

**HELMHOLTZ**

RESEARCH FOR GRAND CHALLENGES

**HelmholtzZentrum münchen**  
German Research Center for Environmental Health



Technische Universität München

---

Modeling spatiotemporal single cell patterns in  
tissue maintenance and repair

---

Valerio Lupperger

December 2021



TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

**Modeling spatiotemporal single cell  
patterns in tissue maintenance and repair**

**Valerio Lupberger**

Vollständiger Abdruck der von der TUM School of Life Sciences der  
Technischen Universität München zur Erlangung des akademischen  
Grades eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation.

**Vorsitzende(r):** Prof. Dr. Dmitrij Frischmann

**Prüfer der Dissertation:**

1. Prof. Dr. Dr. Fabian J. Theis
2. Prof. Dr. Christina Kuttler

Die Dissertation wurde am 29.01.2021 bei der Technischen Universität  
München eingereicht und durch die TUM School of Life Sciences am  
20.10.2021 angenommen.



## Danksagung

Zunächst möchte ich mich bei Carsten bedanken, der mich mit seiner positiven und ermutigenden Einstellung durch meine Doktorarbeit geführt hat. Besonderer Dank geht an Fabian, dessen Feedback oft neue Perspektiven eröffnet hat.

Ich möchte mich bei meinem langjährigen Bürokollegen Hans bedanken, der die Tage meiner Doktorarbeit immer bereichert hat, sei es mit entspannten Kaffeepausen oder intensiven Diskussionen über unsere Projekte. Außerdem bin ich auch sehr dankbar für die fruchtbaren Diskussionen mit den Kollegen und Freunden aus dem Marr-Lab und dem gesamten ICB.

Meinen Eltern danke ich dafür, dass sie mich immer unterstützt haben, an mich geglaubt haben und mich meinen Weg haben finden lassen.

Mein tiefster Dank gilt natürlich meiner Frau Anastasia, die mich auch in schwierigen Zeiten immer bestärkt hat und einen erheblichen Teil dazu beigetragen hat, dass ich dort stehe wo ich jetzt bin. Auch möchte ich meine zwei kleinen Wirbelwinde erwähnen, die es geschafft haben, meine Doktorarbeit noch außergewöhnlicher zu machen.

## Abstract

Understanding single cell behavior in a tissue context is pivotal for insights into tissue maintenance and repair. Spatial statistics and mechanistic modeling can help to identify cellular patterns and test hypotheses on underlying rules. The aim of this cumulative doctoral thesis was to develop, modify and apply computational methods to (i) analyze division patterns of neural stem cells (NSCs) in the zebrafish brain, (ii) deliver mechanistic insight into the regrowth of the lateral organ on the zebrafish skin, and (iii) better comprehend fibroblast movement during scar formation in mice. Each aspect was addressed in a separate study and published in a peer-reviewed article.

Firstly, we modified the Ripley's K statistic, a measure to identify spatial patterns of point processes, to a discrete setting and applied it to the division patterns of NSCs. A spatiotemporal likelihood-based model allowed us to describe the detected aggregated patterns in terms of strength and extend, and a non-spatial cell division model was used to infer cell cycle parameters. We finally fed these parameters into an agent-based model to simulate growth and maintenance of an NSC population and could show that observed cell redivisions suffice for the emergence of spatiotemporal patterns similar to the ones in experimental data.

In the second project, we used tracking data of single multipotent and progenitor cells to investigate spatial factors that are crucial for regrowing the lateral organ in zebrafish. We extracted spatial and spatiotemporal features and used random forest classifiers to predict the lineage tree of starting cells and cell types that daughter cells adopt after a division during the regeneration process. For additional insight, we determined which features contributed most to the prediction and found that distance to the organ center as well as distance to mantle cells, one particular cell type in this tissue, is crucial.

Finally, we developed a measure based on a neighborhood model to quantify single cell movement during scar formation in mice. This measure enabled the comparison of conditions by calculating movement velocity in three dimensional space. Additionally, we determine the direction of movement of individual cells in relation to neighboring cells to identify swarm-like behavior.

In summary, we have developed spatiotemporal models and modified existing ones for two- and three-dimensional single cell data with different time resolutions. Application of our methods provided novel insights into the role of single cells in NSC maintenance, during lateral organ regeneration, and in scar formation.

## Zusammenfassung

Das Verständnis des Verhaltens einzelner Zellen in einem Gewebe ist entscheidend für Einblicke in dessen Erhaltung und Reparatur. Räumliche Statistik und mechanistische Modelle können dazu beitragen, Muster zu identifizieren und Hypothesen zugrunde liegender Regeln zu testen. Das Ziel dieser kumulativen Doktorarbeit war es, Methoden zu entwickeln, zu modifizieren und anzuwenden, um (i) Organisation und Muster neuronaler Stammzellen (NSCs) im Zebrafischgehirn zu analysieren, (ii) mechanistische Einblicke in die Regeneration des Seitenlinienorgans auf der Zebrafischhaut zu bekommen und (iii) Fibroblastenbewegung während der Narbenbildung bei Mäusen besser zu verstehen. Jeder Aspekt wurde in einer separaten Studie behandelt und in einem Peer-Review-Artikel veröffentlicht.

Zunächst haben wir die Ripley-K-Statistik, ein Maß zur Identifizierung räumlicher Muster von Punktprozessen, an eine diskrete Verteilung angepasst, um räumliche Teilungsmuster von NSCs im Zebrafischgehirn zu identifizieren. Ein raumzeitliches Interaktionsmodell ermöglichte es uns, die erkannten aggregierten Muster quantitativ zu beschreiben, und ein nicht-räumliches Zellteilungsmodell erlaubte es uns Zellzyklusparameter zu approximieren. Wir haben schließlich die geschätzten Parameter für ein agentenbasiertes Modell benutzt, um das Wachstum und die Aufrechterhaltung einer NSC-Population zu simulieren. Wir konnten zeigen, dass wiederholte Zellteilungen ausreichen, um raumzeitliche Muster zu erzeugen, die denen in experimentellen Daten ähneln.

Im zweiten Projekt haben wir Tracking-Daten einzelner multipotenter Zellen und Vorläuferzellen verwendet, um räumliche Faktoren zu finden, die für die Regeneration des Seitenlinienorgans im Zebrafisch entscheidend sind. Wir haben räumliche und raumzeitliche Merkmale extrahiert und verwendeten einen Random Forest Algorithmus, um den phylogenetischen Baum der Zellen die schon zu Beginn der Regeneration vorhanden waren vorauszusagen. Außerdem sagten wir jene Zelltypen vorher, die Tochterzellen nach der Teilung während des Regenerationsprozesses annehmen. Für zusätzliche Erkenntnisse haben wir bestimmt, welche Merkmale am meisten zur Vorhersage beitragen, und herausgefunden, dass die Entfernung zum Organmittelpunkt sowie die Entfernung zu Mantelzellen, einem bestimmten Zelltyp in diesem Gewebe, entscheidend ist.

Schließlich entwickelten wir ein Maß, das auf einem Nachbarschaftsmodell basiert, um die Bewegung einzelner Zellen während der Narbenbildung bei Mäusen zu quantifizieren. Dieses Maß ermöglichte den Vergleich von unterschiedlichen Bedingungen durch Berechnung der Bewegungsgeschwindigkeit

im dreidimensionalen Raum. Wir verglichen die Bewegungsrichtung einzelner Zellen mit benachbarten Zellen um schwarmartiges Verhalten aufzudecken. Zusammenfassend haben wir raumzeitliche Modelle entwickelt und vorhandene Modelle für zwei- und dreidimensionale Einzelzelldaten mit unterschiedlichen Zeitaufösungen modifiziert. Die Anwendung unserer Methoden lieferte neue Einblicke in die Rolle einzelner Zellen bei der Aufrechterhaltung der NSC Population, während der lateralen Organregeneration, und bei der Narbenbildung.

## List of contributed articles

### Publications in the context of my doctoral studies

All results of my doctoral thesis have been published in peer-reviewed journals. These publications are listed below in chronological order.

- **Lupperger, V.**, Buggenthin, F., Chapouton, P., Marr, C. (2018). **Image analysis of neural stem cell division patterns in the zebrafish brain.** *Cytometry Part A*, 93(3), 314-322.
- Viader-Llargués, O., **Lupperger, V.**, Pola-Morell, L., Marr, C., López-Schier, H. (2018). **Live cell-lineage tracing and machine learning reveal patterns of organ regeneration.** *Elife*, 7, e30823.
- **Lupperger, V.**, Marr, C., Chapouton, P. (2020). **Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatio-temporal patterns.** *PLoS Biology*, 18(12), e3000708.
- Jiang, D., Christ, S., Correa-Gallegos, D, Ramesh, P., Gopal, S. K., Wannemacher, Mayr, C. H., **Lupperger, V.**, J, Yu, Q., Ye, H., Mück-Häusl, M., Rajendran, V., Wan, L., Liu, J., Mirastschijski, U., Volz, T., Marr, C., Schiller, H. B., Rinkevich, Y. (2020) **Injury triggers fascia fibroblast collective cell migration to drive scar formation through N-cadherin** *Nature Communications*, 11(1), 1-13.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Patterns . . . . .	1
1.2	Spatiotemporal pattern analysis in cell biology . . . . .	3
1.3	Research questions . . . . .	4
1.4	Analysis workflow . . . . .	4
<b>2</b>	<b>Methods</b>	<b>11</b>
2.1	Single cell identification pipeline (SCIP) . . . . .	11
2.2	Ripley's $K$ for spatiotemporal single cell analysis . . . . .	15
2.2.1	Original Ripley's $K$ . . . . .	15
2.2.2	Spatiotemporal discrete Ripley's $K$ . . . . .	16
2.2.3	Pattern sampling . . . . .	17
2.3	Model-based quantification of spatiotemporal patterns . . . . .	20
2.3.1	Spatial interaction model . . . . .	20
2.3.2	Spatiotemporal interaction model . . . . .	21
2.3.3	Confidence measurement . . . . .	22
2.4	Likelihood-free parameter estimation . . . . .	22
2.4.1	Problem definition . . . . .	23
2.4.2	Non-spatial division model . . . . .	24
2.5	Agent-based cell population model . . . . .	25
2.6	Neighborhood coordination model . . . . .	26
<b>3</b>	<b>Summary of contributed articles</b>	<b>29</b>
3.1	Image analysis of neural stem cell division patterns in the zebrafish brain . . . . .	29
3.2	Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatiotemporal patterns . . . . .	31

3.3	Live cell-lineage tracing and machine learning reveal patterns of organ regeneration . . . . .	32
3.4	Scar formation is driven by N-Cadherin dependent collective fibroblast migration . . . . .	33
<b>4</b>	<b>Discussion &amp; future perspectives</b>	<b>35</b>
4.1	Spatiotemporal analysis in single cell biology . . . . .	35
4.2	Spatial analysis in medical research . . . . .	39
4.3	Methodological perspectives . . . . .	40
	<b>Bibliography</b>	<b>43</b>
<b>A</b>	<b>Image Analysis of Neural Stem Cell Division Patterns in the Zebrafish Brain.</b>	<b>59</b>
<b>B</b>	<b>Live cell-lineage tracing and machine learning reveal patterns of organ regeneration.</b>	<b>69</b>
<b>C</b>	<b>Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatio-temporal patterns.</b>	<b>93</b>

# Chapter 1

## Introduction

### 1.1 Patterns

Interactions shape our everyday life. Reoccurring interactions lead to patterns and embedding these interactions into time and space give rise to spatiotemporal patterns. These patterns contain an enormous amount of information, which is often not obvious, but it is worth searching and interpreting them. Back in 1854 during a cholera outbreak in London, the physician John Snow analyzed the locations of cholera cases in the city. His analysis showed a cluster of cases in Broad Street (see Figure 1.1) through which he identified a contaminated water pump in the center of this cluster [1]. The deactivation of the pump was widely recognized as the end of the outbreak.

