle lying inside the surface.

by neighboring border points and the center c:

$$w(p_i, d) = 1 - (d^2\pi)^{-1}\big(d^2\alpha\pi - \sum_{i=1}^{n-1} A(p_i, p_{i+1}, c) + A(J_1, p_1, c) + A(p_n, J_2, c)\big) \tag{3.9}$$

where p_i is the ith border point inside the circle, α is the angle between I_1, center and I_2 and $A(x, y, z)$ is the area of the triangle defined by the three point coordinates of x, y and z.

To account for the fact that our surface is no rectangle our approach differs from the normally used one [30]. Normally the edge correction is calculated depending on proportion of the circumference of the circle that fall into the study area, while our approach calculates the proportion of the circle area falling into the study surface.

Including the edge correction in Equation (3.8) we get:

$$\hat{K}(d) = N^{-1} \sum_i w(p_i, d) \sum_{j \neq i} I(d'_{ij} < d). \tag{3.10}$$

3.4.5 Active regions

To get a visual impression of the regions where dividing cells are over-represented, we tiled the 2D projection and calculated sub-images. To create these sub-images the image was divided in 16x16=256 squares (using a squared image).

For every sub-image i the number of cells was counted (C_i) and the fraction of dividing cells to all cells was determined. To take into account that the cells are not distributed equally we sampled the number of dividing cells on all cell locations ($n_s = 1000$). For every sample n we counted the number of cells inside every square D_i again. Having a background distribution of dividing cells per sub-image we can calculate an empirical p-value [31] for the over-representation of dividing cells per sub-image over n_s simulations:

$$p_i = n_s^{-1} \sum_n I\left(\frac{D_i}{C_i} > \frac{D_{in}}{C_i}\right) \tag{3.11}$$

with D_{in} being the number of dividing cells in sub-image i in simulation n and D_i the respective number of dividing cells in the original image.

4 Results

4.1 Image processing pipeline to identify single cells in 3D confocal image stacks

To quantitatively estimate cell patterns we first had to identify single cells and make sure that the identified cells were not biased by image artifacts. Furthermore we had to calculate distances between cells, for which we approximated the hemisphere surface on which the cells were located. This was necessary since the hemisphere was not flat and we needed to account for unevenness on it. To solve this tasks, we implemented a pipeline (see Figure 4.1), that we explain exemplary for experiment ZF1L in the following.

4.1.1 Exclusion of uninformative regions

Due to experimental setup and imaging conditions some areas of the image stack were not of interest for further analyses. We thus excluded non-

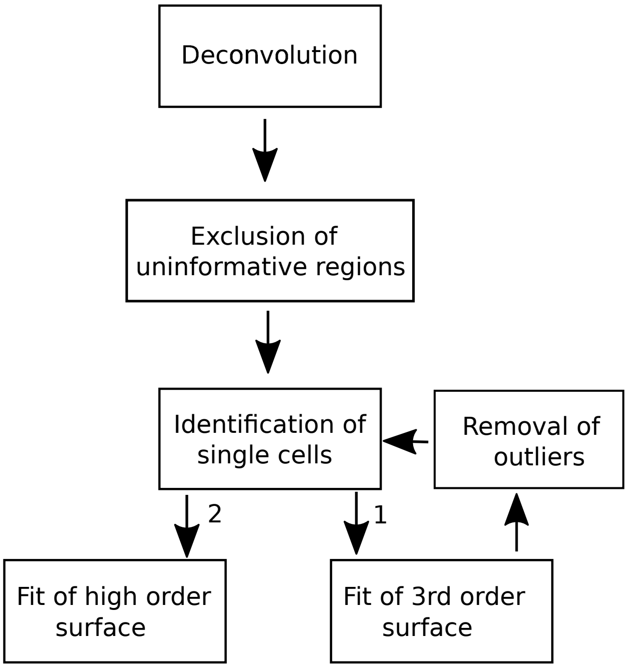

Figure 4.1: Overview of the pipeline that describes all preprocessing steps which were explained in detail in Section 4.1.

hemisphere regions of the image, which showed microscopic artifacts or the other hemisphere. We also excluded overexposed regions, where single cells cannot be distinguished (see Figure 4.2A). The masked area for ZF1L was ∼ 15% of the whole image, while it covered ∼ 45% and ∼ 42% in ZF2L and ZF2R, respectively, due to large non-hemisphere regions.

All pixels that were part of the masked area were set to background intensity (see Figure 4.2B), which had an intensity of 3 for an 8 bit image. The mask identified on the MIZP was processed to the whole image stack.

4.1.2 Identification of single cells

After isolating informative hemisphere regions single cells need to be identified in the images, to derive their exact coordinates in 3D for later analysis. A first attempt to segment cells with a 3D method (see Section 3.2.1) failed, as only around 60% of all cells were correctly identified (see Figure 4.3). This could be the consequence of the image quality, the distance between the image stacks or the low contrast of single cell staining in each single slice. Consequently we developed an approach, based on MIZP of the full image stack, where the single cell staining is intensified in the projection and the according intensity profile in z.

After performing the Laplace of Gaussian method (see Section 3.1.2) on the

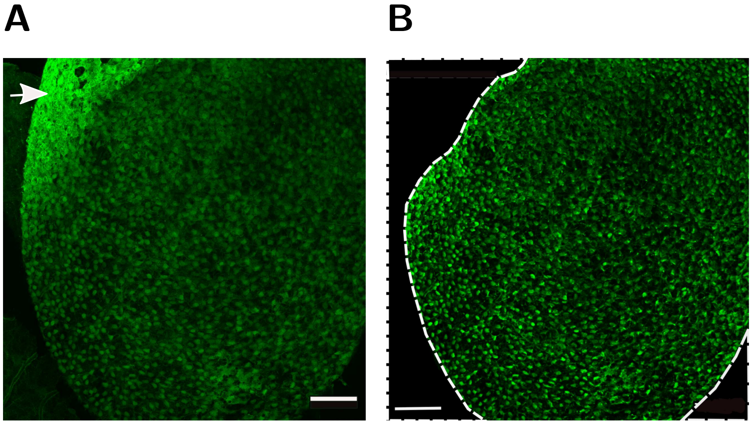

Figure 4.2: Masking of uninformative regions. **(A)** The GFP stained (green) image stack of ZF1L as MIZP. The arrow marks an overexposed region, which had to be excluded. **(B)** Same image after masking. On the top left side overexposed regions were excluded. On the bottom left as well as on the bottom right side uninformative regions were masked (dashed line), to prevent inclusion of image artifacts. Roughly 15% of the image are masked. Scale bars: 50 μm

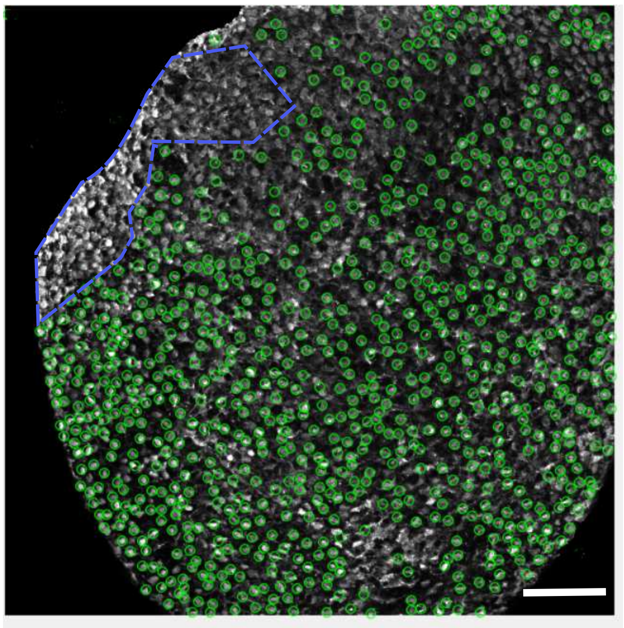

Figure 4.3: The result of the 3D segmentation overlapped with the MIZP (grey background image). Green circles denote identified cell locations. The 3D segmentation did not identify cells that could be identified on 2D projection by an expert, especially in the blue marked area, where only $\sim 60\%$ of cell can be identified. Scale bar: 50 μm

MIZP images 3380 blobs (cell candidates) were identified. From every detected blob we used its x,y centroid, for further analyses. The next step was to find a z-coordinate for each blob. A search for the maximum intensity in z direction could lead to false positives due to noise in the image acquisition e.g. via single glowing pixels with high intensity. To discard possible signals from neighboring cells, we only used pixel information inside a cell sized circular area of radius $r_c = 2\mu$m around the centroids in every image (Figure 4.4A). The combination of circles of all slices formed a tube in z-direction (Figure 4.4B). On every image we calculated the mean intensity inside the respective circle and fitted a Gaussian distribution in z-direction to it (see Figure 4.4C). We took the mean of the distribution as the z coordinate of the cell. Additionally the standard deviation was used to filter whether the detected blob was able to represent a real cell or arose only from an image artifact like glowing single pixels. If the range of two standard deviations σ_z was smaller than an average cell radius r_c the blob was excluded as a false positive.

Analogously we fitted a two dimensional Gaussian distribution in x-y dimen-

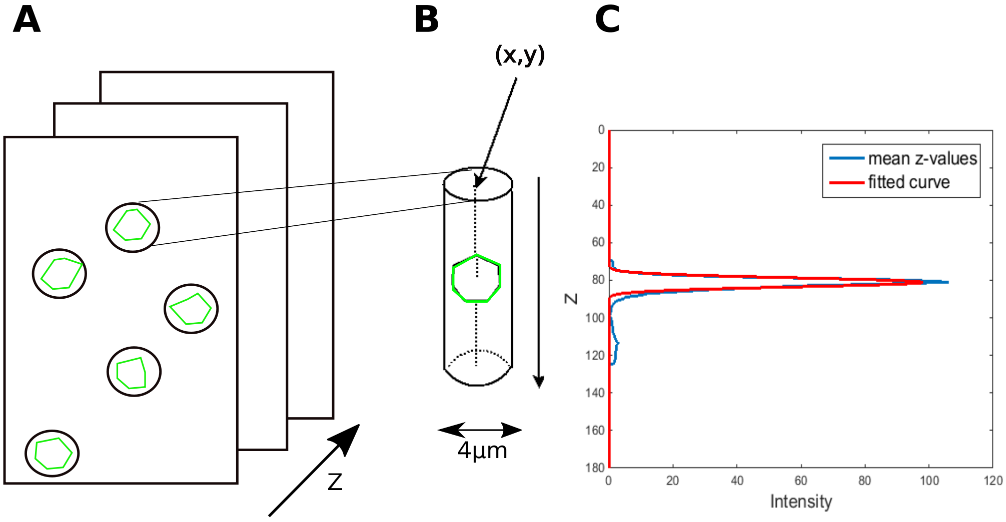

Figure 4.4: Detection of cell centers in z-direction. For every identified blob (circles in **(A)**) a tube in z-direction was formed **(B)**. **(C)** The mean intensities inside the circle per slice in z-direction were calculated (blue line). We fitted a Gaussian distribution to this curve (red line). Using the mean of the distribution in z-direction as z-coordinate of the cell, we robustly determined cell centers in z.

sion we excluded artificial blobs with $2\sigma_i < r_c$ with $i \in (x, y)$.
Compared to the direct 3D segmentation method this approach can identify 100% of all manually counted cells. Additionally cells being close to each other, like recently divided cells, can be identified as two cells while the 3D segmentation would exclude them as one too big object.