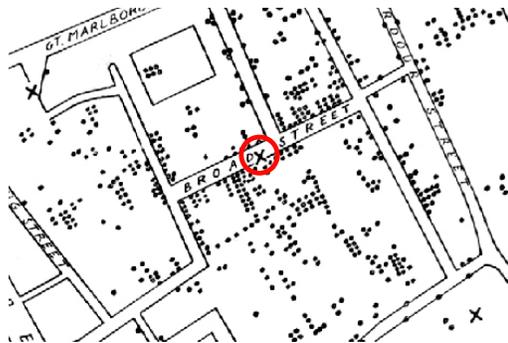


Figure 1.1: Map of cholera deaths (black dots) in London 1854. The contaminated pump in the middle of a cluster of cholera death cases is marked with a red circle. Image taken from [https://de.wikipedia.org/wiki/John\\_Snow\\_\(Mediziner\)](https://de.wikipedia.org/wiki/John_Snow_(Mediziner)) (15.09.2020)

During the coronavirus pandemic that started at the end of 2019, it was crucial to analyze the spreading of the virus not only from a biological point of view but also from a spatial perspective. Infection patterns were for example analyzed at a restaurant outbreak [2]. The study concluded that the tables were distant enough ( $>1$  meter) to allow only for a minor transmission. However, the dispersion supported by the air conditioner resulted in an infection of nine other guests along the airflow. Another infection pattern resulted from a choir practice in the State of Washington [3]. There, one so-called super spreader distributed the disease through aerosol transmission to neighboring singers, whose chairs were arranged with 15-25 cm distance. The transmission was particularly supported by singing through which the range of aerosol distribution increases.

Besides analyzing virus spreading on a local (person to person) level, in a pandemic it is pivotal to look at global patterns [4]. To stop or contain the global dispersion it is crucial find out which measures are most suitable for this task. While border control policies slowed the spreading of the virus, they did not suffice to contain the global expansion [5].

Even more interesting is the consideration and analysis of spatiotemporal patterns that appear during dynamical processes. Common examples are the weather forecast [6], traffic predictions [7, 8], or the location information on our phones that creates a large amount of spatiotemporal data that can be collected and interpreted [9, 10]. For fighting the Corona pandemic, spatiotemporal analyses and according intervention could help to control and contain the virus. One approach involves the tracking of people's locations through an app. This allows to quickly identify contact persons of a newly detected virus carrier and inform or quarantine contacts that were at the same locations as the carrier for more than ten minutes [11, 12]. One of the first countries to implement this strategy is Germany where the app has already over 22 million downloads (As of 13.11.2020). For this app, it could be enough to identify persons that were close to a virus carrier. However, for more complex settings including e.g. spatiotemporal movement patterns relevant information is often hidden and not obvious. Statistical methods are able to reveal this information.

Spatiotemporal patterns emerge on all scales, from the emergence of pathogens and diseased people, over human activity on a map [9, 10] to the clustering of species in ecological systems [13]. Their analysis often allows to draw conclusions about underlying system mechanisms. In particular for biological systems, where many cellular and molecular mechanisms are still unknown, quantitative statistical analyses of patterns have been successful in the past [14, 15].

## 1.2 Spatiotemporal pattern analysis in cell biology

Spatial and spatiotemporal pattern analysis of cells in a tissue setting is an important factor as pattern appearance can be observed in adult tissue as well as early development [16, 17].

All animals start as a single cell. From this cell, a complex developmental process leads to the final organism with about  $10^{13}$  cells [18] and multiple cell types. Tissue development and maintenance relies on stem cells able to develop into any other cell type, a process called differentiation [19]. Stem cells can also renew themselves producing new stem cells through divisions [20]. From stem cells more specialized cells arise like mature blood, neural, or muscle stem cells [21]. One factor that determines if, when and how stem cells differentiate is the so-called niche. A niche is the spatial microenvironment that can influence differentiation as well as support stem cell maintenance [22]. Within such a niche spatial features play an important role: Stem cells interact with their direct neighbors, like other stem cells, differentiated cells or dedicated niche cells [23]. In rats for instance, it could be shown that cell-cell contact is crucial for cell differentiation in connective tissue [24]. Another important differentiation factor is the interaction of stem cells with the so-called extracellular matrix, i.e. molecules and structures filling the space between cells. This interaction seems to play a major role in regulating stem cell differentiation in skin [25] and embryonic stem cells [26]. In conclusion, spatiotemporal analyses are inevitable to shed light on cell-cell interactions and cell decision making.

To study such fundamental aspects of stem cell properties model organisms are used. Complex processes in human tissue are thought to be simpler there and experimentally accessible [27]. One model organism stands out for *in vivo* analysis of embryonic and organ development: The zebrafish (*Danio rerio*) has morphological and physiological similarity to mammals [28] and convenient features such as high fertility and transparent embryos and larvae. This makes it a perfect model organism in drug discovery and toxicity testing [29] and tissue regeneration [30]. Regeneration in the zebrafish can be examined in many different tissues including heart and brain [31, 32, 33]. Brain tissue in particular can easily be analyzed *in vivo* due to its transparency [34]. It is even possible to track neural stem cells (NSCs) over time and quantify their regeneration and differentiation potential [35].

However, little is known about spatial arrangement of NSC divisions. To address this point we analyzed images of zebrafish brains in single cell resolution. We also took this analysis one step further and modeled the maintenance of the NSC population by recreating the observed patterns artificially.

To see how a system reacts to disturbances, we investigated images of the regeneration of the lateral organ (neuromast) in the zebrafish after ablation to only 4-6 cells. The question is if one can identify spatial features, which are crucial to determine the regeneration outcome. A less drastic intervention into a cell population is a lesion which triggers a spatial response of cells surrounding the wound. The open question here is how coordinated is the movement of cells in response to the injury.

### 1.3 Research questions

Issues raised at the end of the previous paragraph are important for the respective scientific field and can be tackled with spatiotemporal methods. Within the scope of this cumulative thesis the following research questions are addressed to generate quantitative insights into cell population maintenance, population regeneration, and cell movement coordination.

- I. How can we identify and model spatial and spatiotemporal patterns of single cell division locations?
- II. What spatial features provide insight into the mechanism behind the regeneration of zebrafish neuromasts?
- III. How can we appropriately quantify whether cells move in a spatially coordinated fashion within a restricting tissue?

### 1.4 Analysis workflow

To answer the research questions raised in the previous chapter with spatial and spatiotemporal methods we followed a general analysis workflow (Figure 1.2).

#### Raw image data

The road starts from raw image data (Figure 1.2A). Investigating spatial and spatiotemporal neural stem cell (NSC) division patterns as well as modeling NSC maintenance (**Research question I.** in section 1.3) commences with

3D image stacks (see Figure 1.3A) of zebrafish brain hemispheres acquired at the Helmholtz Zentrum München by Prisca Chapouton in the lab of Hernan Lopez-Schier. Image slices comprise 2048x2048 pixels and were imaged with three different channels. In the first channel NSCs are captured that emit green light (*gfap:GFP*) as we use a transgenic zebrafish line [36], while in the other two channels cells are marked being in S-phase with either BrdU (cyan cells in Figure 1.3) or EdU (magenta cells in Figure 1.3) two thymidine analogs, which are incorporated into the DNA during replication in S-phase. These S-phase markers are applied with time intervals from 9h up to 3 days. Gaining knowledge about organ regeneration (**Research question II.** in section 1.3) is based on 3D movie data (see Figure 1.3B). Neuromasts were ablated with a laser to a few remaining cells (4-8). Then the neuromast regeneration was recorded over three days and images were taken every 15 minutes in the lab of Hernan Lopez-Schier. Green cells in the zebrafish neuromast videos are sustentacular cells, while mantle cells are red and hair cells show no color.

The final data set to answer the question about coordinated cell movement (**Research question III.** in section 1.3) originates from back-skin that was collected from new-born mice and cultured in a dish [37] in the lab of Yuval Rinkevich (see Figure 1.3C). Fibroblasts were labeled by crossing the

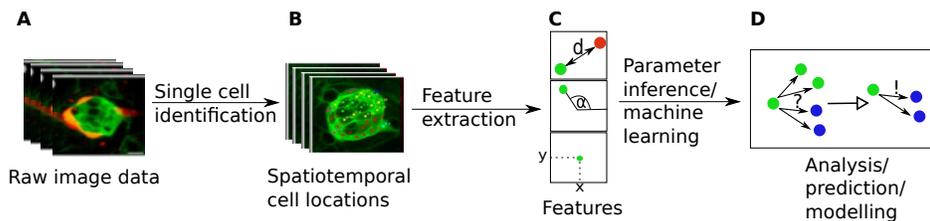


Figure 1.2: Work steps performed in this thesis, from raw (image) data to the final analysis. (A) Raw image data for a specific biological research question is collected. (B) Spatiotemporal cell locations are extracted from the raw data either manually or via cell identification algorithms. (C) Locations combined with cell phenotypes are used to extract spatial or spatiotemporal features like distances and angles between cells or cell types. (D) With these features machine learning algorithms are trained and spatial interaction parameters are inferred with mechanistic models. The gained knowledge about spatial relationships can be used for further analyses, such as predicting cell and image specific properties or the explicit modeling of spatial and spatiotemporal processes.

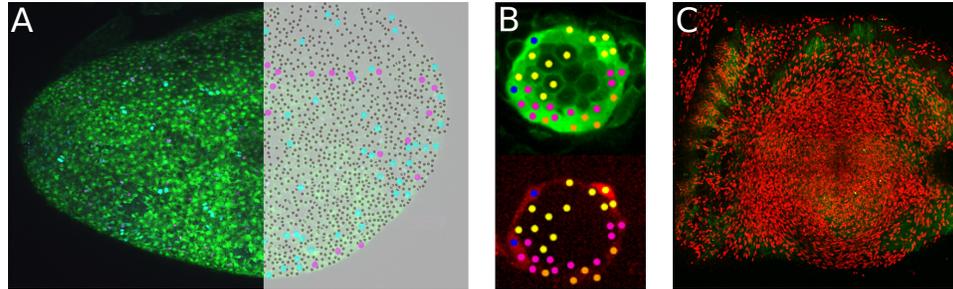


Figure 1.3: Overview of the data used in this thesis. (A) Maximum intensity projection of a zebrafish hemisphere showing the original image on the left side and identified NSCs on the right side. NSCs are green (left side) and black (right side), the first S-phase marker is cyan and the second S-phase marker is magenta. (B) Middle slice of a zebrafish neuromast stack with tracked cells on top (small colored circles). Cells that appear in the green channel (top) are sustentacular cells, while cells in the red channel (bottom) are mantle cells and cells without marker (in the middle of the neuromast) are hair cells. (C) The image shows the first frame of a movie about mouse skin tissue after injury. Red cells are fibroblasts.

fibroblast lineage-specific promoter mice to a nuclear mCherry reporter line [38] and are thus red. Fibroblast tissue is then imaged every 15 min over roughly 15 hours.

### Spatial and spatiotemporal features

To extract numbers from raw image data single cells need to be identified either manually or automatically (Figure 1.2B). Additionally, orientation in space or relationships between cells like distances and angles between them can often be meaningful (exemplary shown in Figure 1.2C). Although more complex features could add further spatial information, they are generally problem or question specific and it is important crucial to detect the correct and meaningful features as needed to answer **Research question II.** (in section 1.3). Furthermore, a temporal component can increase the feature space tremendously. Using cell movement over time is able to create so-called spatiotemporal features, such as movement direction or movement speed, which can be employed to address **Research question III.** (in section 1.3).

All spatial and spatiotemporal features mentioned above can be extracted manually for each object, but there are also methods with which features

are extracted automatically. For example, cell coordinates along with cell-specific information can be used to identify spatial features automatically, as done in a study in which neighborhood characteristics of bone marrow cells were automatically extracted from images [39]. First they segmented single cells and then used a variety of follow-up methods such as cell-specific clustering of similar neighborhoods or spatial correlation of different cell type locations with one another.

Convolutional neural networks (CNNs) can be used to extract spatial features automatically [40]. CNNs are artificial neural networks that are commonly used in computer vision to solve tasks such as image classification or object detection [41]. One recent study uses a three dimensional CNN to extract features in magnetic resonance imaging (MRI) data and automatize the diagnosis of attention deficit hyperactivity disorder (ADHD) [42]. Another medical application based on CNNs is a sleep stage analysis using spatial features from electroencephalogram (EEG) data [43], which is not image data but EEG signals that are concatenated make the input data structure image-like.

### Basic methods for spatial pattern analysis

Having determined all spatial and spatiotemporal features one needs the right follow-up methods to answer the raised research questions (Figure 1.2D). To answer the first aspect of **Research question I.** (in section 1.3) spatial and spatiotemporal patterns need to be identified. A basic way to assess spatial patterns is the usage of exploratory methods. They describe or visualize the distribution of spatial events to provide a first impression of the data and to create the basis for more complex spatial analyses. An example of such a method are Fry plots, which were originally used by Hannah and Fry to determine the spatial pattern of mineral deposits in Dyfed, Wales [44, 45]. The method takes a centered snapshot of every event (e.g. a certain mineral location) with its surrounding. Then these snapshots are merged by overlaying the respective centers and the emerging visual pattern is interpreted. Fry plots are still used e.g. to identify mineralization patterns in Iran to reveal possible mineral reservoirs [46, 47, 48].

However, more complex statistical approaches can provide even more power to analyze spatial data. One more involved methods is the Ripley's  $K$  statistic [49], where the amount of events in the neighborhood of an event is evaluated. Ripley's  $K$  has been applied in many different fields e.g. for spatial crime analysis in the city of Chicago [50], for analysis of spatial patterns in

neurons where they identified aggregation of pyramidal neurons [51] or in breast cancer research to detect abnormal areas in mammograms [52]. The open problem with aforementioned methods is that applied to cell data they would aim to analyze patterns of all cells not patterns of a subpopulation (dividing cells) in the context of the whole cell population as needed to answer **Research question I.** (in section 1.3).

### Spatial and spatiotemporal models

One approach that often provides more detailed insight into spatial problems compared to spatial statistics is the application of spatial models. Spatial models use spatial information and include additional prior information to either model spatial effects such as emerging clusters or to derive non-spatial parameters as done for instance with the spatial influence on voting outcomes in the United States [53]. Even the likelihood of nations participating in war has been described using a spatial model based on proximity to other nations that are already at war [54]. Such an approach perfectly fits the need to model observed spatial cell patterns based on problem and cell specific prior knowledge as needed to address the second aspect of **Research question I.** (in section 1.3).

Methods to analyze spatiotemporal patterns are often extensions of existing spatial methods. For example, the Ripley's  $K$  statistic can easily be extended by considering events within spatial and temporal distance [55]. This approach has been applied e.g. to investigate the spatiotemporal relationship between damage to trees by pests and later forest fires [56] and earthquake trends in Tehran [57]. Similar to spatial models, spatiotemporal models are a good way to get valuable information from spatiotemporal data sets [58]. In general, spatiotemporal models are based on spatiotemporal features such as movement vectors (e.g. for traffic analysis) or changes in certain locations over time (e.g. for weather analysis). Spatiotemporal features can be combined and used as an input for complex statistical methods like neural networks, as done for instance to predict bus arrival times [59]. Likewise, to predict the parking occupancy rate for street blocks within the next hour, machine learning models (decision trees) have been used based on spatiotemporal features [60] such as week day or occupancy rate for nearest parking street blocks. The combination of spatiotemporal features and machine learning is one possible way to address **Research question II.** (section 1.3).

Also in the case of the corona pandemic employing spatiotemporal features such as the transmission rate based on the interaction time for spatiotemporal models can improve insights into viral transmission. Researchers from Germany have published a tool that simulates the spread of the Corona virus within a small city, predicts outbreak events and evaluates containment measures based on e.g. school or inter-regional travel guidelines or hygiene parameters [61]. Such knowledge gained from spatiotemporal research enables politicians and other decision-makers to take appropriate measures [62].



## Chapter 2

# Methods

### 2.1 Single cell identification pipeline (SCIP)

Before spatial analyses can be applied to image data with single cell resolution, cellular coordinates have to be determined. This can be done manually or in an automated fashion using computational tools. However, manual identification is tedious and also very challenging in three dimensional (3D) image stacks [63]. Thus, single cell segmentation in 3D has been performed successfully using automated image processing methods e.g. in vitro cell images [64] and also in the context of challenging data sets including touching cells using machine learning [65, 66] and deep learning [33, 67]. These methods generally need a large amount of labelled training data to let the algorithm learn the right features. Whenever labelled data is missing an informed approach, using prior knowledge of the system at hand, should be used [68]. In case of neural stem cells (NSCs) in the zebrafish brain we observed that all NSCs are located on top of the brain hemisphere and form a continuous surface. Our algorithm, called single cell identification pipeline (SCIP), exploits this observation. It applies a two step approach that first identifies  $x$  and  $y$  coordinates of possible cell centroids on a two dimensional (2D)  $z$ -projection and then checks whether these candidates can be found on the NSC surface by fitting the optimal  $z$  coordinate and evaluating its distance to the surface.

SCIP is visualized in Figure 2.1. We start from a 3D image stack of a zebrafish brain hemisphere (Figure 2.1A) of which a 2D maximum intensity projection is calculated (Figure 2.1B). As cells are placed on a slightly bent surface and within the imaged field of view no or only negligible overlap is observed in  $z$  direction. Thus, a blob detection via Laplace of Gaussian

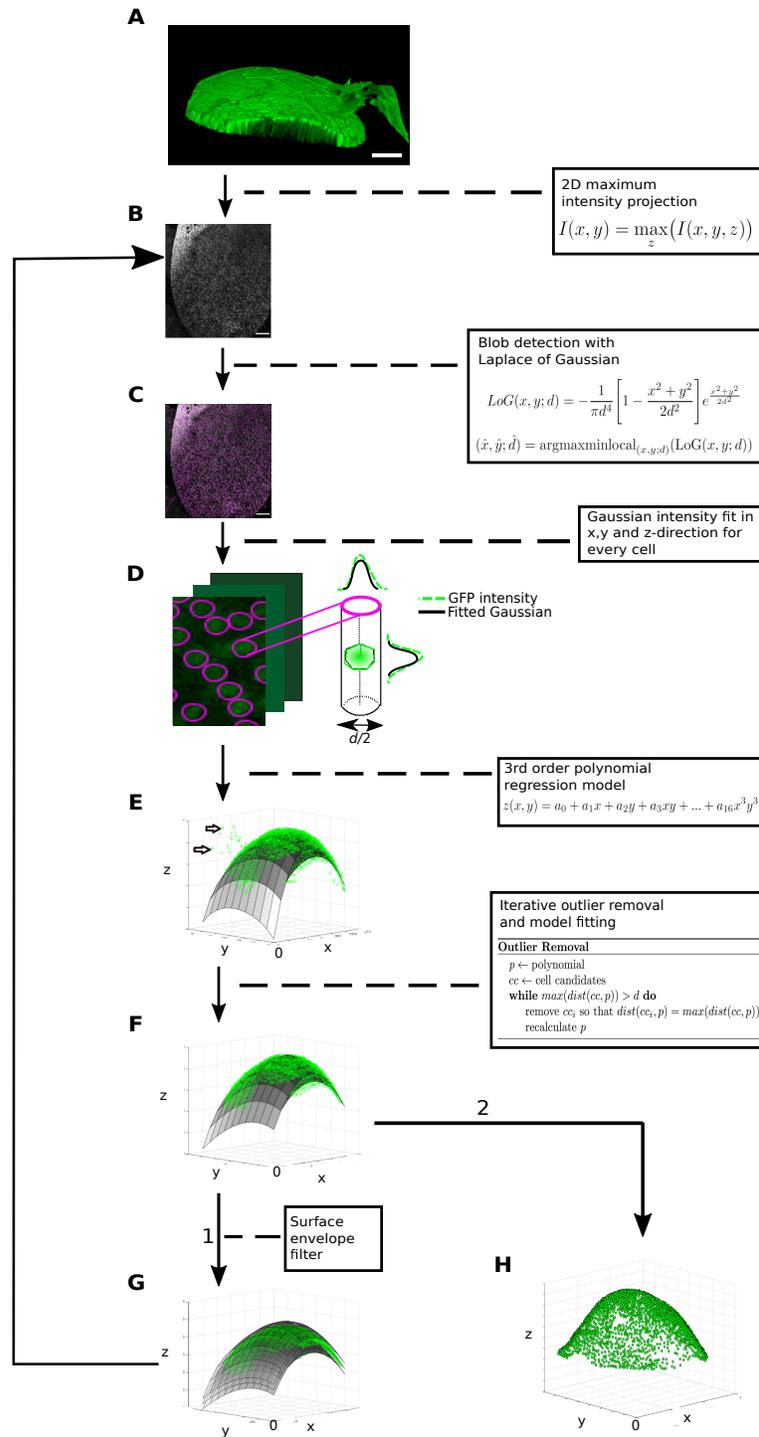


Figure 2.1: (Caption on the following page.)

Figure 2.1: Single cell identification pipeline (SCIP) for neural stem cells (NSCs) in the adult zebrafish brain. Raw three dimensional data (A) is converted into two dimensional (2D) images (B) via 2D maximum intensity projection. Cell centers display a high intensity and are used for identification. (C) A blob detection algorithm using LoG identifies cell candidates shown in magenta. (D) A Gaussian curve is fitted to the intensity profile of a cylinder with diameter  $d/2$  along  $z$  of every cell candidate. The mean of the fitted Gaussian is employed as the  $z$ -coordinate of each cell candidate centroid. (E) A 3rd order polynomial regression model is fitted to all centroids building a surface along the hemisphere top. (F) Cell candidates that are further away than approximately two cell diameters ( $2d$ ) are excluded step by step by removing iteratively the most distant outlier and recalculate the polynomial. (G) To remove remaining image artifacts an envelope is placed in  $2d$  distance around the polynomial. All pixels outside this envelope are set to background intensity. Afterwards the pipeline redoes steps (B) to (F) utilizing the filtered image stack without image artifacts. (H) The resulting three dimensional cell centroids can now be used for further analyses. Scale bars:  $50 \mu\text{m}$ .

(LoG) (Equation 2.1) followed by a local maximum/minimum search (Equation 2.2) is able to propose possible cell candidates (magenta circles in Figure 2.1C,D) with a diameter of roughly  $d = 8 \mu\text{m}$  in  $x, y$  on the  $z$ -projection.

$$\text{LoG}(x, y; d) = -\frac{1}{\pi d^4} \left[ 1 - \frac{x^2 + y^2}{2d^2} \right] e^{\frac{x^2 + y^2}{2d^2}} \quad (2.1)$$

$$(\hat{x}, \hat{y}; \hat{d}) = \text{argmaxminlocal}_{(x,y;d)} \text{LoG}(x, y; d) \quad (2.2)$$

To find the respective  $z$  coordinate of a  $x, y$  centroid, a Gaussian curve is fitted to the respective intensity profile in  $z$  direction. The fitted mean values define  $z$  coordinates for every cell candidate. To make this step more robust, the Gaussian is fitted to the mean pixel intensities of a cell sized  $(d/2)$  circular area around the center of cell candidates (Figure 2.1D). Additionally, the fitted Gaussian standard deviation can be used to filter cell candidates by their extent in  $z$ . To that end, cell candidates having a standard deviation smaller than  $d/4$  in  $z$  direction are excluded, as these cell candidates are typically image artifacts e.g. single glowing pixels. Analogously a 2D Gaussian curve can be fitted in  $x$  and  $y$  direction, where cell candidates with small

standard deviations in any dimension (below  $d/4$ ) are filtered out. The 3D coordinates of remaining cell candidates are used to fit a polynomial regression model  $p$ . As  $p$  should be flexible enough to fit the overall shape a low order (3rd order) polynomial is chosen.

$$p(x, y) = a_0 + a_1x + a_2y + a_3xy + \dots + a_{16}x^3y^3 \quad (2.3)$$

As all cells should be located on a surface, outlying cell candidates (examples marked by black arrows in Figure 2.1E) need to be removed. This is approached with an iterative process in which the cell candidate being most distant to  $p$  is removed and subsequent  $p$  is recalculated until no cell candidate is further away from  $p$  than a certain threshold (compare Figure 2.1F). In the case of zebrafish NSCs a threshold of one cell diameter ( $d$ ) is applied. In some cases, cell candidates are masked (i.e. overlaid) in the  $z$ -projection. This happens when bright image artifacts share the same  $x$  and  $y$  coordinates with a cell candidate but are located in a different  $z$  layer (see the right part of Figure 2.1A). To address this possibility, steps shown in sub-figures 2.1B-F are redone with a sub part of the image stack, which includes only the area around the hemisphere approximated by  $p$ . To this end, the original image stack is cropped by an envelope, around  $p$  with  $2d$  distance (Figure 2.1G). Image pixels outside the envelope are set to background intensity, which is calculated as the mean intensity of all eight image stack corner pixels. Finally the remaining cell candidates are returned as true NSC centroid 3D coordinates (Figure 2.1H).