4.1.3 Approximation of the hemisphere surface

The hemisphere of a zebrafish is not planar but curved and has also grooves on it (Figure 4.5A). As euclidean distances do not represent the real distances between cells on the surface, it was important to take the structure of the hemisphere surface into account. We used the x,y,z centroids for every cell and fitted a general polynomial regression model to all centroids to approximate the hemisphere surface.
For choosing the right polynomial order we had to consider two issues. (i) Image artifacts (marked white in Figure 4.5A) could result in outliers and attract the fitted surface (as shown in Figure 4.5B); (ii) lower orders could be too general to approximate the unevenness of the hemisphere. Hence, we decided to apply a two step approach. First we used the approximation of

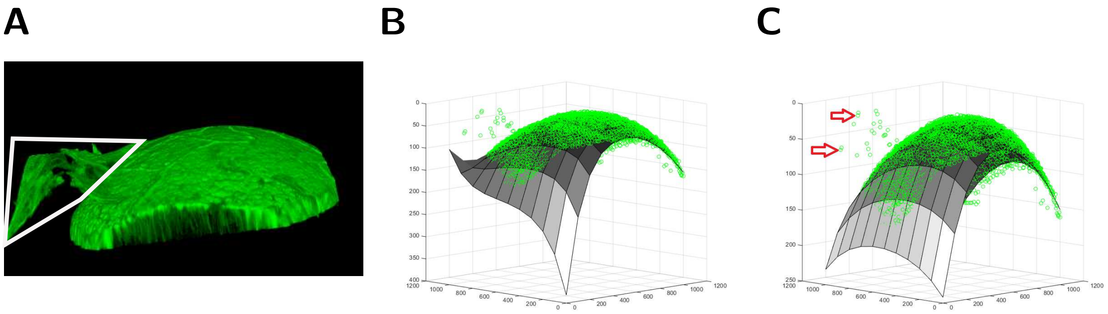

Figure 4.5: Identification and filtering of segmentation artifacts. **(A)** An exemplary region of erroneously segmented foreground objects (inside white border) on a 3D rendered image stack, which could influence subsequent analyses if they were not excluded. **(B)** The 4th order polynomial could be attracted to outliers resulting from image artifacts, which was the reason for using a two step approach. We started with fitting a 3rd order polynomial **(C)** to the data, which only roughly approximated the surface. We then excluded the outliers (arrows) and fitted a higher order polynomial to the remaining data.

the hemisphere surface with a low order (n=3) to exclude image artifacts and afterwards we fitted a higher order polynomial to the surface to account for local unevenness.

The previously mentioned outlier problem was fixed by repetitively excluding the cell furthest from the surface (see marked spots in Figure 4.5C) and recalculating the surface until no spot was further away than 8μm (approximate two cell diameters) and the median of all distances to the surface was smaller than 0.1 μm. The resulting surface was then used to eliminate all remaining image artifacts by applying an upper and a lower bound to the calculated surface with z-distance of 12 μm (approximate three cell diameters) each. Every pixel outside this borders was set to background intensity (intensity of 3 for a 8-bit image).

Afterwards the segmentation (Section 4.1.2) was redone and a higher order polynomial was fitted to the data to consider unevenness of the brain surface.

To evaluate the order we applied a lasso fit (see Section 3.3.3) to the centroids. The lasso fit should reveal which low orders were too general and which higher orders lead to overfitting. We applied the lasso fit to combinations of polynomial terms up to 30th order (see Table 3). In the best case lasso excluded all high order polynomial combinations and we could use the highest lasso order as the polynomial order for our fitting. Unfortunately the fitted lasso term still included high order terms (up to 28th order) and the resulting mean square error of all centroids was only slightly lower than the one calculated for the fitted surface of order 8. Hence the lasso term was not usable and we had to define the order manually. We did this by looking for increasing orders and evaluating whether they cover all visible cells on z-projection. The smallest order covering all cells was taken. The order for ZF1L was 8, for ZF2L 10 and for ZF2R 12.

4.1.4 Identification of dividing stem cells

After the identification of all GFP expressing stem cells we applied the same algorithm on EdU stained dividing cells. Stem cells with EdU will be called double positive cells in the following. To identify them we compared the locations of all dividing cells with the positions of all stem cells and if they were closer than one cell diameter (4μm) they were labeled as double positive.

4.1.5 Application of the pipeline to three hemispheres

We applied the final pipeline to a total of 3 hemispheres. In Figure 4.6 all three hemispheres are displayed showing the identified GFP, EdU and double positive cells. We obtained 2668 GFP positive cells, 86 EdU positive cells and 17 double positive cells for experiment ZF1L. Experiment ZF2L shows 2466, 90 and 14 respectively, while ZF2R results in 2349 GFP positive cells, 175 EdU positive cells and 12 double positive cells (Table 4). The cell numbers were comparable between the different experiments except for the EdU count of ZF2R, which was about the twice as much as in the two other hemispheres. This could influence the comparability of the results between the experiments in the following analyses.

Table 4: Number of stem cells (GFP positive), dividing cells (EdU positive) and dividing stem cells (GFP and EdU positive), quantified in three independent image stacks. Names of experiments follow the scheme ZFx[LR], where x is the number of the zebrafish while L/R stands for left/right hemisphere.

	#GFP	#EdU	#GFP+EdU
ZF1L	2668	86	17
ZF2L	2466	90	14
ZF2R	2349	175	12

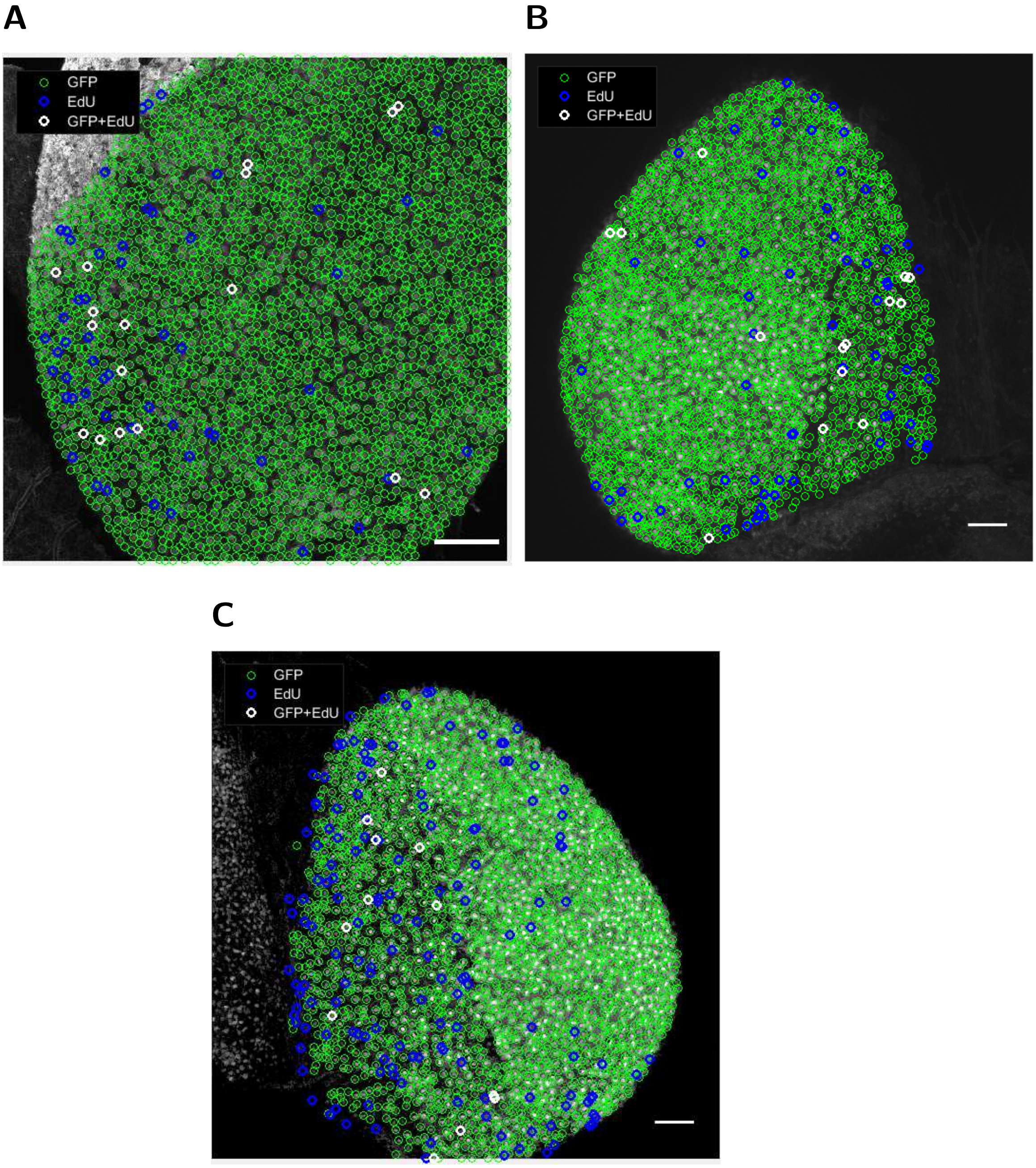

Figure 4.6: Results of the segmentation for the three hemispheres displayed on the MIZP raw image stack. **(A)** Experiment ZF1L with 2668 GFP positive cells (green circles), 86 EdU positives (blue circles) and 17 double positives (white circles). **(B)** Experiment ZF2L with 2466 GFP cells, 90 EdU cells and 14 double positives. **(C)** Experiment ZF2R with 2349 GFP positive cells, 175 EdU positive cells and 12 double positive cells. Scale bars: 50 μm

4.2 Doublet statistics

Within the group of EdU positive cells, many cells close by appear, which we defined as doublets that were already split and could be identified on the MIZP of the EdU staining as close neighboring dividing cells (see Figure 4.7A). Doublets can be composed of three different cell combinations: (i) Both stay stem cells, (ii) both developed to progenitors or (iii) one stem cell and one progenitor cell (see Figure 4.8). We found that doublets, which consisted of two progenitor cells clearly form the majority with 80%, 75% and 80% for ZF1L, ZF2L and ZF2R respectively. Asymmetric divisions appeared to be the second largest combination (15%, 8% and 13%). The third combination were two stem cells with 5%, 17%, 6% respectively.

Interestingly divisions leading to two progenitors did not necessarily have a stem cell as mother cell (compare Figure 4.8B,C). The observed doublet results can still change as stem cells that are still in the division process could give rise to additional progenitor cells.

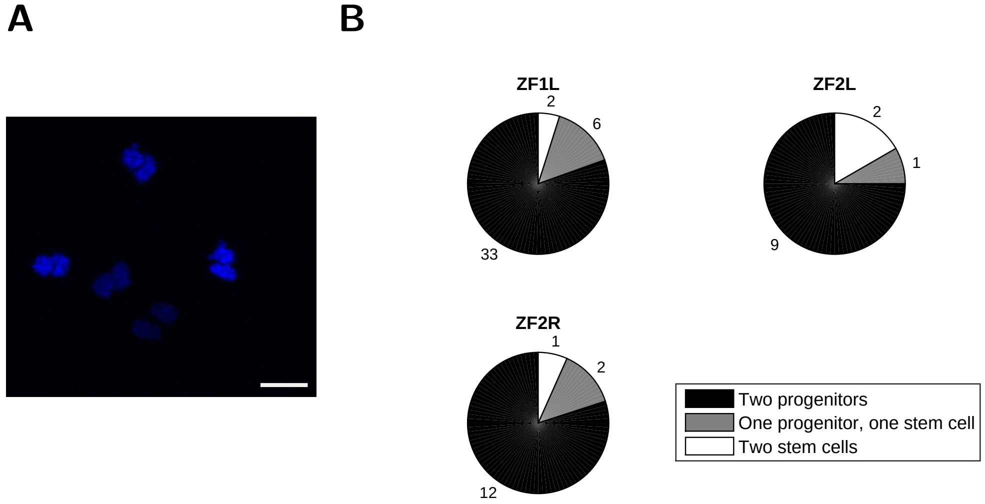

Figure 4.7: Cell divisions could give rise to three different cell combinations. **(A)** Cutout of a z-projected image stack showing EdU stained doublets on the maximum intensity projection (blue blobs). Scale bar: 10 μm **(B)** Two progenitors clearly form the majority with 80%, 75% and 80% for experiments ZF1L, ZF2L and ZF2R respectively, while asymmetric divisions tend to be the second largest combination (15 %, 8 % and 13%). The third combination were two stem cells with 5%, 17%, 6% respectively.

To verify the results of the cell identification we compared our results to the observations of an unpublished study on zebrafish stem cells by Barbosa et al. We estimated the average length of the S-phase using their results. The study states that around 12% of stem cells undergo asymmetric division and 1% symmetric divisions within two weeks. Transferred to experiment ZF1L, where we observed 2668 stem cells we expect a number of 2668x0.13 = 346.84 stem cell divisions within 14 days. The time required to observe 17 dividing stem cells is thus (17x14)/346.84 = 0.68 days which is about 16h to incorporate EdU. The respective results for ZF2L and ZF2R were about 18h and 20h. Since EdU can be incorporated during the whole S-Phase the stem cell can be at the beginning of the S-Phase (Figure 4.8E) or at the end of the S-Phase (Figure 4.8D) and thus the temporal window increases to approximately the length of two S-Phases. Consequently we calculated the length of two S-Phases and the resulting S-Phases are 8h, 9h and 10h long.

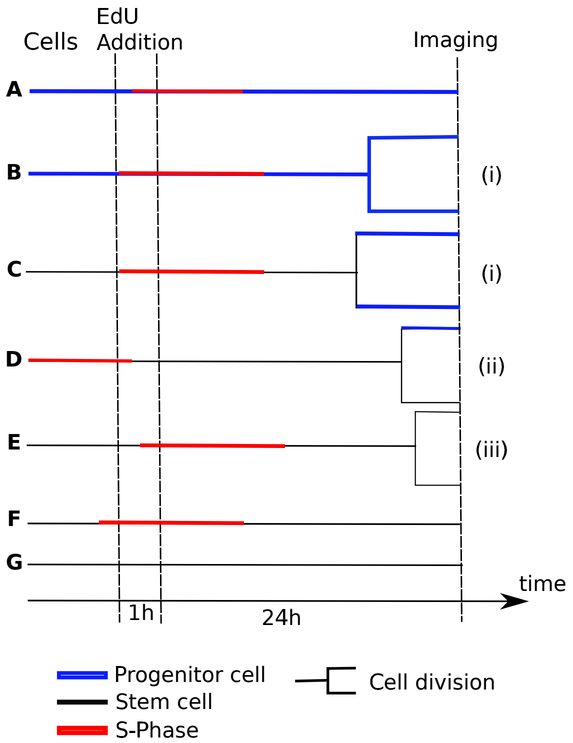

Figure 4.8: Cells observed at the imaging time point could have different origins. Observing an EdU stained cell could either be a progenitor cell that already passed its S-Phase **(A)** or was at least one part of an earlier division **(B)-(D)**. A dividing stem cells could arise from an undivided cell **(F)**, from a symmetric stem cell division **(E)** or from an asymmetric division **(D)**. A GFP only cell could just be the consequence of a quiescent stem cell **(G)**.

4.3 Distances to surface

To examine if dividing cells are still located on the hemisphere surface as addressed from Chen et all [32], we calculated the shortest distance of every dividing cell to the surface .

As shown in Figure 4.9 we found 12 clearly visible neuroblasts residing under the hemisphere surfaces (> 20 μm which is about 5 cell diameter) probably moving in direction of the inner brain to a region where new neurons were needed [32]. 95% of dividing stem cells were found under the hemisphere, but they were located only up to three cell diameters (≈ 12 μm) below the surface, which could be a consequence of the division process. Possibly the cells need more space to divide and move away from the neighboring cells for this purpose. To evaluate the surface fitting we also compared the locations of non-dividing stem cells to the himsphere surface. As expected they were found equally distributed under and above the calculated hemisphere surface with a maximum distance of 2 cell diameter.

Noticeably was a subset of dividing cell residing clearly above the surface (up to 30 μm) in ZF2R. This could be a consequence of the lack of stem cells in the lower left region of the hemisphere (see Figure 4.6C). As there were not enough stem cell the fitted surface drops before it reaches the outer regions. Following all dividing cells in the lower left regions were found above the calculated surface.

4.4 Spatial patterns of dividing cells

To categorize the spatial patterns of dividing cells on the hemisphere surface we applied three statistical methods. We applied all methods on two cell populations: dividing stem cells (identified by a GFP and EdU signal) and all dividing cells (identified by an EdU signal). The first method compares statistical properties of nearest neighbor distance distributions. The second method reveals relationships between randomly distributed points on the surface to locations of dividing cells. Ripley's K as the third method describes more complex relationships between the dividing cells by taking into account that cells at the border of the hemisphere can have no adjacent divisions outside the hemisphere.

4.4.1 Spatial pattern simulation

We compare the observed patterns with three different scenarios: (i) the division rate p is identical for every cell, which means that every cell has the

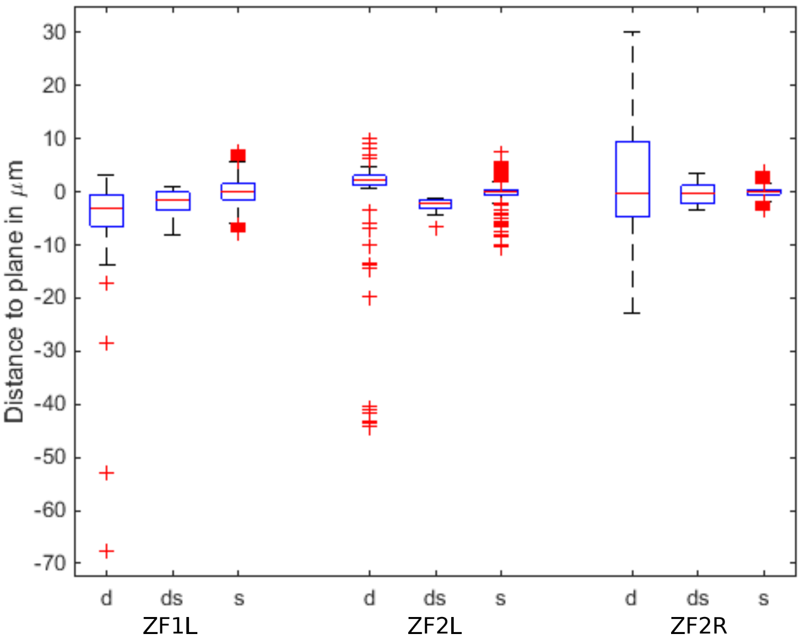

Figure 4.9: Distances of cells to the surface of all three experiments for all dividing cells (d), dividing stem cells (ds) and stem cells (s). 12 dividing progenitors can be clearly found below the surface (distance $> 20\ \mu$m ≈ 5 cell diameters) while non-dividing stem cells were located near the surface. In contrast 95% of dividing stem cells were found under the hemisphere with a distance of three cell diameters ($\approx 12\ \mu$m) at most.

same probability to divide at every time point (complete spatial randomness (CSR) scenario). (ii) A cell division increases the chance that a cell divides nearby (attractive scenario). (iii) Cells that divide decrease the probability of neighboring cells to divide (repulsive scenario).
The impact of a division on other cells can be influenced by the two parameters "influence radius" r and "influence strength" s. Thus in the attractive scenario, the division probability p_i for a certain cell c_i is a function of

$$p_i = ps \sum_{j \neq i} I(d(c_i, c_j) < r) \tag{4.1}$$

with p being the CSR division probability and $d(c_i, c_j$ denotes the distance between cell c_i and cell c_j. In the repulsive scenario the division probability

p_i for a certain cell c_i is a function of

$$p_i = p\Big(s \sum_{j \neq i} I(d(c_i, c_j) < r)\Big)^{-1} \tag{4.2}$$

As input the simulation gets the locations of all cells (regardless of their state) and the surface on which they were located (see Section 4.1.3). The simulation works by drawing a random not yet divided cell and simulates a division, which simply increases or decreases the probability of neighboring cells to be drawn. The drawing of a cell takes into account that cells can have different probabilities to divide and adjusts the drawing accordingly.
With the simulated spatial cell distributions of dividing cells we were able to compare different settings to real data and fit the two model parameters to the data by optimizing the influence of radius and strength.

4.4.2 Nearest neighbor statistics

To analyze the patterns of dividing stem cells and all dividing cells on the zebrafish hemisphere we first looked at the nearest neighbor distances (d_{NN}). Figure 4.11A displays distributions of d_{NN} for dividing stem cells on ZF1L. We observed 10 (59%) small distances (< 20 μm ≈ 5 cell diameters) but also one (3%) large one (> 50 μm ≈ 12 cell diameters). Figure 4.11B shows 63 small distances (73%) and 6 (7%) large ones for all dividing cells. Comparing both cases we could state that d_{NN} of all dividing cells contained more extreme values.
We next compared the observed d_{NN} to a background distribution. For the respective number of points in the patterns n_p, we randomly sampled n_s=1000 cells distributed over all possible cell positions to generate a background distribution of d_{NN} (see Section 3.4.2). Using this distribution we could compare mean, median and standard deviation between observed and sampled distances.
An empirical p-value was calculated by counting the values of the background distribution being higher or lower than the observed value and dividing it by the number of samples n_s [31]. As shown in Figure 4.11C,G the observed median and mean of ZF1L were below all sampled values, suggesting that the empirical p-value was smaller than 0.001. Similar empirical p-values were calculated for mean (empirical p-value = 0.011) and median (empirical p-value $<$ 0.001) of d_{NN} for all dividing cells (Figure 4.11C,G). Both empirical p-values suggested that we observe significantly more small d_{NN} than by distributing cells randomly. Consequently every cell that was part of the observed pattern had at least one dividing cell nearby, but they not

necessarily needed to cluster all together. Looking at standard deviation of d_{NN} we observed different patterns (Figure 4.11E-F). Standard deviations of dividing stem cells tended to be smaller than the randomly sampled distribution (empirical p-value = 0.013), while d_{NN} of all dividing cells had a significantly greater standard deviation compared to random sampled ones (empirical p-value < 0.001). This was surprising, as this implied that dividing stem cells cluster always together while all dividing cells include some cells that were not near any other division.

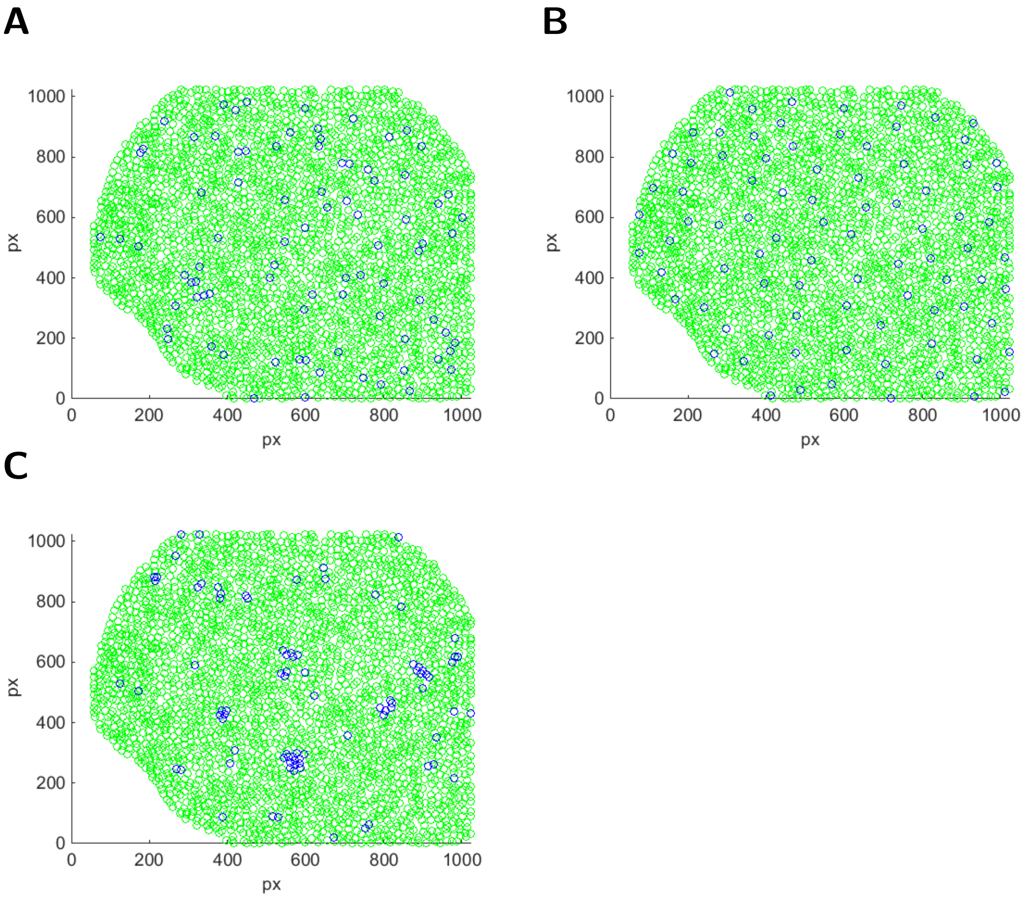

Figure 4.10: Simulation cell division patterns for three different scenarios. Green circles represent stem cells while blue circles represent division events **(A)** The division appearances of cells were sampled randomly e.g. without influencing each other. **(B)** Cells that divided decrease the division probability for their neighboring cells and a repulsive pattern appears. **(C)** An attractive pattern appears by increasing the division probability for cells located near other divisions.

The observed d_{NN} of dividing stem cells for experiment ZF2L showed 8 (57%) small distances and 3 (21%) large ones, while for all dividing cells 47 (52%) d_{NN} were small and 3 (3%) large (see Figure 4.12A,B). Comparing mean d_{NN} to the background distribution we observed small values in both cases with p-values of 0.012 and 0.013 for dividing stem cells and all dividing cells, respectively (Figure 4.12C,D). The standard deviations were in this case significantly higher than expected (empirical p-values of 0.019 for dividing stem cells and 0.001 for all dividing cells, Figure 4.12E,F). The biggest deviation between dividing stem cells and all dividing cells in ZF2L was found comparing d_{NN} medians. While the dividing stem cells showed a significantly smaller median (empirical p-value < 0.001), the median of all dividing cell was not significantly different from the sampled distribution (empirical p-value of 0.111, Figure 4.12G,H).