## 2.2 Ripley's $K$ for spatiotemporal single cell analysis

Given a set of events  $N$  (e.g. cells of a particular tissue) with the information about their coordinates (2D or 3D) on a certain area  $A$ , one can ask the question how the events are spatially related i.e.: Do the events in  $N$  form a certain spatial pattern or are they distributed randomly? This question is relevant in many different settings as in geographic information science [55], in the context of spatial economic analysis [69] and archaeological studies, where one can e.g. assess social strategies from settlement patterning [70]. In our context of analysing zebrafish neural stem cell (NSC) divisions we apply Ripley's  $K$  to investigate spatial and spatiotemporal distributions of S-phase NSCs.

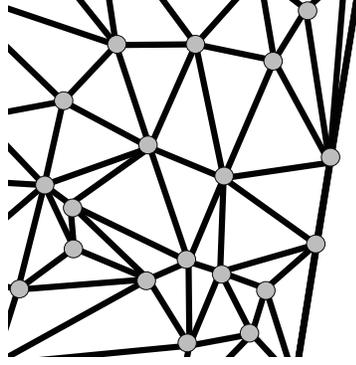
To answer the question above the Ripley's  $K$  statistic [49] can be applied. The original Ripley's  $K$  method assumes that events can happen everywhere within a fixed area  $A$ . However, in our problem setting events are not happening within an area but on a subset  $S \subset N$  of fixed possible events  $N$  that is, the coordinates of NSCs in the brain. Therefore, the original Ripley's  $K$  method needs to be adapted to fit the given task. This adaptation does not affect the calculation of the  $K$  values itself (see Section 2.2.1) but becomes important when comparing the observed  $K$  values to randomly sampled patterns (see Section 2.2.3).

### 2.2.1 Original Ripley's $K$

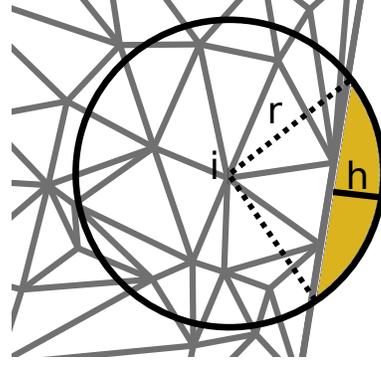
The original Ripley's  $K$  method evaluates the spatial distribution of events by counting nearby events. Events that are further away are taken into account step by step. To that end, one iterates over all events  $S$ . One starts by counting for one event  $i$  the number of other events  $j$  within a radius  $r$ . The Ripley's  $K$  value for a spatial pattern is then calculated by increasing  $r$  and is normalizing the counts by the number of events  $|S|$  and area  $A$ :

$$K(r) = \left(\frac{|S|}{A}\right)^{-1} \sum_{i=1}^S w(i, r) \sum_{j=1, j \neq i}^S \frac{I(\text{dist}(i, j) \leq r)}{|S|}. \quad (2.4)$$

Here  $i, j \in S$ , the edge correction term  $w(i, r)$  is applied to the event  $i$  and radius  $r$ , and  $I$  is the indicator function (being 1 if the distance  $\text{dist}$  between event  $i$  and  $j$  is smaller or equal to  $r$ ). Area  $A$  is determined by calculating a Delaunay triangulation [71] on  $N$  (Figure 2.2.1) and summing up the area



2.2.1: Visualisation of the triangulation result. Triangulation applied to event coordinates (gray circles) to calculate the overall area  $A$  by summing up all triangle areas.



2.2.2: Visualisation of the edge correction calculation. The term is calculated by determining the fraction of the circle with radius  $r$  around event  $i$  being outside  $A$  (yellow area) with  $h$  being the difference between  $r$  and the closest distance to the edge of  $A$ .

of all triangles. The edge correction  $w(i, r)$  for event  $i$  and radius  $r$  can be approximated by the fraction of the circle around  $i$  with radius  $r$  being inside  $A$ . This is done by determining and then subtracting the yellow area in Figure 2.2.2 from the area of the circle with radius  $r$  and dividing it by the whole circular area:

$$w(i, r) = \frac{(r^2\pi - r^2 \arccos(1 - \frac{h}{r}) - (r - h) * \sqrt{(2rh - h^2)})}{r^2\pi}. \quad (2.5)$$

Here,  $h$  is the difference between  $r$  and the shortest path to the edge of  $A$ .

## 2.2.2 Spatiotemporal discrete Ripley's $K$

With Ripley's  $K$  also spatial relationships between two sets of events, e.g. observed at different time points, can be analysed. Now, two disjoint sets of events  $S_1$  and  $S_2$  in  $N$  and the spatial influence from events  $S_1$  on the events  $S_2$  can be investigated. Since in our case the event sets occur one after the other, we call the analysis spatiotemporal Ripley's  $K$ . To this end, the spatiotemporal Ripley's  $K_{ST}$  is calculated:

$$K_{ST}(r) = \left(\frac{|S_1|}{A}\right)^{-1} \sum_{i=1}^{S_1} w(i, r) \sum_{j=1}^{S_2} \frac{I(\text{dist}(i, j) \leq r)}{|S_2|}. \quad (2.6)$$

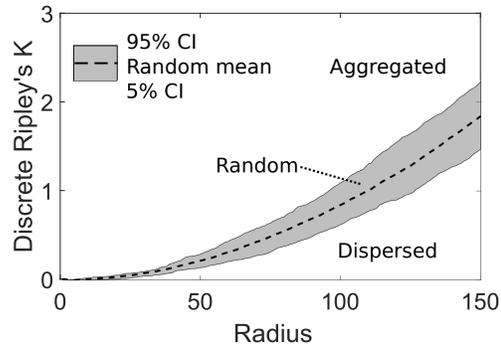


Figure 2.3: Illustration of the Ripley's  $K$  pattern interpretation. The observed discrete Ripley's  $K$  values are compared to the 5% and 95% confidence intervals. In case they are in between them, the pattern is called random, below 5% dispersed and above 95% aggregated.

The major difference to Equation 2.4 is that for every event in  $S_1$  all events in  $S_2$  within radius  $r$  are summed up. Otherwise  $K_{ST}$  is calculated analogously to the spatial  $K$  value.

### 2.2.3 Pattern sampling

Ripley's  $K$  values alone are not enough to evaluate the randomness of observed spatial or spatiotemporal patterns. To this end, the observed values have to be compared to randomly sampled  $K$  values.

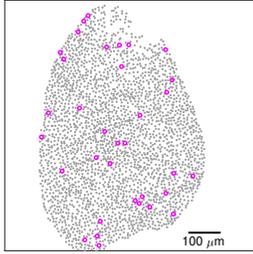
Originally,  $|S|$  events are sampled randomly on  $A$  [49]. For the discrete case however events are drawn randomly from  $N$  repeatedly and the respective  $K$  values are calculated. For the spatiotemporal case ( $K_{ST}$ ) the  $S_1$  coordinates are kept fix and  $|S_2|$  random events are drawn from  $N \setminus S_1$ .

To determine the observed pattern type the 5% and 95% confidence intervals (CI) are calculated from sampled  $K$  values and compared to the observed  $K$  values. In the case that the observed  $K$  values are within the CIs the observed pattern is called a random pattern, while being above the 95% CI indicates an aggregated and below the 5% CI a dispersed pattern (Figure 2.3).

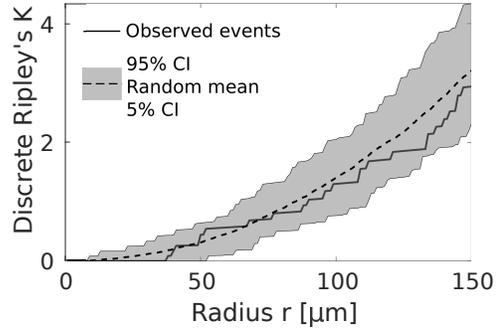
To evaluate the discrete Ripley's  $K$  method, different spatial patterns are simulated and compared to randomly sampled  $K$  values. We use cell coordinates (in microns) from a real experiment [72]. For a random pattern (Figure 2.4.1) 33 events (a number comparable to observed event frequency in [72])

are selected randomly from all  $\sim 2600$  possible event coordinates in  $N$ . As can be seen in Figure 2.4.2 the observed Ripley's  $K$  curve runs completely between the 5% and 95% CIs. To simulate aggregated patterns 33 events are added one by one. After every new event the event probability (starting at 1 for every coordinate) is increased by a factor of 1.6 for the surrounding coordinates within a radius of  $100 \mu\text{m}$ , resulting in an aggregated pattern (Figure 2.4.3). The respective Ripley's  $K$  curve runs above the 95% CI indicating an aggregated pattern (Figure 2.4.4). Similarly, dispersed events are sampled by altering the event probability (here, by a factor of 0.2 within a radius of  $100 \mu\text{m}$ ) after every added event (Figure 2.4.5). Again the Ripley's  $K$  curve, running below the 5% CI (Figure 2.4.6), confirms the simulation. Interestingly, the  $K$  curve deviates from the CI already at  $r \approx 50 \mu\text{m}$  and thus underestimates the simulated interaction radius.

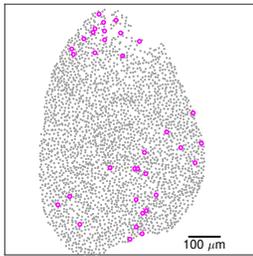
2.2. RIPLEY'S  $K$  FOR SPATIOTEMPORAL SINGLE CELL ANALYSIS 19



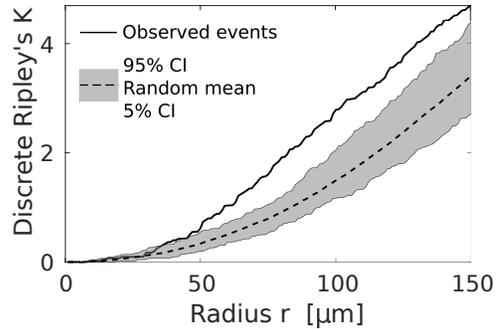
2.4.1: Randomly sampled events (magenta points) on 2600 possible coordinates.



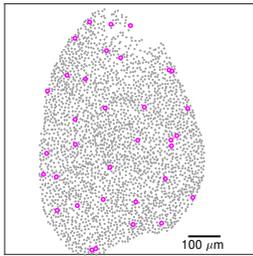
2.4.2: Discrete Ripley's  $K$  correctly identifying a random pattern for the simulation in Figure 2.4.1.



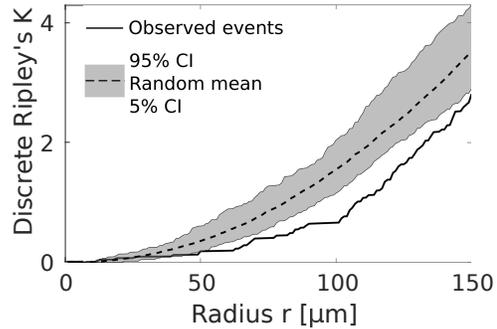
2.4.3: Events sampled in an aggregated manner within a radius of  $100 \mu\text{m}$ .



2.4.4: Discrete Ripley's  $K$  correctly identifying a aggregated pattern for the simulation in Figure 2.4.3.



2.4.5: Events sampled in a dispersed manner within a radius of  $100 \mu\text{m}$ .



2.4.6: Discrete Ripley's  $K$  correctly identifying a dispersed pattern for the simulation in Figure 2.4.5.

## 2.3 Model-based quantification of spatiotemporal patterns

A qualitative assessment of spatial and spatiotemporal patterns is described in the previous section using a modified version of the Ripley's K statistics. However, if the observed patterns deviate from random patterns the next step is of course the quantification of the observed patterns. To approach this problem one could define a problem-specific interaction model and find the best model parameters using maximum likelihood estimation [73].

### 2.3.1 Spatial interaction model

We start from the same problem setting as in Section 2.2, having a set of event coordinates  $N$  and a set of events  $S \in N$ . To identify whether these events are spatially related, one has to quantify the spatial relationship. This aim can be achieved by determining how pronounced the spatial relationship is (called interaction strength  $g \in \mathbb{R}$  further on) and how far the interaction spreads (described as interaction radius  $r \in \mathbb{R}^+$ ). To infer which parameter combination describes the observed spatial pattern best, a log-likelihood function  $L$  is defined taking into account events and non-events:

$$\log L(g, r) = \sum_{i=1}^S \log g^{\sum_{j=1, j \neq i}^S I(\text{dist}(i, j) \leq r)} / \sum_{i=1}^S \log g^{\sum_{k=1, k \neq i}^{N \setminus S} I(\text{dist}(i, k) \leq r)} \quad (2.7)$$

Equation 2.7 iterates over all events  $S$  twice while comparing non-identical event coordinates. In case the distance between event  $i$  and event  $j$  is smaller or equal to  $r$  the indicator function  $I(\text{dist}(i, j) \leq r)$  is 1. Thus it sums up all events in  $S$  within radius  $r$ . The denominator normalizes the equation by summing up all remaining possible events within  $r$ . The optimal  $g$  value is in this setting also a indicator on how the events are distributed, i.e.  $g < 1$  represents a dispersed pattern as one would expect less than one event within  $r$  around other events. On the contrary,  $g > 1$  means events are aggregated within  $r$  while  $g = 1$  indicates random events. Best parameters are inferred by minimizing Equation 2.7 with a multi-start optimization approach [74]. This means that the equation is minimized for multiple times from different initial parameters. Every run will either end up in a local minimum or will find the global one. Comparing the resulting log-likelihood values local minima can be separated from the global minimum.