Comparing d_{NN} on ZF2R we found 2 (17%) small distance and 4 (34%) large ones for dividing stem cells (Figure 4.13A). For all dividing cells we obtained 100 (58%) small d_{NN} and 3 (2%) large ones (Figure 4.13B). The mean of d_{NN} of dividing stem cells was significantly smaller than the background distribution (empirical p-value of 0.003, Figure 4.13C). The mean d_{NN} of all dividing cells in contrast was insignificant (empirical p-value $=$ 0.143, Figure 4.13D). Also the standard deviation of d_{NN} of dividing stem cells was not significant (empirical p-value $= 0.149$, Figure 4.13E) while it was significantly larger than expected by the background distribution for all dividing cells (empirical p-value $= 0.019$, Figure 4.13F). The empirical p-values calculated for the median of dividing stem cells was insignificant with 0.045 and also insignificant for all dividing cells (empirical p-value $= 0.386$). In summary we could state that means of dividing stem cells were significantly smaller than expected by the background distribution in all three experiments, while the standard deviations showed no conclusive tendency in the three experiments. For all dividing cells only the standard deviation showed a conclusive tendency by being significantly higher than expected by the background distributions for all hemispheres. This indicates that dividing stem cells tend to form predominant cluster but also include a 1-2 cells that stand alone. All dividing cells also form cluster but they include much more cells that have no near neighbor.

4.4.3 F-function by Andrey et al.

Nearest neighbor statistics only consider the nearest neighbor distances but not the hemisphere as a whole and the location of the dividing cells on it. To analyze non-local patterns on the hemisphere we applied the F-function, which uses the distances of all non-dividing stem cells to its nearest cell in

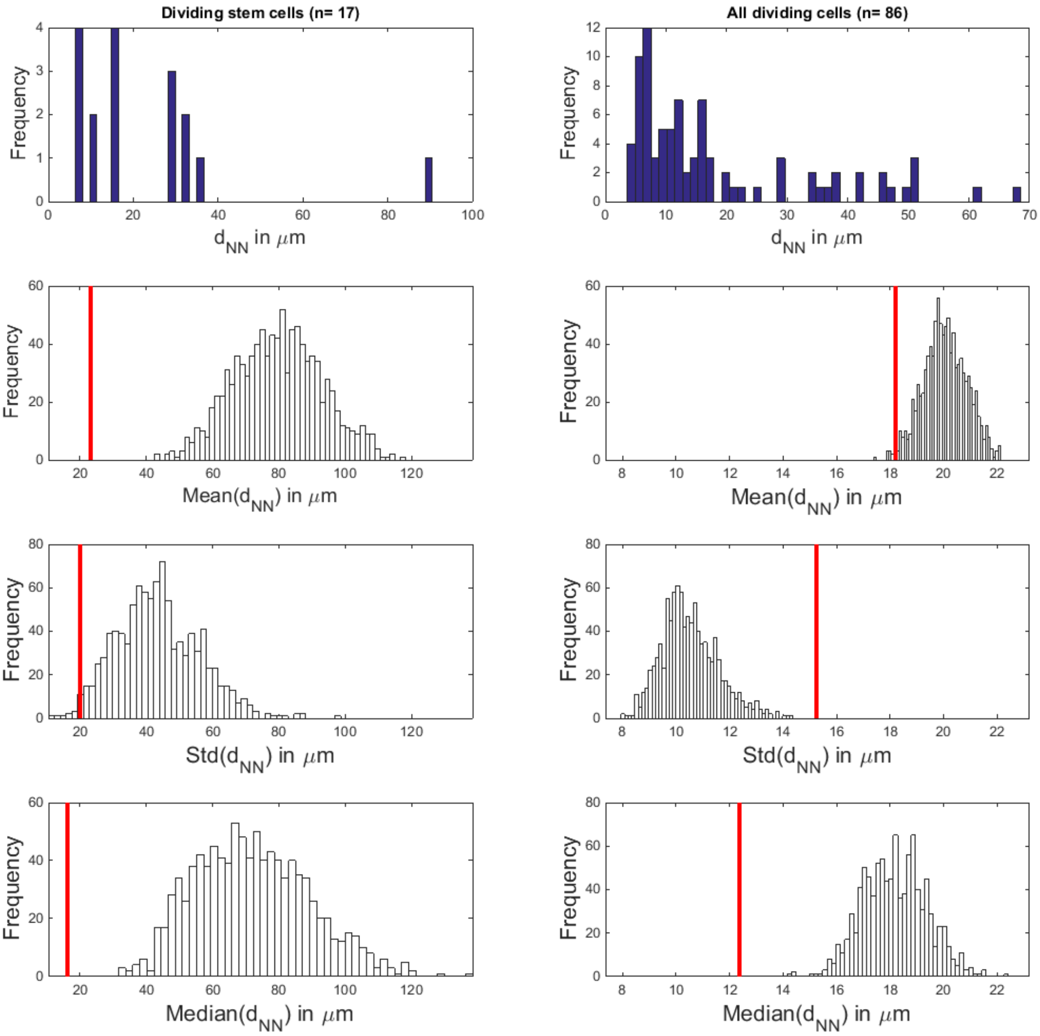

Figure 4.11: Nearest neighbor statistics performed on experiment ZF1L. **(A)-(B)** Nearest neighbor distances d_{NN} observed for both patterns, showing 10 (59%) small distances (< 20 μm ≈ 5 cell diameter) but also one (3%) large one (> 50 μm ≈ 12 cell diameter) for dividing stem cells. The d_{NN} for all dividing cells contained 63 small (73%) and 6 (7%) large distances. In **(C)-(H)** the red line indicates the observed value while the white bars are the background distribution. **(C)-(D), (G)-(H)** Mean and median of the nearest neighbors distances compared to background distributions. Both patterns showed significantly smaller mean and median values compared to the background distribution (empirical p-values < 0.001, except **(D)** with p-value $= 0.011$). **(E)-(F)** The standard deviation differs between both pattern. The dividing stem cells tended to have a small standard deviation (empirical p-value $= 0.013$) compared to background distribution while all dividing cells showed a significantly larger standard deviation (empirical p-value < 0.001).

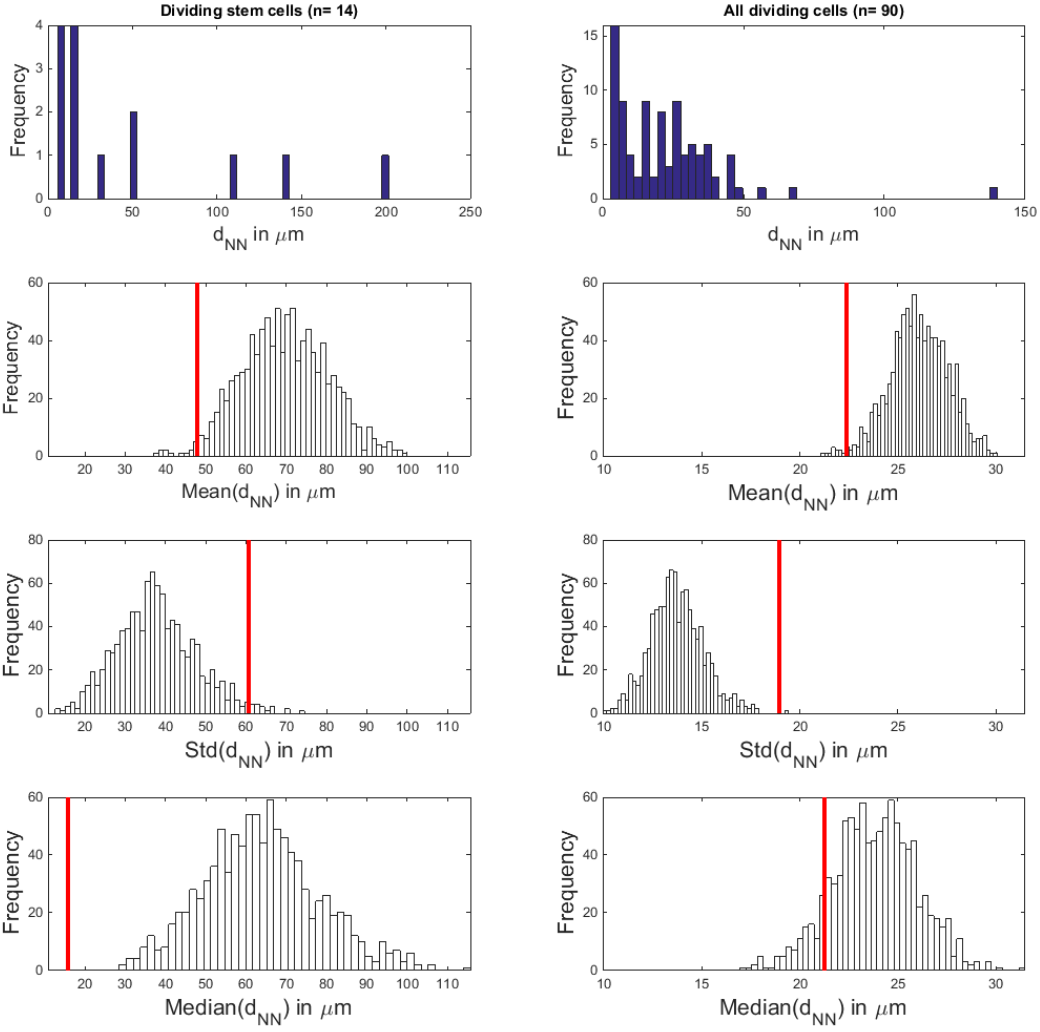

Figure 4.12: Nearest neighbor statistics performed on experiment ZF2L. **(A)-(B)** Nearest neighbor distances d_{NN} observed for both pattern, showing 8 (57%) small distances (< 20 μm ≈ 5 cell diameter) but also 3 (21%) large ones (> 50 μm ≈ 12 cell diameter) for dividing stem cells. The d_{NN} for all dividing cells had 47 small (52%) and 3 (3%) large distances. In **(C)-(H)** the red line indicates the observed value while the white bars are the background distribution. **(C)-(D)** Mean of the d_{NN} compared to background distributions. Both patterns showed significantly smaller mean values as the background distribution with empirical p-values of 0.012 and 0.013 for dividing stem cells and all dividing cells respectively. **(E)-(F)** The standard deviations were significantly higher than expected resulting in empirical p-values of 0.019 for dividing stem cells and 0.001 for all dividing cells. **(G)-(H)** Median distances differed for both cases as dividing stem cells had a significantly lower median d_{NN} as expected (empirical p-value < 0.001) and all dividing cells had an insignificant median (empirical p-value $= 0.111$).

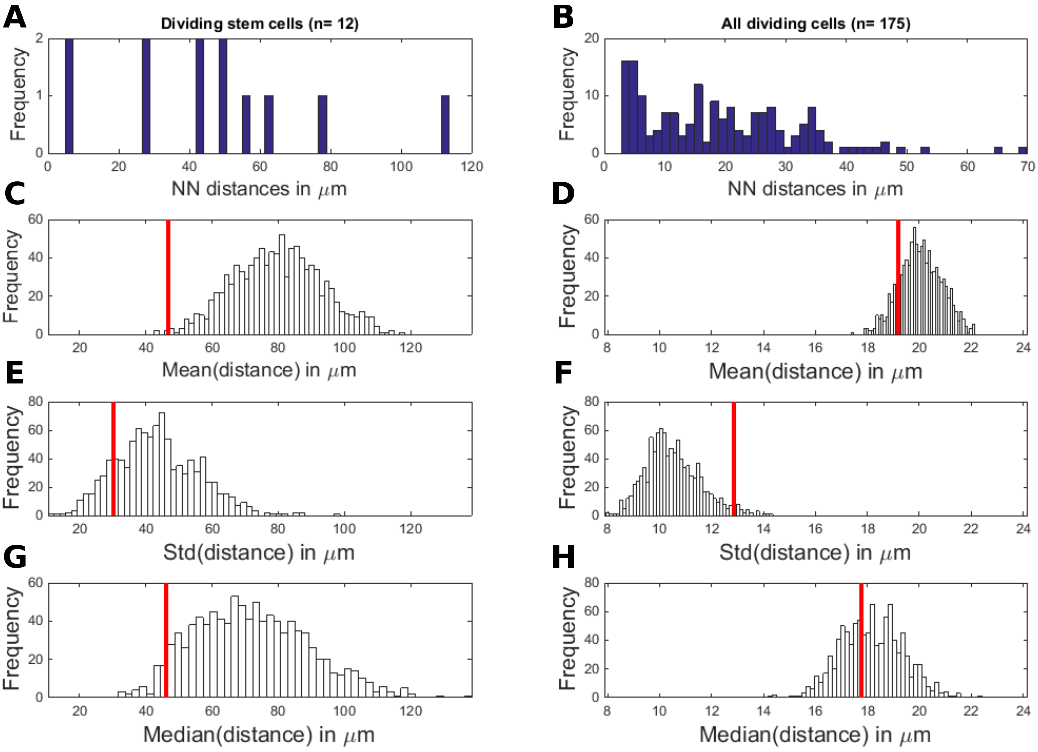

Figure 4.13: Nearest neighbor statistics performed on experiment ZF2R. **(A)-(B)** Nearest neighbor distances d_{NN} observed for both patterns, showing 2 (17%) small distances (< 20 μm $\approx$ 5 cell diameter) but also 4 (34%) large ones (> 50 μm $\approx$ 12 cell diameter) for dividing stem cells. The d_{NN} for all dividing cells had 100 (58%) small and 3 (2%) large distances. In **(C)-(H)** the red line indicates the observed value while the white bars are the background distribution. **(C)-(D)** The mean of d_{NN} of dividing stem cells were significantly smaller than the background distribution showing an empirical p-value of 0.003. The mean value of all dividing cells in contrast was insignificant (empirical p-value = 0.143). **(E)-(F)** The standard deviation of d_{NN} of dividing stem cells was insignificant (empirical p-value = 0.149) while it was significantly higher than expected for all dividing cells (empirical p-value = 0.019). **(G)-(H)** The empirical p-value calculated for the median of dividing stem cells was significantly smaller than expected with 0.045, while it was insignificant for all dividing cells (empirical p-value = 0.386).

the pattern and evaluates the cumulative distribution function (cdf) of these distances by comparing it to the cdf of completely spatial random (CSR) pattern [28]. For more details see Section 3.4.3.
For experiment ZF1L the resulting cdfs (Figure 4.14A,B) implied that for both patterns the cdfs differs from the cdf of a CSR, as the cdf function all dividing stem cells lay outside the 95% envelope of CSR cdf and the cdf of dividing stem cells cut the envelope only in a small interval for $d_{NN} >100$.
For experiment ZF2L the cdf of dividing stem cells cut the 95% envelope of the CSR cdf in the small interval for $d_{NN} >150$ (Figure 4.14C), while the cdf of all dividing cells lay to the right of the whole CSR cdf (Figure 4.14D). For experiment ZF2R the cdfs of dividing stem cells as well as all dividing cells were found completely to the right of the CSR cdf (Figure 4.14E,F). Cuts for $d_{NN} < 5$ were negligible as until this distance no neighboring cell can be found and thus the cdf curves had to overlay.
In summary cdfs of all dividing cells were found to right of the CSR cdfs and thus tend to form attractive pattern. Cdfs of dividing stem cell were predominately found right of CSR cdfs and we can conclude that they tend to form only slight attractive pattern.

4.4.4 Ripley's K

One weakness of the F-function is that it can not distinguish between the cases of having two equally sized clusters and having a small and a huge cluster. The distances between evaluation points and nearest points in the clusters would be nearly identical for both cases although the patterns are different. We thus applied a more complex spatial method called Ripley's K [29] (see Section 3.4.4), that takes the environment of every cell in the pattern into account. Specifically, all divisions around every cell in the pattern are counted within circles around every point of the pattern:

$$\begin{aligned} K(d) &= (n_p/A)^{-1}[\text{number of extra divisions within} \\ &\quad \text{distance d of an dividing cell}] \end{aligned} \tag{4.3}$$

with n_p being the number of points in the pattern and A the hemisphere area. Applying Ripley's K to dividing cells we were able to compare them to simulated completely spatially random (CSR) distributed, repulsively distributed and attractive distributed cells.
To evaluate the Ripley's K properly, we needed to estimate the whole area of the surface the cells were located on. For this purpose we applied the Delaunay triangulation algorithm (see Section 3.3.2 and Figure 4.15).

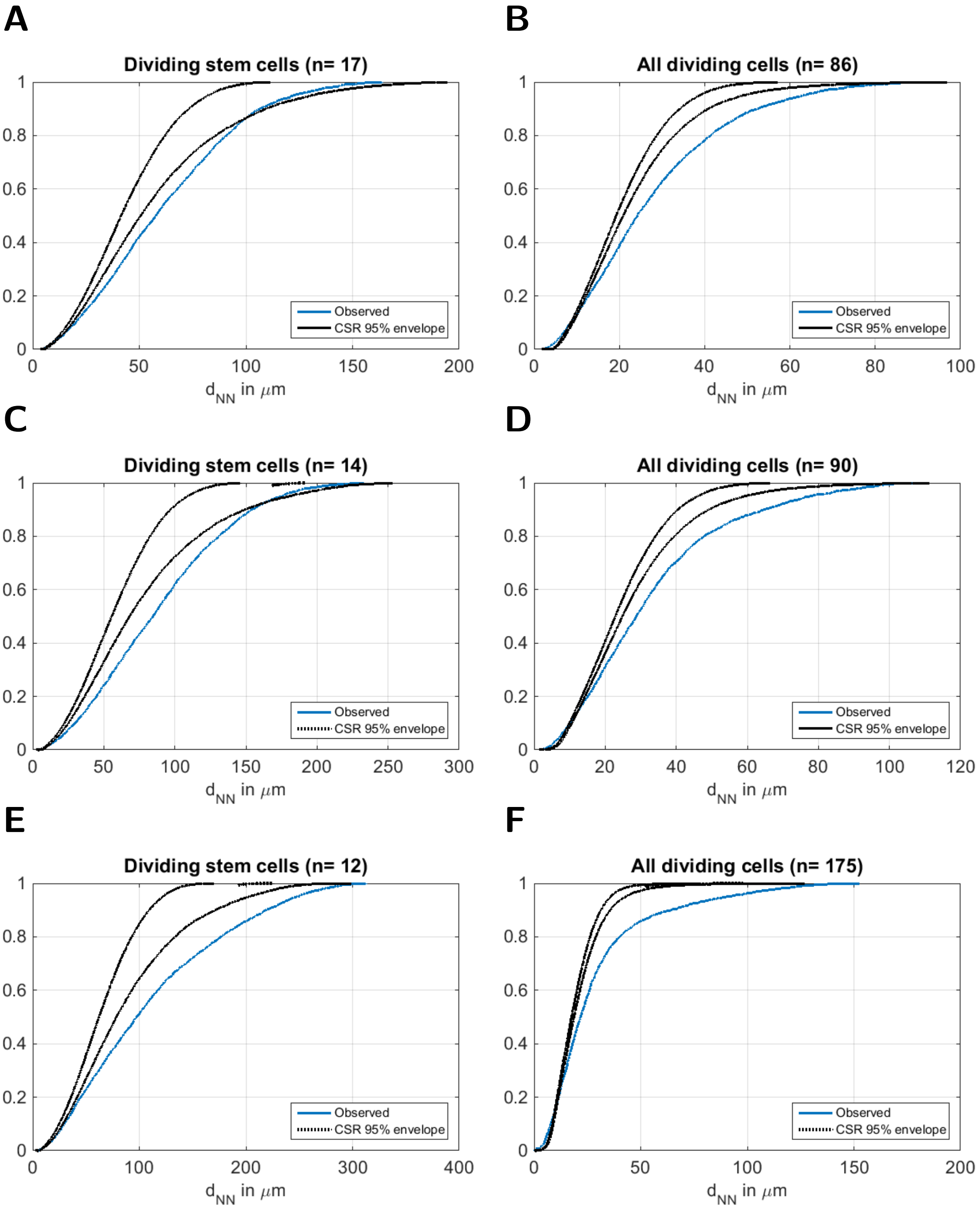

Figure 4.14: The cumulative distribution function (cdf) of the F-function distances. The blue line indicates the observed distances, while the black dashed line represent randomly sampled mean distances. The cdf of dividing stem cells as well as all dividing cells arise right to the respective random sampled mean cdf, which indicates that both patterns show a trend to form an attractive pattern. **(A),(B)** In experiment ZF1L the maximum distances between both curves were about 24 μm each. **(C),(D)** For experiment ZF2L we observed 46 μm and 31 μm maximum deviation for dividing stem cells and all dividing cells respectively, while **(E),(F)** ZF2R even showed 101 μm and 74 μm deviation.

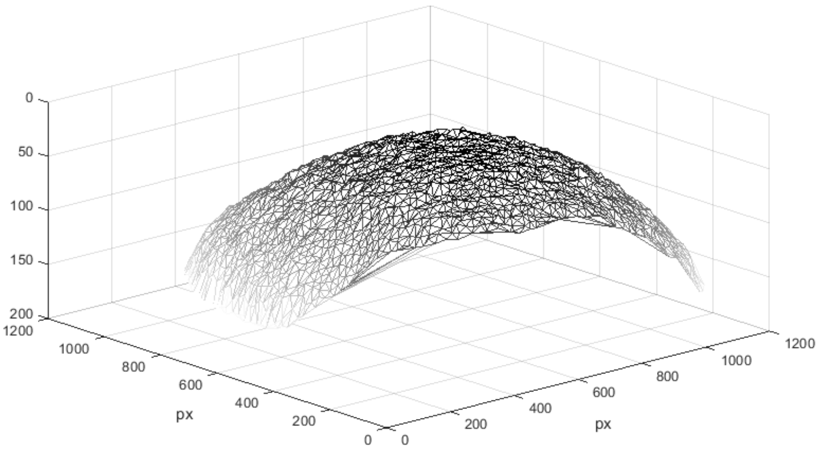

Figure 4.15: Visualization of the Delaunay triangulation of the hemisphere surface of experiment ZF1L. The triangles were used to calculate the area of the surface on which the cells were located.

We first used Ripley's K to analyze whether the locations of stem cells on ZF1L were following a CSR pattern. We thus compared Ripley's K of all stem cells to CSR distributed points on an area with equal size as the area determined by triangulation. It can be seen in Figure 4.16A that the difference in $K(d)$ is only barely visible due to the large range of the Y axis. But we can calculate $L(d) = (K(d)/\pi)^{(}1/2)$ [30] as $var(L(d))$ is approximately constant under CSR and under CSR $L(d) - d = 0$. Looking at $L(d) - d$ two departures of CSR can be observed (Figure 4.16B). The difference for small d values can be explained by the cell diameter around $4\mu m$ and therefore constituted the minimal observed distance between cells. The deviation for $d > 10$ indicated a difference between the stem cell distribution and completely spatial randomness. To take this uneven distribution of cells into account, we used the stem cell locations as possible points where divisions could appear for further calculations. Additionally, since the $L(d) - d$ function was way more practical to compare distributions than the $K(d)$ function we restricted the following analyses to the $L(d) - d$ function.

We first tried to optimize the parameter of the attractive and repulsive sampling (see Section 4.4.1) for dividing stem cells and all dividing cells (see Figure 4.17). We evaluated several influence strength and radius combinations (s, r) to detect the one with the lowest root mean square distance

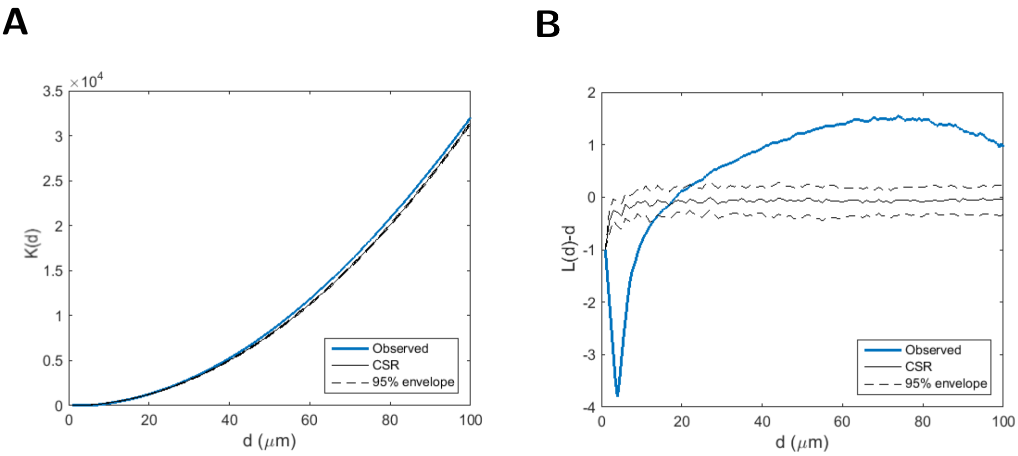

Figure 4.16: Ripley's K and $L(d)-d$ function performed on all stem cells of ZF1L (blue line) compared to completely random distributed locations with the mean as black line and the 95% envelope as dashed lines. **(A)** Ripleys'K compared to a completely random distribution of the same number of points can not be distinguished on the $K(d)$ plot but comparing $L(d) - d$ **(B)** a deviation is visible. The small observed distance (< 2) is huge enough to state that the stem cells are not completely spatially random distributed.