The mechanism behind Equation 2.7 can be explained in a simplified 2D graphical example shown in Figure 2.5. Magenta circles are events while

black circles are coordinates without event. Evaluating Equation 2.7 for two events, a and b, within a radius  $r$  three other events are counted resulting in  $g^1$  for event a and  $g^2$  for event b.

This approach extends existing interaction models such as Gibbs or Cox point processes [75] with the condition that events (points) are restricted to predefined or discrete coordinates.

### 2.3.2 Spatiotemporal interaction model

Apart from the spatial parameters inferred within a set of events the spatial relationship between two sets of events  $S_1$  and  $S_2$  within  $N$  is worth to be explored. When events in  $S_1$  appear earlier than events in  $S_2$  they can be described as spatially and temporally correlated. Again the spatial relation can be described with two parameters, radius  $r$  and strength  $g$ . In contrast to Equation 2.7 the outer sum now loops over events from  $S_1$  while the second sum loops over  $S_2$ . Analogously to the spatial case the term is normalized by dividing through the sum over all remaining possible events  $N \setminus \{S_1, S_2\}$ :

$$\log L(g, r) = \sum_{i=1}^{S_1} \log g^{\sum_{j=1}^{S_2} I(\text{dist}(i,j) \leq r)} / \sum_{i=1}^{S_1} \log g^{\sum_{k=1, k \neq i, j}^{N \setminus \{S_1, S_2\}} I(\text{dist}(i,k) \leq r)} \quad (2.8)$$

The main idea of Equation 2.8 can again be explained with a simple two dimensional sketch. In Figure 2.6 cyan circles are  $S_1$  events while  $S_2$  events are shown in magenta and coordinates in  $N$  without events in black. Exem-

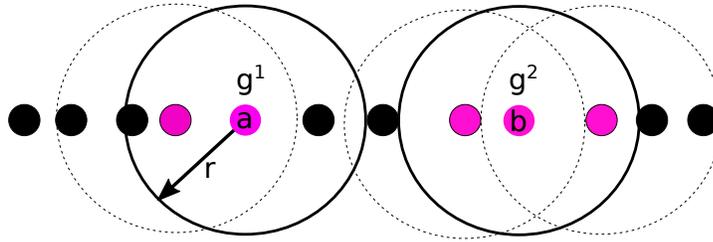


Figure 2.5: Illustration of the spatial interaction model. All filled circles represent coordinates at which events can occur. Events are observed for magenta circles, while no events are observed for black circles. Starting from two example events a and b, other events within radius  $r$  are counted to determine the exponent of  $g$ , leading to  $g^1$  for event a and  $g^2$  for event b. For all other events (magenta circles that are neither a nor b) the respective radii are shown with dashed lines.

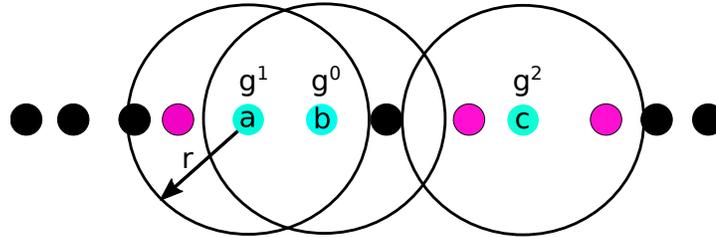


Figure 2.6: Illustration of the spatiotemporal interaction model. Cyan circles represent  $S_1$  events, magenta circles  $S_2$  events and black circles represent remaining coordinates of  $N$ . To exemplarily determine the exponent of  $g$ ,  $S_2$  events within distance  $r$  of  $a$  and  $b$  are counted, leading to  $g^1$  and  $g^2$  for  $a$  and  $b$  respectively.

ply  $S_2$  events around events  $a, b$  and  $c$  within  $r$  are counted determining the exponential of  $g$ . For the the example events  $a, b$  and  $c$  this leads to  $g^1$ ,  $g^0$  and  $g^2$  respectively.

### 2.3.3 Confidence measurement

To assess the robustness of inferred parameters it is common to determine confidence intervals (CIs) per parameter [76] suggesting a range for plausible parameter values. CI borders that are close to the most likely parameter indicate a well defined optimal parameter value while big CI ranges hint towards either a flat log-likelihood value landscape around the maximum or the existence of similar log-likelihood values for other parameter combinations [77]. CIs can be calculated with posterior sampling e.g. employing Markov-chain Monte Carlo methods [78]. Besides calculating the CI, posterior sampling allows to determine whether the posterior values show a single mode or are multi modal, i.e. show one or more equally likely parameter combinations [79].

## 2.4 Likelihood-free parameter estimation

Estimating parameters from snapshot measurements is challenging, as they describe a system state at one single time point and thus provide only incomplete information. To infer the missing parameters, a maximum likelihood approach is often employed [80]. However, sometimes there is not enough

information available about the underlying system or it is too complex to formulate a likelihood function. To approach this issue one can simulate complete time-solved data and take virtual snapshot measurements which can be compared to the observed snapshots. For such problems Approximate Bayesian Computation (ABC) methods can be applied as they can be employed to infer posterior distributions where likelihood functions are computationally unsolvable or too expensive to evaluate [81]. Instead of calculating the likelihood they compare observed and simulated data. To that end, ABC iteratively compares both data and thins out the distributions of parameter combinations resulting in very narrow distributions from which the most likely parameter sets can be obtained.

For instance, ABC was successfully applied to infer model parameters to match experimentally observed differentiation probabilities in hematopoiesis [82]. Another study used genotype data to analyze the transmission parameters of tuberculosis [83]. There, the authors successfully estimated the net transmission rate, the reproductive value and the doubling time of the bacterial population. Furthermore ABC was used in the context of HIV tracing, where efficiency of the HIV detection system was evaluated [84]. ABC in a spatial context was recently explored with spatial point patterns [85] and also used to model annual maximum temperature in South Australia [86]. Being such an important tool in so many different fields ABC algorithms got more sophisticated, easier to apply to all kinds of data types [87] and faster, over time [88, 89].

### 2.4.1 Problem definition

The problem setting we are confronted with is as follows: at two time points snapshot measurements of dividing cells, labelled with an S-phase marker are performed. Let's assume  $N$  cells are labelled at both time points. These cells are either still in division at the second measurement (double labelled S-phases, DLS) or are again in division (that is, re-dividing). The respective fractions of cells can be measured for a range of time intervals. However, the real re-division rate cannot be inferred by these snapshot measurements and need to be approximated along with cell cycle length and S-phase length. As it is often difficult to compare the full observed data set  $D$  and simulated data set  $\hat{D}$ , it is common to use summary statistics of the data sets instead [90]. Here, we employ the mean and standard deviation of observed re-divisions and DLS cells. The best parameter combinations can be inferred via minimizing the problem specific distance function based on those two

measures. This happens through iteration over every single time interval  $i$  and minimizing the distance  $\rho$  between simulated and observed data:

$$\rho(\hat{D}, D) = \sum_i \left( |\text{mean}(\text{obs}_i) - \text{mean}(\text{sim}_i)| + |\text{std}(\text{obs}_i) - \text{std}(\text{sim}_i)| \right) \quad (2.9)$$

### 2.4.2 Non-spatial division model

To evaluate a certain parameter combination via Equation 2.9, ten thousand trees are simulated for six generations, as shown in the division tree simulation algorithm, and virtual snapshot measurements are generated. The tree simulation algorithm is based on a cell cycle model (Figure 2.7) defined by the re-division probability  $p_{\text{rediv}}$ , cell cycle length and S-phase length. Cell cycle length and S-phase length are modelled with a lag exponential function [91] defined by a minimum length ( $\text{min}$ ) and scaling parameter for the exponential function ( $\beta$ ), respectively. The algorithm builds a tree of maximal depth  $t_{\text{depth}}$  by iterating via depth-first search over the whole tree and checking after every division whether a cell continues to divide using  $p_{\text{rediv}}$ . On the created tree one can measure the simulated re-division fraction by taking virtual measurements. An exemplary measurement is shown in Figure 2.8 where at two time points T1 and T2 the observed re-division

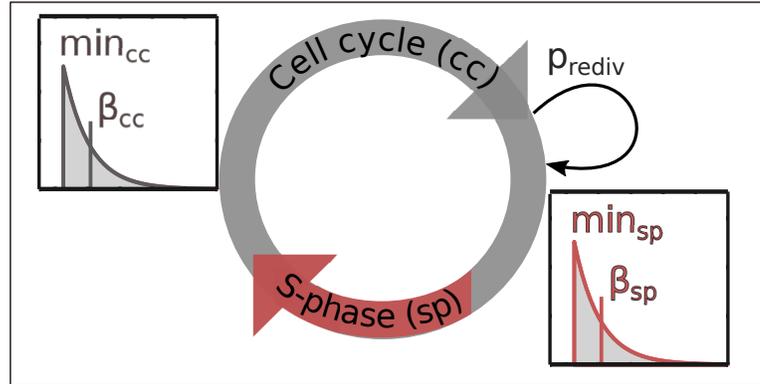


Figure 2.7: Illustration of the cell division model. The model describes by which parameters one cell division is influenced. First, the cell cycle length, which is using a lag exponential distribution defined with  $\text{min}_{\text{cc}}$  and  $\beta_{\text{cc}}$ , second the S-phase length, which again is using a lag exponential distribution defined with  $\text{min}_{\text{sp}}$  and  $\beta_{\text{sp}}$  and finally the re-division probability  $p_{\text{rediv}}$ .

frequency can be measured. From ten divisions, with proceeding S-phases (indicated with red lines in Figure 2.8) only two cells are found in S-Phase both at T1 and T2 which leads to an observed re-division fraction of 0.2. Analogously, the virtual measurement at T2' detects two cells (cell 1 and 2) as DLS with respect to the measurement at T1. To plugin the simulated measurements in Equation 2.9 one has to average over thousands of trees taking measurements for every single time interval  $i$ .