(RMSD) between sampled mean $K(d)$ values and observed ones:

$$RMSD = \left(N^{-1}\sum_{d}(K_s(d) - K_o(d))^2\right)^{0.5} \tag{4.4}$$

with $K_s(d)$ being the sampled K function for distance d, $K_o(d)$ the observed K function for distance d and N the maximal considered distance.
The RMSD for attractive sampling for all dividing cells showed no clear tendency in one direction but small values above certain r, s combinations. (Figure 4.17A). The optimal parameters used for the simulation were $s=$ 2.1 and r 21 μm. RSDM optimized for repulsive sampling on all dividing cells showed a slight drift towards small radii and small intensities, but the observed RMSDs were about twice as high as for the attractive sampling (Figure 4.17B). The resulting chosen parameters were $s=$ 2 and $r=$ 6 μm. Attractive sampling on dividing stem cells showed a tendency for high radii combined with high intensity (Figure 4.17C). The chosen parameters were $S=$ 3.8 and $R=$ 80 μm. The optimization for repulsive sampling on dividing stem cells revealed no trend and no such small RMSDs as for the attractive sampling (Figure 4.17D). The parameter combination with the lowest RMSD within the sampled grid were $S = 3.3$ and $r = 5$ μm. For the following simulations we used these parameters.

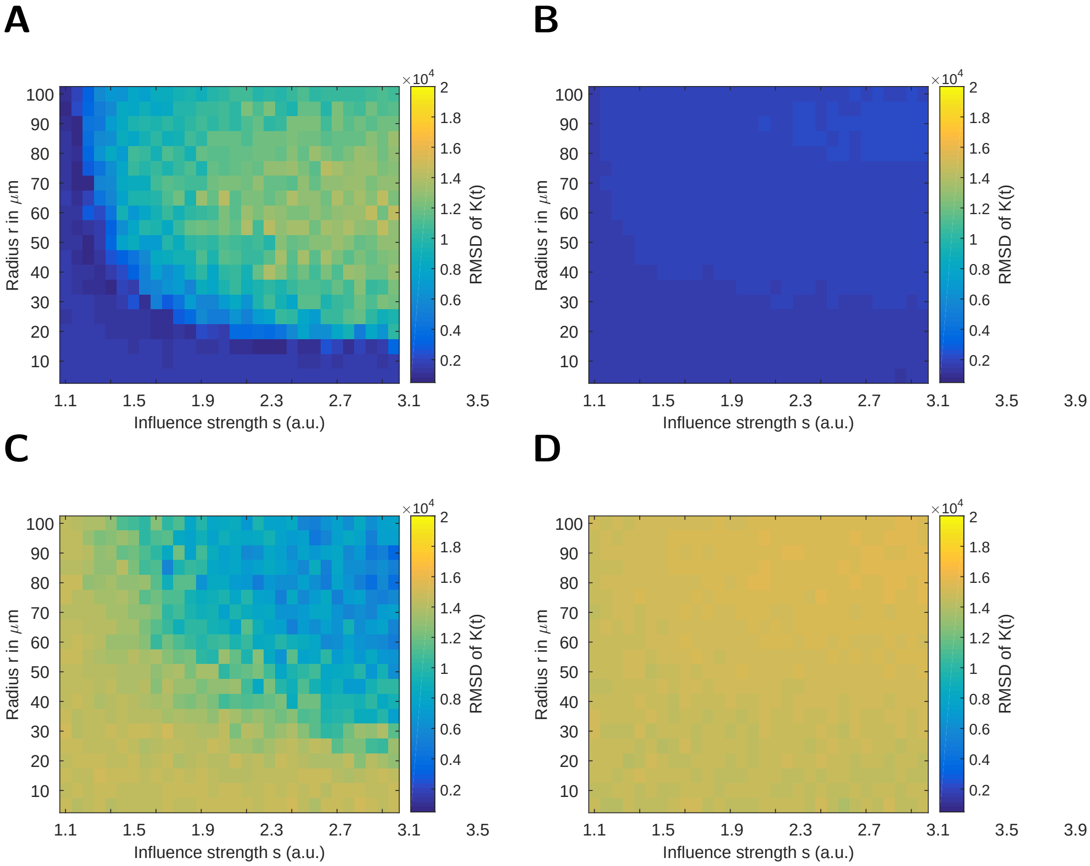

Figure 4.17: Optimizing sampling parameters.**(A)** The root mean square deviation (RMSD) from the observed $K(d)$ values is shown for attractively sampled $K(d)$ values with different influence strength (s) and influence radii (r) for all dividing cells. No clear tendency in one direction was visible. The chosen parameters were $s=$ 2.1 and $r=$ 21 μm. **(B)** Repulsive sampling on all dividing cells showed a slight RMSD drift towards small radii and small intensities, but the observed RMSDs were about twice as high as for the attractive sampling. The resulting parameters were $s=$ 2 and $r=$ 6 μm. **(C)** Attractive sampling on dividing stem cells show a tendency for high radii combined with high intensity. The resulting parameters were $s=$ 3.8 and $r=$ 80 μm. **(D)** The optimization for repulsive sampling on dividing stem cells revealed no trend and no such small RMSDs as for the attractive sampling were observed. The parameter combination with the lowest RMSD were an intensity of 3.3 and a radius of 5 μm.

The $L(d) - d$ function of the observed dividing stem cells clearly differs from the CSR as well as from repulsive sampling as the curve is located completely above the 0.95% envelopes from the samplings (Figure 4.18A). Compared to the attractive scenario the curve cuts the envelope in the region between 5 and 25 μm. Comparing the CSR and repulsive sampling $L(d) - d$ function of all dividing cells to the observed pattern we observed no cut of the 95 % envelopes (Figure 4.18B). The 95% envelope of attractive sampling in contrast included the complete observed curve. The huge span of the envelope of attractive sampling was probably the result of the possible chain effect that could arise. One cell divides and the region around it granted twice the probability for further divisions. If another cell divides nearby the regional probability doubles again. This can lead to a huge cell cluster in a small region. When the radius around one cell reaches this region the K(d) value increases strongly.

The same analysis performed on ZF2L showed similar results as ZF1L for CSR and repulsive sampling for dividing stem cells (Figure 4.19A) as the observation curve was located completely above both 95% envelopes. Comparing the attractive scenario to the observed pattern we saw a difference to ZF1L as the $L(d) - d$ function of all dividing cells cut the envelope of the attractive sampling $L(d) - d$ function in the regions from 5 to 70 μm with small peak outside the envelope around 8 μm. The $L(d) - d$ function of all dividing cells lay again nearly completely inside the 95% envelope (Figure 4.19B). Only the small intervals between 0 and 8 μm as well as between 18 and 22 μm were found outside it. Around 50 μm the curve even reached the middle of the envelope and stayed there. Compared CSR and repulsive sampling the observed $L(d) - d$ function was located clearly above their 95% envelopes, respectively.

Ripley's K performed on ZF2R showed different results compared to ZF1L and ZF2L for CSR and repulsive sampling for dividing stem cells (Figure 4.20A) as the $L(d) - d$ curve cuts both 95% envelopes between 15 and 60 μm. Compared with attractive sampling we observed a small cut of the 95% envelope between 10 and 32 μm. The $L(d) - d$ function of all dividing cells was found again nearly completely inside the 95% envelope for attractive sampling (Figure 4.20B). Only the small interval between 5 and 8 μm was found outside. Surprisingly we observed also a long cut between the observed $L(d) - d$ function and the CSR 95% envelope. Starting at 20 μm the envelope covered the curve except for two small intervals between 27 and 28 μm and 35 and 42 μm. The 95% envelope of repulsive sampling instead cut the observed curve only in small regions between 0 and 4 μm, which was negligible as in this region no neighboring cell could appear. Noticeable here is the form of the CSR curve as it raises straight up to a $L(d) - d$ value of 10 instead

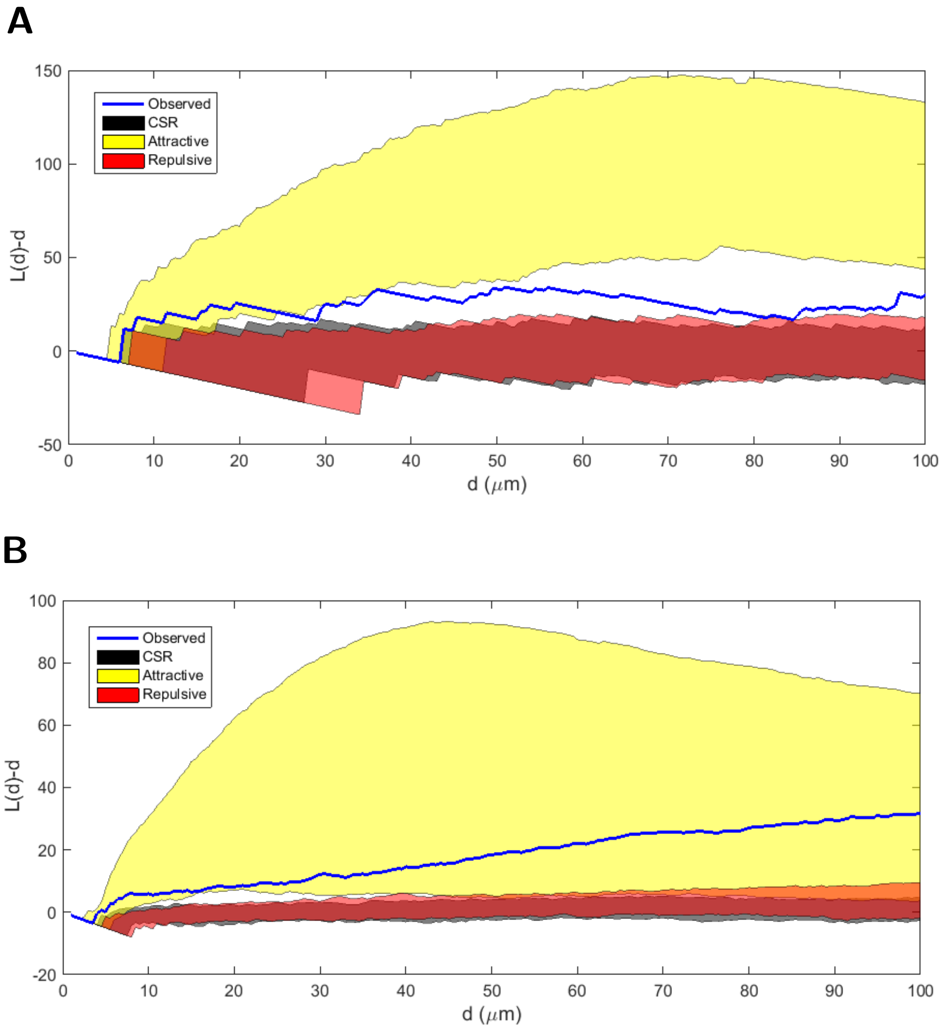

Figure 4.18: The $L(d) - d$ function (blue line) for ZF1L dividing stem cells and all dividing cells compared to three sampled patterns. The different colored patches represent the 95% envelope for the three compared pattern. **(A)** The $L(d) - d$ function of the dividing stem cells differs from the CSR and the repulsive sampling as the curve is located completely above the 0.95% envelopes. Compared to the attractive $L(d) - d$ function the curve cuts the envelope in the region between 5 and 25 μm. **(B)** Comparing the CSR and repulsive sampling $L(d) - d$ function of all dividing cells to the observed pattern we observed no cut of the 95 % envelopes with the observed curve. The 95% envelope of attractive sampling in contrast completely includes the observed curve.