---

**Algorithm 1: Division tree simulation**


---

**Input:**

$tdepth$	▷ Tree depth
$t$	▷ Empty tree
$p_{rediv}$	▷ re-division probability
$min_{cc}$	▷ minimum cell cycle length
$\beta_{cc}$	▷ average cell cycle length
$min_{sp}$	▷ minimum S-phase length
$\beta_{sp}$	▷ average S-phase length

```

procedure RECURSEBRANCHING( $tdepth, t$ )           ▷ depth-first search
  if rand(0,1)  $\leq p_{rediv}$  &  $tdepth > 0$  then
     $cc \leftarrow min_{cc} + \text{Exp}(\beta_{cc})$            ▷ Draw cell cycle length
     $sp \leftarrow min_{sp} + \text{Exp}(\beta_{sp})$        ▷ Draw S-phase length
     $t.append(branch(cc, sp))$                  ▷ extend tree by new branch
    RECURSEBRANCHING( $tdepth - 1, t$ )         ▷ Call branching for first
  daughter
    RECURSEBRANCHING( $tdepth - 1, t$ )         ▷ Call branching for second
  daughter
  end if
end procedure

```

---

## 2.5 Agent-based cell population model

The growth and maintenance of cell populations can be simulated by agent-based models [92]. This means that every single cells acts as an agent and is able to interact with the environment and other nearby cells. The model is incorporated into the Morpheus [93] modeling environment and employs cellular Potts models [94] to model cell shape kinetics, as it allows for ar-

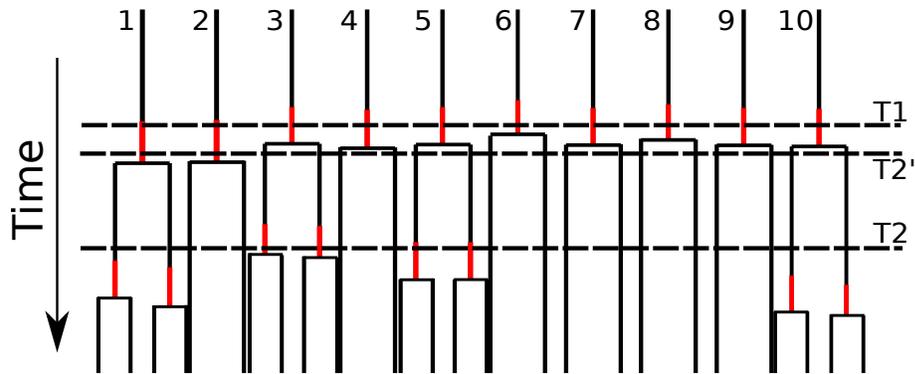


Figure 2.8: Illustration of ten division trees simulated with the Algorithm 1 to take virtual measurements that can be compared to observed ones via ABC. Exemplary measurements taken at time points T1 and T2 capture a re-division fraction of 20% as from ten observed S-phases (red lines) at T1, two S-phases in daughter cells are found at T2 (cells 3 and 5). A second measurement at T2' detects two DLS cells (cells 1 and 2).

bitrary cell shapes and stochasticity in cell movement. Model parameters, which are observable e.g. cell size, differentiation rate, and division probability are approximated and transferred. Cell size is modeled with a sigmoid curve from smallest observed cell size to biggest and differentiation rate and division probability are calculated from the observed cell numbers. Cell cycle specific parameters are adopted from the results of the ABC model in the previous section. This setup allows to compare the influence of different conditions on the cell population over time. We apply this model to compare spatiotemporal division patterns in cell populations with reoccurring cell divisions against the null model having only the base division rate [72].

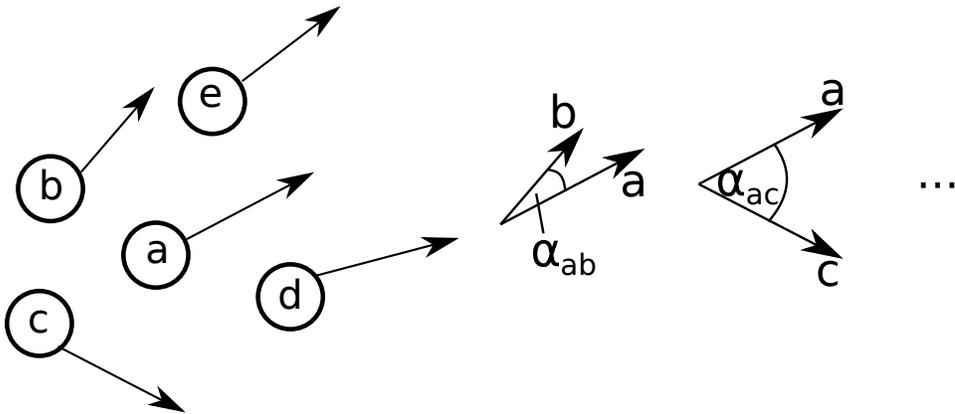
## 2.6 Neighborhood coordination model

Given a set of moving objects it may be of interest to investigate whether these objects are moving in a coordinated vs. an independent fashion. Such analyses were e.g. performed on flocks of birds to investigate swarm behaviour [95, 96]. Coordinated swarm-like behaviour was also analyzed in the context of immune response models where the authors compared different spatial behaviours such as aggregation, dispersion or random movement for immune cells to find out the best immune response strategy [97]. In cell biology cell movement has been analysed by comparing long axes (and thus

movement direction) of cells [98].

While object elongation is a good proxy to estimate movement direction, having measured at least two time points for every object allows to calculate a movement vector per object and time point precisely. An exemplary two dimensional representation of five objects and their movement vector is visualized in Figure 2.9.1. However, to analyse coordination it does not make sense to compare the movement direction of every object against each other: if objects are e.g. arranged in a circle and perform a circular movement the objects move coordinated but not all in the same direction. The pairwise comparison would yield a small coordination. Instead only the movement of neighboring objects should be compared. To identify the neighborhood of an object in two or three dimensions a triangulation [71] can be applied. Cells with a distance from each other below a threshold can be treated as neighbors.

We defined a neighborhood movement similarity model, which returns a score describing to what extent one cell moves in the same directions as its direct neighbors. The neighborhood movement similarity is determined by calculating the angle  $\alpha$  of movements vectors of two neighboring cells (as shown in Figure 2.9.2). For object  $a$   $\alpha_i$  is calculated for all neighbors  $n$  (e.g. in Figure 2.9.1 neighbors of  $a$  are objects  $b$ - $e$ ) and then averaged to get one



2.9.1: Two dimensional representation of objects  $a$ - $e$  moving in a direction indicated by the respective arrow.

2.9.2: Neighborhood movement similarity is represented by calculating the intermediate angle  $\alpha$  between the movements arrows of two objects. The smaller the angle the higher the coordination.

value reflecting the coordination of this object with its neighbors:

$$\bar{\alpha}_a = \frac{1}{n} \sum_{j=1}^n \alpha_{a,j} \quad (2.10)$$

Analogously the neighborhood coordination can be calculated for objects moving in three dimensions. For two 3d vectors  $u$  and  $v$  this is:

$$\alpha_{u,v} = \cos^{-1} \left( \frac{u_x v_x + u_y v_y + u_z v_z}{\sqrt{u_x^2 + u_y^2 + u_z^2} \sqrt{v_x^2 + v_y^2 + v_z^2}} \right) \quad (2.11)$$

With the movement similarity calculated for every object it is easy to identify spatial regions in 3D in which objects move coordinated (small  $\alpha$ ), independent ( $\alpha \sim 90^\circ$ ) or even apart from each other ( $\alpha$  close to  $180^\circ$ ).

We apply this method to analyze the coordinated behaviour of fascia fibroblasts during wound healing and identify synchronized regions over time (see Section 3.4) [99].

## Chapter 3

# Summary of contributed articles

### 3.1 Image analysis of neural stem cell division patterns in the zebrafish brain

To investigate stem cell regeneration and proliferation in the adult body stem cells we analyzed neural stem cells (NSCs) in the zebrafish brain and published the first results 2018 in Cytometry A [100] (IF 3.1). For this project we collaborate with Prisca Chapouton from the labs of Jovica Ninkovic ("Neurogenesis & Regeneration") and Hernan Lopez-Schier ("Sensory Biology & Organogenesis"), who performed all experimental tasks.

It is known that NSCs divide to sustain the stem cell population and differentiate into neural progenitor cells, which finally refill the pool of mature neurons and glial cells. However, to what amount NSC division and differentiation is spatially coordinated i.e. what spatial patterns can be identified remains an open question (**Research question I** in section 1.3). To quantify the patterns of NSC divisions, NSCs need to be identified, then one has to determine NSCs that divide within a certain time window and finally the spatial coordination of those divisions need to be analyzed. I developed a bioimage informatics pipeline that is able to automatically identify GFP expressing NSCs of zebrafish brain hemispheres in three-dimensional image stacks from whole-mount preparations.

The pipeline exploits the fact that NSCs are located on a slightly curved surface, which is almost two-dimensional. On this surface I identify around 2000 NSCs in six zebrafish brain hemispheres. To determine the position of dividing NSCs in the hemisphere Prisca Chapouton injected EdU, which is

incorporated into cells undergoing S-phase. I use the positional information to calculate all pairwise NSC distances with three alternative metrics. One metric that takes into account the bent surface, the second one is the Euclidean distance in three dimensions and at last a network based distance deduced from Delaunay triangulation on all NSCs. Finally, I fit a probabilistic model, evaluating influence radius and strength, to the observed spatial NSC patterns that is able to account for the non-homogeneous distribution of NSCs.

We are able to answer the spatial component of **Research question I** by applying a probabilistic model, which is not only able to identify the observed pattern but also quantify it with an interaction radius and strength. We detect a weak positive coordination between NSCs in S-phase for all three tested metrics and conclude that neither strong inhibitory nor strong enhancement signals drive NSC proliferation in the adult zebrafish brain.

**My contribution:** First of all I did all computational work. I implemented the single cell identification pipeline (SCIP) to extract x, y and z coordinates of single cells (<https://github.com/marrlab/SCIP>). To that end, I combined a blob detection method from Fiji that generates cell candidates with an iterative Matlab script. Iteratively, outliers are sorted out by either a size filter or a distance filter, which excludes candidates that are too distant from the majority of other cell candidates.

To process the obtained NSC coordinates I implemented an algorithms to analyze the spatial patterns of dividing cells via parameter inference on three different distance metrics. The algorithm infers the most likely inference radius and strength to explain a certain spatial pattern. A slightly modified version is even able to combine multiple experiments (brain hemispheres) to obtain the most likely parameters that explain all observed spatial patterns. For quality control, I created artificial data with certain combinations of radius and strength to ensure that the probabilistic model can recreate them. Finally, I compared my cell identification method against two other state of the art methods (one Fiji plugin and the Imaris software) to show that SCIP performs similar or better.

### 3.2 Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatiotemporal patterns

We continued the collaboration with the both labs and extended the analysis of NSCs in the zebrafish brain. The resulting paper "Agent-based modeling reveals that reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatiotemporal patterns" [72] was published in 2020 in PLoS Biology (IF 7.6). In the paper we investigate the regulatory mechanisms of cell cycle entry and quiescence as they are poorly understood. However, they are crucial for the maintenance of stem cell populations. Especially, it is unclear whether cells divide with a certain spatiotemporal pattern and whether a model is able to quantify and reproduce the spatiotemporal activity of single stem cells within a population (**Research question I** in section 1.3). To gain more insight into these processes we analyzed division events in a population of adult neural stem cells (NSCs) within the zebrafish telencephalon. To that end, I developed and applied spatiotemporal statistics and mathematical models to over 80,000 NSCs in 36 brain, which could show that NSCs have weakly aggregated, non-random division patterns in space and time. Analyzing divisions at two time points ranging from 9h to 72h allowed me to infer S-phase and cell cycle length computationally. Strikingly, we observed fast cell cycle re-entries in about 15% of newly born NSCs. I employed agent-based simulations to model NSC populations and could show that the observed re-entries suffice to induce aggregated spatial and spatiotemporal division patterns that are consistent with the ones observed in experimental data. I compared the results to simulations that omit re-divisions, which leads to random spatial and spatiotemporal distributions of dividing cells. We conclude that spatiotemporal aggregated patterns of dividing NSCs can emerge simply from the cell's history.

**My contribution:** For this paper I identified NSCs with SCIP and merged the x, y and z coordinates with information whether each cell is positive or negative for both applied S-phase markers and how long the respective markers have been applied before imaging the brain. These feature combination allowed me to perform spatial statistics such as Ripley's  $K$  on a spatial and spatiotemporal level. Spatial analysis means investigating the patterns of S-phase positive cells for one of the two markers, while spatiotemporal analysis examines the patterns occurring between S-phase positive cells of

one marker and the S-phase positive cells of the second marker, which was applied later in time. I implemented both versions in Matlab with an experiment specific edge-correction method since the edges of a hemisphere are not easily determined (see Methods 2.2).

To quantify the observed spatiotemporal patterns I extended the interaction model from spatial interactions (see previous section) to spatiotemporal interactions. Furthermore, I designed a cell division model that takes cell cycle length, S-phase length and re-division probability as input parameters and simulates cell proliferation. After several cell cycles virtual measurements are performed such that they are comparable with observed measurements. Using an ABC implementation in python I optimized the input parameters to fit the observed measurements. Next I used these optimal parameters to feed an agent-based model, which I calibrated to simulate the growth and maintenance of a two-dimensional NSC population. Finally I evaluated the output of the agent-based model with the same spatial statistics and models as before to compare it to the observed patterns in experimental data.

Code to reproduce analysis, parameter inference and result figures of the paper can be found at <https://github.com/marrlab/spatiotemporalAnalyses>.

### 3.3 Live cell-lineage tracing and machine learning reveal patterns of organ regeneration

In the paper "Live cell-lineage tracing and machine learning reveal patterns of organ regeneration" [101] published 2018 in eLife (IF 7.1) we analyzed the regrowth of sensory organs in the model organism zebrafish. This project was done in collaboration with biologists from the Lopez-Schier lab ("Sensory Biology & Organogenesis") from the Helmholtz Center Munich.