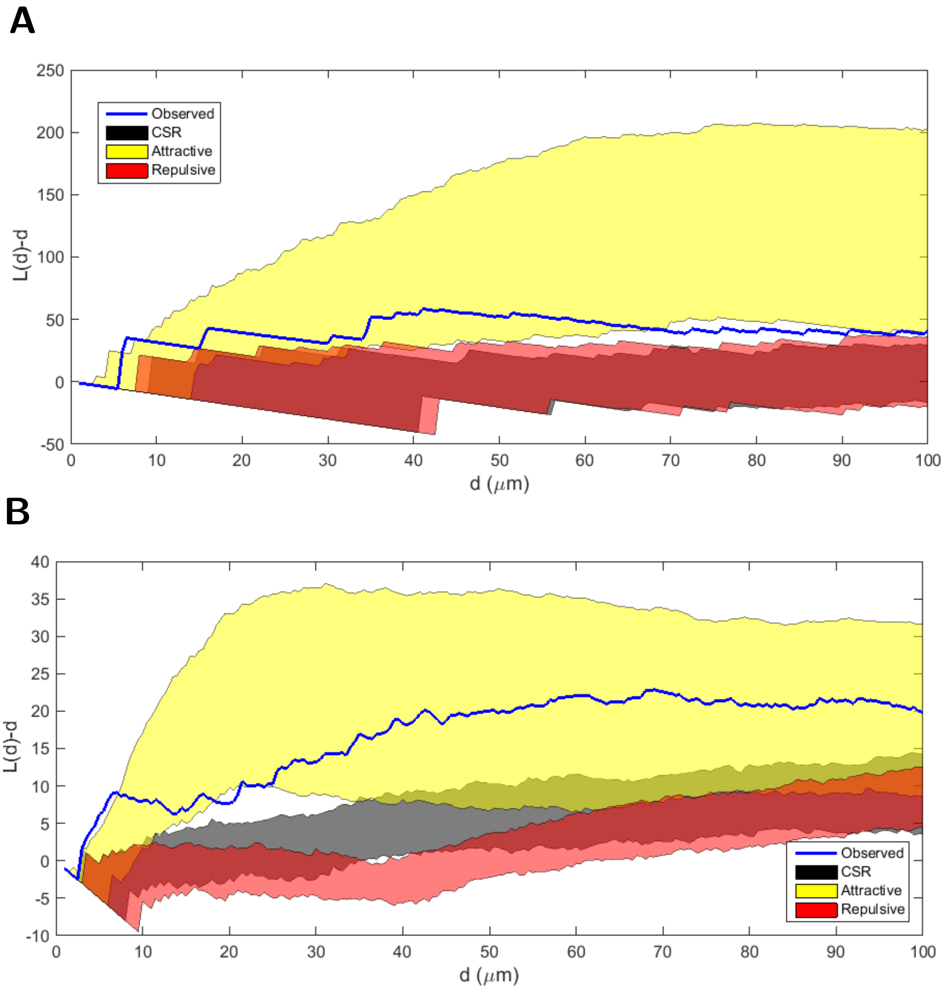

Figure 4.19: The $L(d) - d$ function (blue line) for ZF2L dividing stem cells and all dividing cells compared to three sampled patterns. The different colored patches represent the 95% envelope for the three compared pattern. **(A)** The $L(d) - d$ function of the dividing stem cells differs from the CSR and the repulsive sampling as the curve is located completely above the 0.95% envelopes. Compared to the attractive $L(d) - d$ function the curve cuts the envelope in the region between 5 and 8 μm and from 10 to 70 μm . **(B)** Comparing the CSR and repulsive sampling $L(d) - d$ function of all dividing cells to the observed pattern we observed no cut of the 95 % envelopes with the observed curve. The 95% envelope of attractive sampling in contrast includes the nearly the complete observed curve (except the region between 18 and 22 μm.

staying around 0. This could be an effect of the overall cell distribution on the hemisphere in experiment ZF2R.

In summary we can conclude that only for all dividing cells a consistent result for all three experiments was obtained. We found that CSR and repulsive sampling was unlikely, as the $L(d) - d$ functions for observed values were never inside the respective 95% envelopes, while the 95% envelopes of attractive sampling often covered the $L(d) - d$ function for observed values. Dividing stem cells in contrast behaved completely different experiment-wise as their $L(d) - d$ curve cut the attractive scenario 95% envelope in all three experiments while it cut also the CSR and repulsive sampled 95% envelope in ZF2R. Additionally, as we obtained only small optimal radii (5 and 6 μm) for repulsive sampling, we observed similar curves for repulsive sampling and CSR.

4.4.5 Active hemisphere regions

Ripley's K describes the whole data in terms of statistical pattern properties. To visualize these features on the images we implemented a method that detects and displays regions of the hemisphere that were more active than one would expect at random (see Section 3.4.5). Specifically we calculated an empirical p-value for the presence of dividing cells for 256 sub-images of the image and visualize them using a contour plot (see Figure 4.21A,B).

On the left side of the ZF1L hemisphere low p-value regions arose. Figures 4.21C-F display that both other experiments show significantly lower empirical p-values as expected at the medial sides. The medial side is the side of the hemisphere that is near to the other hemisphere and it is assumed that the medial site of the hemisphere is more active than the dorsal one (personal communication). Our results showed that the cell location probably also impacts the division rate.

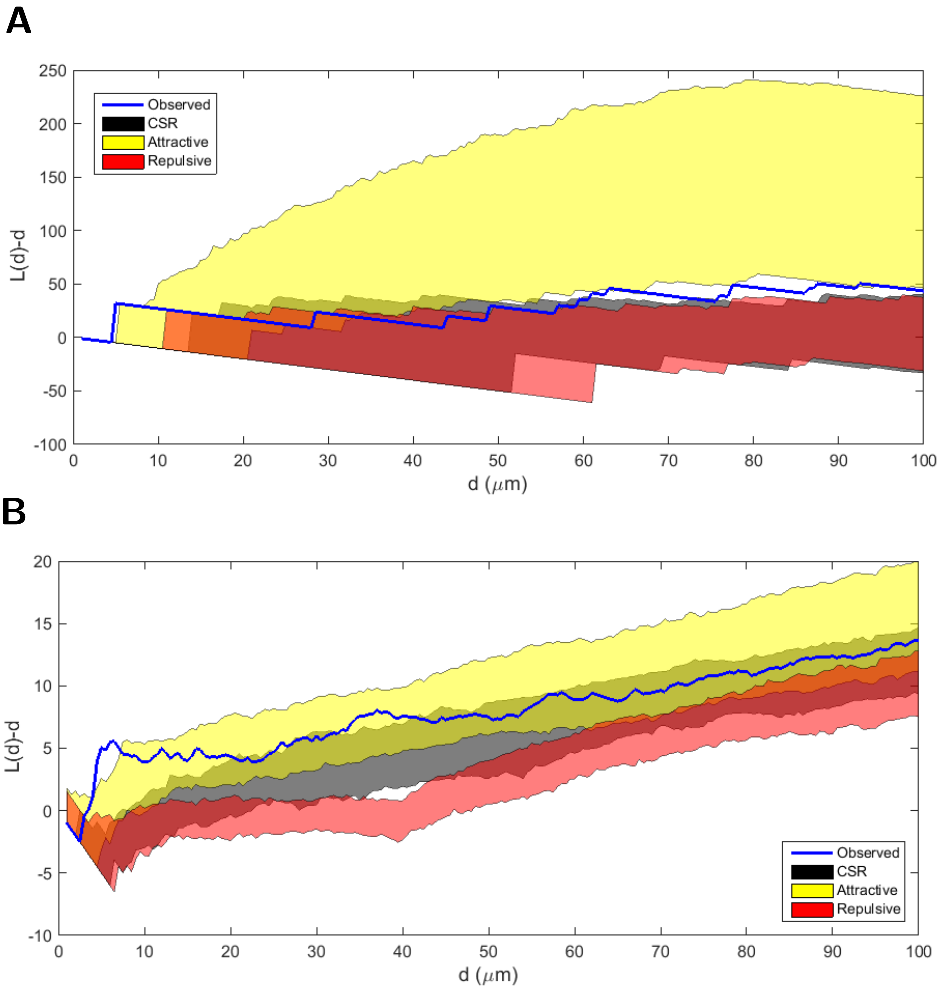

Figure 4.20: The $L(d)-d$ function (blue line) for ZF2R dividing stem cells and all dividing cells compared to three sampled patterns. The different colored patches represent the 95% envelope for the three compared pattern. **(A)** The $L(d)-d$ function of the dividing stem cells cuts the CSR and the repulsive sampling 95% envelope between 15 and 60 μm. Compared to the attractive $L(d)-d$ function the curve cuts the envelope in the region between 9 and 32 μm. **(B)** Comparing the repulsive sampling $L(d)-d$ function of all dividing cells to the observed pattern we observed only a small cut of the 95 % envelopes with the observed curve between 0 and 5, while we observed two small cuts with the CSR sampled $L(d)-d$ function between 20 and 35 μm and a long cut starting at 40 μm. The 95% envelope of attractive sampling in contrast includes nearly the complete observed curve (except for the region between 5 and 8 μm).

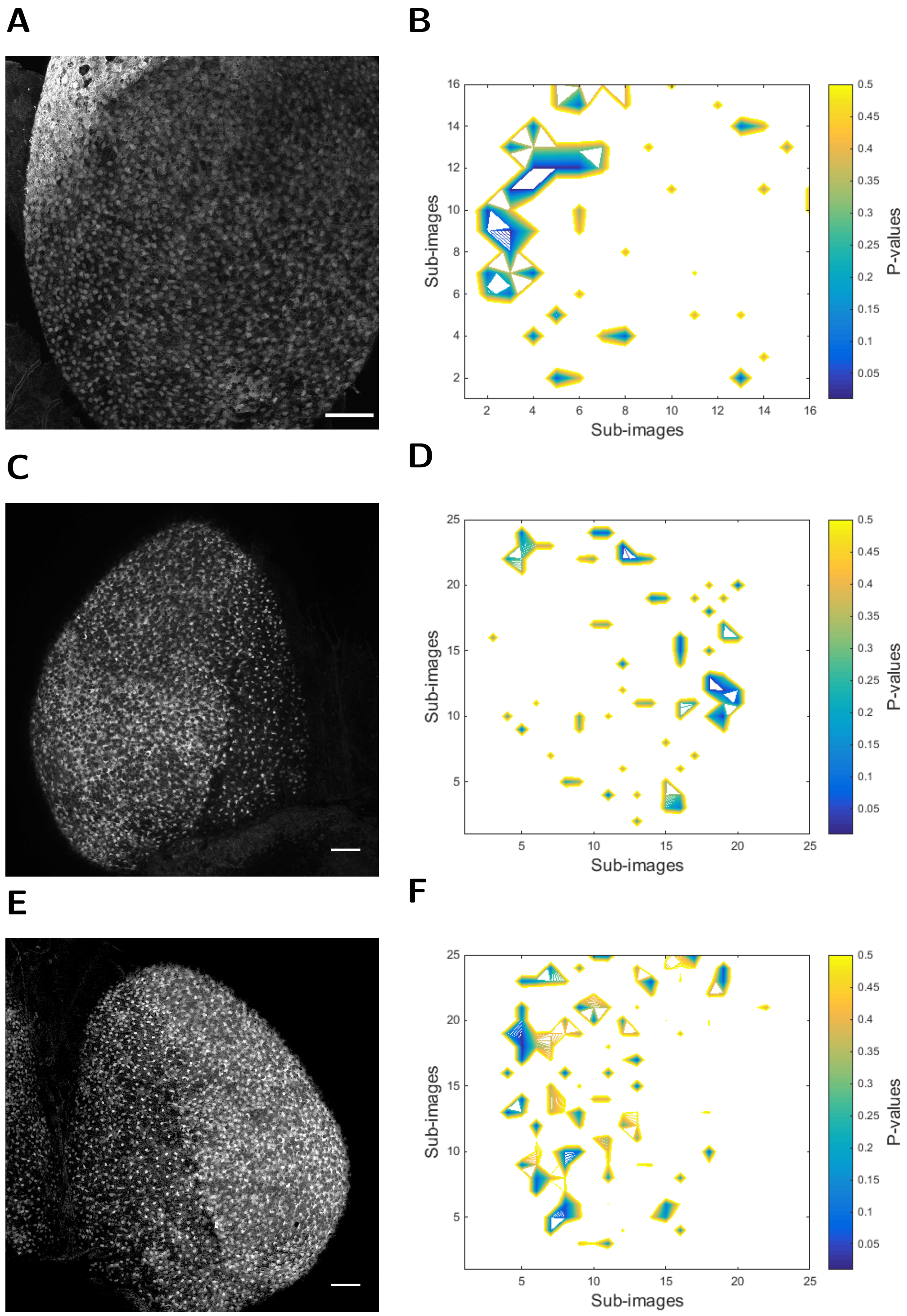

Figure 4.21: (Caption next page.)