Sensory organs in zebrafish are called neuromasts and have the crucial ability to regain normal architecture during repair to maintain their functionality after they got damaged severely. We use an approach that joins live cell-lineage tracking and classification with random forests to reveal how cell type and localization are adapted during neuromast regeneration. To that end, Oriol Viader from the lab of Hernan Lopez-Schier utilized superficial neuromasts in larval zebrafish, which consists of three cell classes arranged in a single planar-polarity axis and radial symmetry. He analyzed cell-fate transitions at high temporal resolution and revealed that neuromasts regenerate isotropically to reclaim cell type proportions, geometric order and polarity with extraordinary precision. I extracted spatial and spatiotemporal features from every single neuromast regeneration movie aiming to predict cell fate

### 3.4. SCAR FORMATION IS DRIVEN BY N-CADHERIN DEPENDENT COLLECTIVE FIBROBLAST MI

and cell lineage compositions (**Research question II** in section 1.3). The most promising feature I identified to predict cell fate is the distance to the neuromast center during regeneration. We suggest a self-regulatory process that leads the regeneration to almost identical result with minimal extrinsic input. The integrated method that we have developed is widely applicable and also quite simple, which should help to determine predictive features that influence the growth of complex tissues.

**My contribution:** I implemented the algorithm to extract cell type and lineage information from neuromast regeneration movie data. As quality control I identified too short cell trajectories and checked the data for cell tracks that appeared to be wrong as they e.g. change their cell type more than once within one cell cycle. I plotted all cell lineages to allow visual analysis of the data and to assign a cell lineage label. The cell lineage label was determined by the cell type composition within the lineage. Additionally I tailored spatial and spatiotemporal single cell features that can be extracted from the data at every measurement time point. I fed a random forest algorithm with the features and predicted cell fate for every dividing cell and lineage composition at the time of the first division. To find out which features contribute the most to the prediction I calculated the feature importance for every single spatial feature by permutation i.e. features are permuted and the more the model error changes per feature the higher is the respective feature importance. I further extracted cell division frequency of every cell type over time and visualized and compared the division onset for the different cell types. I also matched the size of cell type clones over four cell generations and determined the relative starting positions of different cell lineage trees. Additionally I benchmarked all movies with respect to their ablation sites to reveal to what extent the position of the remaining cells influence the process of neuromast regeneration.

### 3.4 Scar formation is driven by N-Cadherin dependent collective fibroblast migration

In the paper "Fascia fibroblasts swarm to drive scar formation through N-cadherin", which is currently under review at Nature Communications (as of 11.09.20) we investigate scar formation after injury in mammals. The experimental work was done in the lab of Yuval Rinkevich from the Helmholtz Center Munich.

Mammalian connective tissue responds to injury by either regeneration or

scarring. As the direct observation in animals is too challenging and current assays fail to represent the real physiological setting, fibroblastic mechanisms and actions that drive connective tissue responses remain unclear. The Rinkevich lab developed a skin explant technique termed scar-in-a-dish (SCAD) that is able to accurately reproduce scar formation and reveals how scarring emerges in incredible detail. Using a traceable scar-progenitor fibroblast lineage in mouse SCAD experiments, they observed new and unseen connections that form between scar-progenitors, which then collectively swarm towards the injury site in a highly coordinated fashion that step-by-step contract the skin around the wound and form scars. To investigate the collective movement (**Research question III** in section 1.3), I compared cell swarming computationally under different conditions and verified that swarming can only be detected in scar progenitors but not in oral cavity fibroblasts, which regenerate scarless. To find out which factors might stimulate cell-cell connections the Rinkevich lab tested several adhesion molecules and found that an up-regulation of N-cadherin in scar progenitors induced the swarming behaviour. Inhibiting N-cadherin binding impeded swarming and skin contraction, and also resulted in less scarring in SCAD and in mice. Ultimately blocking N-cadherin and the swarming ability of scar-progenitors opens up new therapeutic options to reduce pathological fibrotic responses for a broad field of medical applications.

**My contribution:** My part of the paper involved post-experimental single cell data analyses. First I cleaned the data and excluded short cell tracks or misaligned tracks having an impossible movement distance within the measured 15 minute time intervals. For the remaining cell tracks I calculated the fibroblast velocity over time, which was then used to compare cell movement and velocity between wild-type fibroblasts and oral cavity fibroblasts. Furthermore, I implemented an algorithm that calculates collective cell movement by comparing the movement directions of single cells to their neighboring cells (see Methods 2.6). The collective movement can thus be compared between wild-type and oral cavity fibroblasts by a movement similarity score. I found out that the mean movement distance as well as the collective behaviour differs significantly comparing the two cell types. Finally, I visualized cell velocity and movement similarity score in a way that different conditions can be compared easily and aberrations visually stand out.

## Chapter 4

# Discussion & future perspectives

### 4.1 Spatiotemporal analysis in single cell biology

Spatial models are an important tool to gain mechanistic insight into cell population processes. We devised new models and modified and transferred existing ones from other fields to make them applicable to single cell data. Mechanistic models take mechanistic prior knowledge into account. This prior knowledge could either be a biological mechanism such as a particular division behaviour of cells, or well engineered features that are fed to the model. We used these mechanistic models to gain new biological insights and answer open research questions in stem cell biology.

We have expanded the knowledge of spatiotemporal aspects of tissue maintenance. First, we identified spatiotemporal patterns of dividing neural stem cells (NSCs) while considering the surrounding remaining cell population as the substrate of division events. Then we introduced a spatiotemporal model to quantify the spatial reach and strength of the identified patterns [100]. Furthermore, we were able to confirm via modeling that a cell intrinsic mechanism, namely reoccurring divisions, is able to create the identified patterns. To the best of our knowledge, this is the first agent-based model of NSC tissue maintenance with single cell resolution and a potential starting point for more research in this direction [72].

To provide new mechanistic insight into the regeneration of the zebrafish lateral organ, we employed machine learning models based on spatial single

cell features [101]. To that end we analyzed single cells in organ regeneration movie data and extracted spatiotemporal features for every single cell. A machine learning model allowed us to identify those features that are crucial for cell identity, such as distance to certain cell types or the organ center. This is the first time that a data-driven model-informed mechanistic statement can be made for lateral organ regeneration.

We contributed to scarring research by introducing a neighborhood model that can be used to assess how similar individual cells move in a moving population [99]. A similarity score can evaluate cell movement behaviour and can also be compared between cell types. We found that wild type fibroblasts move in a highly coordinated manner towards the injury site while oral cavity fibroblasts don't show any collective movement after injury. We are the first, to our knowledge, to quantify the collective movement of cells during scarring.

Our analyses can be improved and extended in several directions: The spatiotemporal analysis of dividing NSCs in zebrafish hemispheres in sections 3.1 and 3.2 is based on the 3D locations of NSCs. However, we up to now only analyzed 2D cell shapes. Including 3D cell volumes can potentially improve the analysis. To that end, a 3D cell segmentation algorithm is required. Performance of cell segmentation algorithms in 3D images has improved significantly over the last years [63], especially with the emergence of powerful deep learning (DL) methods [67, 102]. The additional 3D volume and morphology information of every single cell could be included in further analyses. Cell volumes could be correlated with hemisphere location to identify specific brain regions. Cell morphology could be compared between young and old animals to model cell and brain growth in a more detailed manner. However, the bottleneck is to obtain good and enough 3D ground truth data to train the DL models.

An exciting addition to our spatiotemporal interaction model (Section 2.3.2) would be a time dependent parameter inference. As implemented now, parameters are inferred for every labeling interval separately (see Figure 4.1, taken from [100]), without information transferred between neighboring labeling intervals. One could include temporal dependencies by inferring all labeling interval parameters at once and infer the slope running through the maximum likelihood values and y-axis shift as additional parameters (e.g. red lines in Figure 4.1). Even more complex models could be investigated to identify e.g. periodicity in strength or radius (e.g. red dashed lines in Fig-

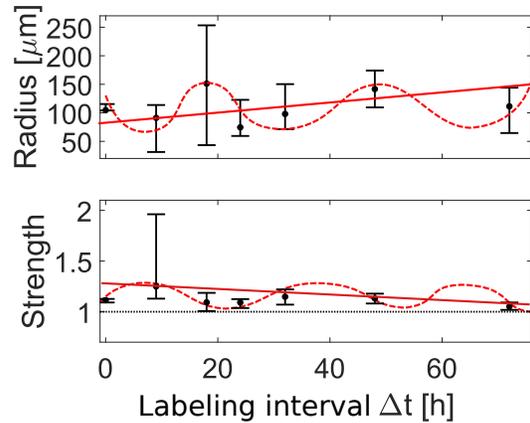


Figure 4.1: Interaction radius and strength inferred from spatiotemporal S-phase data in the zebrafish brain for labeling intervals between 9 h and 72 h (see Section 2.3.2). Red lines show a possible linear time dependency of strength and radius, while the dashed lines depict possible more complex dependencies.

ure 4.1). While this would increase the size of the model by at least four additional parameters (slope and y-axis shift for radius and strength, respectively) and thus computational complexity, it could also generate valuable time dependency information about the temporal extent of the observed correlations.

A biological follow-up experiment, in which single cells are tracked and labeled with a S-phase marker over several days could validate the inferred S-phase and cell cycle length as well as confirm re-divisions. Furthermore, one could follow differentiated cells to determine the time they need to leave the hemisphere surface and compare their movement to NSCs. But time-lapse analyses are challenging and only a few cells can be tracked at once so far, as shown in Barbosa et al. [35]. However, time-lapse data would also open up the possibility to investigate causality in spatiotemporal patterns.

Spatial causality is similar to causality in time series data: Given a spatial pattern  $P$  it causes another spatial pattern  $Q$  if  $P$  provides useful information about  $Q$ . However, identifying causality for patterns that occur based on previous events is difficult to verify and is often confused with correlation [103]. This is the reason why we investigated interaction strength and radius in the zebrafish brain project instead of causality [72]. However, fit-

ting an interaction model does not necessarily mean that interactions are present [104]. With time-lapse data at hand, Granger causality [105] could be a meaningful measure to detect to what extent an earlier S-phase pattern causes a later S-phases pattern. This method has been applied successfully to infer causality from spatiotemporal data for climate change models with respect to the influence of CO<sub>2</sub> and other greenhouse gases on changes in temperature [106] and might be applicable in our context.

To extend the organ regeneration study [101] one could add more cell specific features such as cell area and sphericity. These features could be harvested by a 3D single cell segmentation algorithm as discussed in (1) and could give hints to differentiation decision in tissue context. For instance, a sustentacular cell that will differentiate in either a hair cell or into a mantle cell could already shift its shape into the respective direction. This of course only works when different cell types show different shape characteristics during cell decision process, a hypothesis that could be interesting to address. The fact that the mantle cells observed in our data are rather elongated in contrast to other cell types is a hint in that direction. Again, the bottleneck is 3D cell segmentation or manually annotated ground truth data.

Another way to get more insight into the process of neuromast regeneration is to describe the whole process computationally with an agent-based model. Employing a software that can simulate 3D cell colony growth such as Morpheus [93] could reveal differentiation and division features crucial for healthy neuromast regeneration. One could start the model with a few initial cells and utilize division times (depending on cell type) and location dependent differentiation probabilities. For instance, cells at the colony border might be more likely to become mantle cells and cells in the middle might be prone to become hair cells. To evaluate the final cell colony with 60-70 cells one needs to define characteristics (such as mantle cells being on the outside) and cell type composition, which is defined by roughly 20-25% hair cells, 60% sustentacular cells and 15-20% mantle cells. Parameters could be optimized via ABC (Section 2.4) or via reinforcement learning similar to a recent agent-based model that has been used to simulate vascular tumor growth [107]. This approach would allow to test and compare different division and differentiation strategies e.g. random vs. spatial informed (dependent on the spatial location) division or differentiation. Ultimately, such an approach could allow to simulate the whole neuromast generation process and could inform further research in organ or even tissue regeneration.

## 4.2 Spatial analysis in medical research

Spatial analyses find their way in all kinds of application fields including medical domains such as immunology [108] or digital pathology [109]. Digital pathology is a sub-field of pathology that works with information gained from digitized specimen slides. More and more tissue slides are digitized and thus enable computer algorithms to perform expert tasks by learning from huge amounts of data [110]. With the use of neural networks that extract, amongst others, spatial features from pathology images, computational labs all over the globe aim to support (or partially replace) pathologists by e.g. grading tumor images [111, 112] or predicting cancer from tissue slides [113, 114]. Cancer and immune cell markers can enable further spatial analyses of immune cells with respect to their spatial tumor context and predict e.g. survival for colorectal cancer [115] or lung cancer [116]. In the latter study, survival of lung cancer patients was predicted successfully using spatial interactions between tumor cells and immune cells.