Figure 4.21: (Previous page.) Contour plot showing empirical p-values for cell divisions in sub-images (frames). The number of local divisions was compared to randomly sampled divisions to reveal active regions of the hemisphere. **(A)** Raw image of experiment ZF1L. **(B)** Inn experiment ZF1L on the medial (here left) side clearly visible low p-value regions were obtainable. **(C)** Raw image of experiment ZF2L. **(D)** In experiment ZF2L on the medial (here right) side clearly visible low p-value regions were obtainable. **(E)** Raw image of experiment ZF2R. **(F)** On experiment ZF2R the medial (here left) side clearly visible low p-value regions were obtainable. Scale bars: 50 μm

5 Discussion & Conclusion

Neural stem cells (NSCs), especially NSC divisions and differentiation, play an important role in the treatment and prevention of neurological disorders [33, 34, 35]. We focused in this thesis on detecting spatial patterns of NSC divisions in the zebrafish brain. Our aims were to identify single cells on zebrafish hemispheres surface and use statistical methods, on surface dependent distances between dividing cells, to conclude spatial patterns of dividing cells. The underlying data was provided by the Zebrafish Neurogenetics Department at the Helmholtz Zentrum München.

We developed a cell identification pipeline based on the segmentation of the maximum intensity z-projection of the image stack. The identified 2D cell centroids were utilized to determine the cell positions in 3D using intensity profiles in z direction. The pipeline outperformed an automatic 3D approach by identifying about 60% more cells correctly. We identified stem cells, dividing progenitor cells and dividing stem cells on three hemispheres. The cell numbers were comparable between the different experiments except for the count of dividing cells in one experiment, which was about twice as much as in the two other hemispheres. This huge number could be explained by a short active phase of the hemisphere in which many divisions appeared, since EdU marks only cells in a small temporal interval [36]. Additionally we detected already divided cells, so called doublets, of which 86% divided symmetrically and 14% asymmetrically. We inferred the average length of the NSC S-Phase transferring the proportion of stem cell doublets observed by Barbosa et al. (unpublished data) to our data. The resulting lengths of 8h-10h correspond to the results from Dekens et al. [37].
We used the stem cell locations to estimate the hemisphere surface and compared the locations of dividing cells to it. Observed dividing progenitor cells were located predominately below the hemisphere, probably moving into the brain to further differentiate [32]. Even dividing stem cells were located slightly below the surface. We guess that they possibly need space to divide. It is possible that with increasing image quality a 3D approach would become more practical than our implemented two step method. To increase quality we could consider using light sheet-based fluorescence microscopy (LSFM) for reduced photo-bleaching and photo-toxic effects compared to any other form of microscopy [38]. Using a direct 3D segmentation could determine the 3D structure of each cell, which would allow further investigations by

mapping cell shapes to different cell types.

In the next step we analyzed spatial features of NSC divisions and divisions in general. Our nearest neighbor approach revealed that dividing NSCs and all dividing cells form cluster on all three experiments but all dividing cells also include several spatially isolated divisions. The F-function showed that dividing NSCs and all dividing cells differ from completely spatial randomness and tend to form attractive pattern. Ripley's K evaluated on a surface with non-uniform cell appearance probabilities in contrast indicated an attractive pattern only for all dividing cells, in which dividing cells have an influence radius of 20 μm ($\sim$ 4 cell diameters). Ripley's K found no clear spatial organization of dividing stem cells, which could be the consequence of having only few divisions per experiment. We could extend the pattern simulation for more complex features like combining attractive and repulsive behavior for different distances. Additionally we could adapt the simulation to include explicit biological interaction models like delta-notch signaling [39]. In summary the results conflict with the hypothesis of a repulsive pattern from Chapouton et al. [18] as the repulsive pattern was refused by all applied methods.
We also could consider other reasons for an observed attractive pattern than the influence of a dividing cell on other cells. Theoretically there could be a sub-population of strongly dividing cells where a cell divides and afterwards their daughter cells divide and their daughters as well. The result would be an attractive pattern as observed, since sister cells would divide approximately at the same time when they show similar generation time. This would also explain the observed single divisions as they could represent the first dividing generation. This scenario could be tested with two staining time points approximately one generation apart or by continuous time lapse measurements with appropriate markers.

An interesting observation by Barbosa et al. (unpublished data) is the possibility that stem cells directly differentiate into progenitor cells without prior division. From our perspective it would be interesting to test whether NSCs need to rearrange their DNA in S-Phase before the conversion. To identify such a direct conversion we need the measurement of two time points: At the first time point we need to observe a dividing stem cell, while at the second time point the cell should be marked only as dividing cell without performing a division. If the two time points are far enough from each other so that we can be sure that the cell will not divide at all we would have observed such a direct conversion.

It would especially be interesting to analyze and compare cell division patterns during aging. NSCs of young and old animals could have diverging division patterns due to cellular changes or due to changed requirements of new stem cells or differentiated neural cells. In that way, our analyses can be an important contribution for understanding the parthenogenesis of neurological disorders.

Acknowledgement

Thanks to Dr. Carsten Marr for continuous support and motivational feedback. Thanks to Felix Buggenthin for his advice especially in sharing his knowledge about image processing. Thanks to Prof. Dr. Dr. Fabian Theis for giving me the opportunity to do this work on this interesting topic at the ICB.

References

[1] Paul R Sanberg, David J Eve, L Eduardo Cruz, and Cesar V Borlongan. Neurological disorders and the potential role for stem cells as a therapy. *British medical bulletin*, 101(1):163–181, 2012.

[2] Stephen W Kuffler and John G Nicholls. The physiology of neuroglial cells. *Ergebnisse der physiologie biologischen chemie und experimentellen pharmakologie*, 57(1):1–90, 1966.

[3] Diana L Clarke, Clas B Johansson, Johannes Wilbertz, Biborka Veress, Erik Nilsson, Helena Karlström, Urban Lendahl, and Jonas Frisen. Generalized potential of adult neural stem cells. *Science*, 288(5471):1660–1663, 2000.

[4] Olle Lindvall and Zaal Kokaia. Stem cells for the treatment of neurological disorders. *Nature*, 441(7097):1094–1096, 2006.

[5] Sally Temple. Division and differentiation of isolated cns blast cells in microculture. 1989.

[6] Brent A Reynolds and Samuel Weiss. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *science*, 255(5052):1707–1710, 1992.

[7] CG Craig, CM Morshead, A Roach, and D Van der Kooy. Evidence for a relatively quiescent stem cell in the adult mammalian forebrain. *J Cell Biochem*, 18:176, 1994.

[8] Juergen A Knoblich. Asymmetric cell division: recent developments and their implications for tumour biology. *Nature reviews Molecular cell biology*, 11(12):849–860, 2010.

[9] Jovica Ninkovic and Magdalena Götz. How to make neuronsthoughts on the molecular logic of neurogenesis in the central nervous system. *Cell and tissue research*, 359(1):5–16, 2015.

[10] Jovica Ninkovic and Magdalena Götz. Fate specification in the adult brain–lessons for eliciting neurogenesis from glial cells. *Bioessays*, 35(3):242–252, 2013.

[11] Nathan D Lawson and Brant M Weinstein. In vivo imaging of embryonic vascular development using transgenic zebrafish. *Developmental biology*, 248(2):307–318, 2002.

[12] Jovica Ninkovic, Tetsuji Mori, and Magdalena Götz. Distinct modes of neuron addition in adult mouse neurogenesis. *The Journal of Neuroscience*, 27(40):10906–10911, 2007.

[13] Tina Schrödel, Robert Prevedel, Karin Aumayr, Manuel Zimmer, and Alipasha Vaziri. Brain-wide 3d imaging of neuronal activity in caenorhabditis elegans with sculpted light. *Nature methods*, 10(10):1013–1020, 2013.

[14] Lin Tian, S Andrew Hires, Tianyi Mao, Daniel Huber, M Eugenia Chiappe, Sreekanth H Chalasani, Leopoldo Petreanu, Jasper Akerboom, Sean A McKinney, Eric R Schreiter, et al. Imaging neural activity in worms, flies and mice with improved gcamp calcium indicators. *Nature methods*, 6(12):875–881, 2009.

[15] Sung-Kyun Ko, Xiaoqiang Chen, Juyoung Yoon, and Injae Shin. Zebrafish as a good vertebrate model for molecular imaging using fluorescent probes. *Chem. Soc. Rev.*, 40:2120–2130, 2011.

[16] Fernando Amat, William Lemon, Daniel P Mossing, Katie McDole, Yinan Wan, Kristin Branson, Eugene W Myers, and Philipp J Keller. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nature methods*, 2014.

[17] Nikita Vladimirov, Yu Mu, Takashi Kawashima, Davis V Bennett, Chao-Tsung Yang, Loren L Looger, Philipp J Keller, Jeremy Freeman, and Misha B Ahrens. Light-sheet functional imaging in fictively behaving zebrafish. *Nature methods*, 2014.

[18] Prisca Chapouton, Paulina Skupien, Birgit Hesl, Marion Coolen, John C Moore, Romain Madelaine, Elizabeth Kremmer, Theresa Faus-Kessler, Patrick Blader, Nathan D Lawson, et al. Notch activity levels control the balance between quiescence and recruitment of adult neural stem cells. *The Journal of Neuroscience*, 30(23):7961–7974, 2010.

[19] Rebecca L Bernardos and Pamela A Raymond. Gfap transgenic zebrafish. *Gene expression patterns*, 6(8):1007–1013, 2006.

[20] Adrian Salic and Timothy J Mitchison. A chemical method for fast and sensitive detection of dna synthesis in vivo. *Proceedings of the National Academy of Sciences*, 105(7):2415–2420, 2008.

[21] P. Sarder and Arye Nehorai. Deconvolution methods for 3-d fluorescence microscopy images. *Signal Processing Magazine, IEEE*, 23(3):32–45, May 2006.

[22] William Hadley Richardson. Bayesian-based iterative method of image restoration. *JOSA*, 62(1):55–59, 1972.

[23] Leon B Lucy. An iterative technique for the rectification of observed distributions. *The astronomical journal*, 79:745, 1974.

[24] Sarah Frisken Gibson and Frederick Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *JOSA A*, 9(1):154–166, 1992.

[25] Benjamin Rappaz, Pierre Marquet, Etienne Cuche, Yves Emery, Christian Depeursinge, and Pierre Magistretti. Measurement of the integral refractive index and dynamic cell morphometry of living cells with digital holographic microscopy. *Optics express*, 13(23):9361–9373, 2005.

[26] Boris Delaunay. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7(793-800):1–2, 1934.

[27] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[28] Philippe Andrey, Kiên Kiêu, Clémence Kress, Gaëtan Lehmann, Leïla Tirichine, Zichuan Liu, Eric Biot, Pierre-Gaël Adenot, Cathy Hue-Beauvais, Nicole Houba-Hérin, et al. Statistical analysis of 3d images detects regular spatial distributions of centromeres and chromocenters in animal and plant nuclei. *PLoS computational biology*, 6(7):e1000853, 2010.

[29] Brian D Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, pages 255–266, 1976.

[30] Philip M Dixon. Ripley's k function. *Encyclopedia of environmetrics*, 2002.

[31] Anthony Christopher Davison. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

[32] Sisi Chen and David V Schaffer. Newborn neuroblasts feel the field in the adult brain. *EMBO reports*, 14(2):105–106, 2013.

[33] Olle Lindvall, Zaal Kokaia, and Alberto Martinez-Serrano. Stem cell therapy for human neurodegenerative disorders–how to make it work. 2004.

[34] Fred H Gage and Sally Temple. Neural stem cells: generating and regenerating the brain. *Neuron*, 80(3):588–601, 2013.

[35] Ferdinando Rossi and Elena Cattaneo. Neural stem cell therapy for neurological diseases: dreams and reality. *Nature Reviews Neuroscience*, 3(5):401–409, 2002.

[36] Lixin Sun, Jeongwu Lee, and Howard A Fine. Neuronally expressed stem cell factor induces neural stem cell migration to areas of brain injury. *Journal of Clinical Investigation*, 113(9):1364, 2004.

[37] Marcus PS Dekens, Cristina Santoriello, Daniela Vallone, Gabriele Grassi, David Whitmore, and Nicholas S Foulkes. Light regulates the cell cycle in zebrafish. *Current Biology*, 13(23):2051–2057, 2003.

[38] Francesco Pampaloni, Bo-Jui Chang, and Ernst HK Stelzer. Light sheet-based fluorescence microscopy (lsfm) for the quantitative imaging of cells and tissues. *Cell and tissue research*, 360(1):129–141, 2015.

[39] Sarah J Bray. Notch signalling: a simple pathway becomes complex. *Nature reviews Molecular cell biology*, 7(9):678–689, 2006.