Going one step further in the complexity of spatial analyses leads to spatial transcriptomics. This recent technology enables scientists to measure all gene expression in a tissue sample with  $\mu\text{m}$  resolution [117]. A location specific spatial barcode binds and captures adjacent mRNAs from the tissue. This captured mRNA is sequenced and the spatial barcode makes it possible to localize the origin within the tissue section for each individual mRNA transcript. Spatial transcriptomics allow e.g. detecting which genes are expressed in a certain area of interest or searching for locations where specific genes are expressed [118]. These analyses can refine the identification of cancer and stroma regions or reveal immune cell infiltration [14, 15]. Commercialized spatial transcriptomics technology from 10X genomics is already applied to explore gene function variation across diverse cell types with paired CRISPR knockdowns and rich single cell phenotype readouts [119] or to characterize cell populations and their locations within a tissue [120]. It is also used to visualize the entire transcriptome and to identify new targets and the positions of cells that express them [121] or map the molecular, cellular and spatial composition of distinct bone marrow niches [122].

With all this information about single cells or locations within a tissue one could try to enhance cancer related predictions with a high-dimensional Ripley's  $K$  approach where not only the physical location is used to calculate the distance between locations but also the expression profile [118, 123].

Recently the power of spatial tissue analysis increased with the addition of spatial proteomics [124, 125]. This combination allows to identify cell sub-populations based on both their gene expression and protein level together

with their spatial distributions [126]. In tumor research in particular, spatial analyses in combination with various omics promise a better understanding of the interactions between cells and the spatial restrictions that enable the specialization and development of cancer cells [127].

### 4.3 Methodological perspectives

Medical research can be progressed by adding more omics but also by novel analysis methods. Spatial features extracted via neural networks in combination with a temporal component could advance the understanding of tumor progression over time. The temporal component could be introduced by determining the cancer stage of a tissue sample for each patient or embedding each patient in a low dimensional representation [128]. For every cancer stage descriptive spatial features (compare Section 1.4) are extracted, which are able to characterize each stage. One could then infer cancer growth factors employing spatiotemporal models or machine learning. These factors, together with a healthy initial cell distribution, could be fed into an agent-based spatiotemporal model framework such as Morpheus [93]. The artificial cell culture would then grow until it reaches a certain cancer stage. Based on the previously defined spatial characteristics one could evaluate whether it was possible to reach the aspired tumor stage with the employed model. That modeling of huge cell populations is possible has been shown e.g. by Jagiella et al. [129], where up to  $10^6$  single cells have been modeled. However, they also showed that it represents an enormous computational challenge and thus high energy demands.

With the advent of deep learning (DL) models space takes on an additional meaning, as in latent feature space. Latent feature space is a representation of the input data after it is passed through the deep neural network [130]. Extracting features at the last layer of a convolutional neural net was e.g. used to visualize cell cycle progression and cell cycle phases [131]. Similarly, extracting features from the middle layer of a variational autoencoder can capture biologically relevant features to model cancer gene expression [132]. Within this latent feature space one could apply spatial methods or identify spatial patterns and compare them to patterns within the input data.

One could not only explore space in DL applications but even include the whole model into a DL framework. Some very promising applications are physics informed DL models, which are able to solve supervised learning tasks while respecting certain laws of physics [133]. Recently a DL algo-

rithm was able to figure out that our solar system has to be heliocentric based on observations of sun and mars from earth [134]. A possible biomedical application could be to model the growth of different embryonic stem cells by information about spatial environmental conditions such as surrounding tissue, pH value and temperature. The aim of the model would be to find out which organ grows with the given information.

With the advancements in spatial omics and models being insertable into deep learning architectures this could open up new possibilities to gain insight into disease etiology and tumor growth.



# Bibliography

- [1] John Snow. *On the mode of communication of cholera*. John Churchill, 1855.
- [2] Jianyun Lu, Jieni Gu, Kuibiao Li, Conghui Xu, Wenzhe Su, Zhisheng Lai, Deqian Zhou, Chao Yu, Bin Xu, and Zhicong Yang. Covid-19 outbreak associated with air conditioning in restaurant, guangzhou, china, 2020. *Emerging Infectious Diseases*, 26(7), 2020.
- [3] Lea Hamner. High sars-cov-2 attack rate following exposure at a choir practice—skagit county, washington, march 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69, 2020.
- [4] Mohammad Arashi, Andriette Bekker, Mahdi Salehi, Sollie Millard, Barend Erasmus, Tanita Cronje, and Mohammad Golpaygani. Spatial analysis and prediction of covid-19 spread in south africa after lockdown, 2020.
- [5] Chad R Wells, Pratha Sah, Seyed M Moghadas, Abhishek Pandey, Affan Shoukat, Yaning Wang, Zheng Wang, Lauren A Meyers, Burton H Singer, and Alison P Galvani. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences*, 117(13):7504–7509, 2020.
- [6] Alexandra Diehl, Leandro Pelorosso, Claudio Delrieux, Celeste Saulo, Juan Ruiz, M Eduard Gröller, and Stefan Bruckner. Visual analysis of spatio-temporal data: Applications in weather forecasting. In *Computer Graphics Forum*, volume 34, pages 381–390. Wiley Online Library, 2015.
- [7] Yang Yue and Anthony Gar-On Yeh. Spatiotemporal traffic-flow dependency and short-term traffic forecasting. *Environment and Planning B: Planning and Design*, 35(5):762–771, 2008.

- [8] Alireza Ermagun and David Levinson. Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews*, 38(6):786–814, 2018.
- [9] Serdar Çolak, Lauren P Alexander, Bernardo G Alvim, Shomik R Mehndiratta, and Marta C González. Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record*, 2526(1):126–135, 2015.
- [10] Barton Gellman and Ashkan Soltani. Nsa tracking cellphone locations worldwide, snowden documents show. *The Washington Post*, 4:2013, 2013.
- [11] Nuria Oliver, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Delataille, Marco De Nadai, Emmanuel Letouzé, Albert Ali Salah, Richard Benjamins, Ciro Cattuto, et al. Mobile phone data for informing public health actions across the covid-19 pandemic life cycle, 2020.
- [12] Simon N Williams, Christopher J Armitage, Tova Tampe, and Kimberly Dienes. Public attitudes towards covid-19 contact tracing apps: A uk-based focus group study. *medRxiv*, 2020.
- [13] Marc Dufrêne and Pierre Legendre. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological monographs*, 67(3):345–366, 1997.
- [14] Emelie Berglund, Jonas Maaskola, Niklas Schultz, Stefanie Friedrich, Maja Marklund, Joseph Bergensträhle, Firas Tarish, Anna Tanoglidi, Sanja Vickovic, Ludvig Larsson, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature communications*, 9(1):1–13, 2018.
- [15] Niyaz Yoosuf, José Fernández Navarro, Fredrik Salmén, Patrik L Ståhl, and Carsten O Daub. Identification and transfer of spatial transcriptomics signatures for cancer diagnosis. *Breast Cancer Research*, 22(1):1–10, 2020.
- [16] Jun Li, Julia A Mack, Marcel Souren, Emre Yaksi, Shin-ichi Higashijima, Marina Mione, Joseph R Fetcho, and Rainer W Friedrich. Early development of functional spatial maps in the zebrafish olfactory bulb. *Journal of Neuroscience*, 25(24):5784–5795, 2005.

- [17] Olivier Pourquié. The segmentation clock: converting embryonic time into spatial pattern. *Science*, 301(5631):328–330, 2003.
- [18] Ron Sender, Shai Fuchs, and Ron Milo. Revised estimates for the number of human and bacteria cells in the body. *PLoS biology*, 14(8):e1002533, 2016.
- [19] Jon S Odorico, Dan S Kaufman, and James A Thomson. Multilineage differentiation from human embryonic stem cell lines. *Stem cells*, 19(3):193–204, 2001.
- [20] HE Shenghui, Daisuke Nakada, and Sean J Morrison. Mechanisms of stem cell self-renewal. *Annual Review of cell and Developmental*, 25:377–406, 2009.
- [21] Jonathan MW Slack. *Essential developmental biology*. John Wiley & Sons, 2012.
- [22] Alexander Birbrair and Paul S Frenette. Niche heterogeneity in the bone marrow. *Annals of the new York Academy of Sciences*, 1370(1):82, 2016.
- [23] Dhvani Jhala and Rajesh Vasita. A review on extracellular matrix mimicking strategies for an artificial stem cell niche. *Polymer Reviews*, 55(4):561–595, 2015.
- [24] Jian Tang, Rong Peng, and Jiandong Ding. The regulation of stem cell differentiation by cell-cell contact on micropatterned material surfaces. *Biomaterials*, 31(9):2470–2476, 2010.
- [25] Stephanie J Ellis and Guy Tanentzapf. Integrin-mediated adhesion and stem-cell-niche interactions. *Cell and tissue research*, 339(1):121, 2010.
- [26] Silvia S Chen, Wendy Fitzgerald, Joshua Zimmerberg, Hynda K Kleinman, and Leonid Margolis. Cell-cell and cell-extracellular matrix interactions regulate embryonic stem cell differentiation. *Stem cells*, 25(3):553–561, 2007.
- [27] Rachel A Ankeny and Sabina Leonelli. What’s so special about model organisms? *Studies in History and Philosophy of Science Part A*, 42(2):313–323, 2011.

- [28] VN Sarvaiya, KA Sadariya, MP Rana, and AM Thaker. Zebrafish as model organism for drug discovery and toxicity testing: a review. *Veterinary clinical science*, 2(3):31–38, 2014.
- [29] NF Beliaeva, VN Kashirtseva, NV Medvedeva, IuIu Khudoklinova, OM Ipatova, and AI Archakov. Zebrafish as a model organism for biomedical studies. *Biomeditsinskaja khimiia*, 56(1):120, 2010.
- [30] Matthew Gemberling, Travis J Bailey, David R Hyde, and Kenneth D Poss. The zebrafish as a model for complex tissue regeneration. *Trends in Genetics*, 29(11):611–620, 2013.
- [31] Kenneth D Poss, Lindsay G Wilson, and Mark T Keating. Heart regeneration in zebrafish. *Science*, 298(5601):2188–2190, 2002.
- [32] Kenneth D Poss. Getting to the heart of regeneration in zebrafish. In *Seminars in cell & developmental biology*, volume 18, pages 36–45. Elsevier, 2007.
- [33] Jiaxiang Jiang, Po-Yu Kao, Samuel A Belteton, Daniel B Szymanski, and BS Manjunath. Accurate 3d cell segmentation using deep features and crf refinement. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1555–1559. IEEE, 2019.
- [34] Mehmet Ilyas Cosacak, Christos Papadimitriou, and Caghan Kizil. Regeneration, plasticity, and induced molecular programs in adult zebrafish brain. *BioMed research international*, 2015, 2015.
- [35] Joana S Barbosa, Rosario Sanchez-Gonzalez, Rossella Di Giaimo, Emily Violette Baumgart, Fabian J Theis, Magdalena Götz, and Jovica Ninkovic. Live imaging of adult neural stem cell behavior in the intact and injured zebrafish brain. *Science*, 348(6236):789–793, 2015.
- [36] Rebecca L Bernardos and Pamela A Raymond. Gfap transgenic zebrafish. *Gene Expression Patterns*, 6(8):1007–1013, 2006.
- [37] Donovan Correa-Gallegos, Dongsheng Jiang, Simon Christ, Pushkar Ramesh, Haifeng Ye, Juliane Wannemacher, Shruthi Kalgudde Gopal, Qing Yu, Michaela Aichler, Axel Walch, et al. Patch repair of deep wounds by mobilized fascia. *Nature*, 576(7786):287–292, 2019.
- [38] Simon P Peron, Jeremy Freeman, Vijay Iyer, Caiying Guo, and Karel Svoboda. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783–799, 2015.

- [39] Caleb R Stoltzfus, Jakub Filipek, Benjamin H Gern, Brandy E Olin, Joseph M Leal, Yajun Wu, Miranda R Lyons-Cohen, Jessica Y Huang, Clarissa L Paz-Stoltzfus, Courtney R Plumlee, et al. Cytomap: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. *Cell reports*, 31(3):107523, 2020.
- [40] Neena Aloysius and M Geetha. A review on deep convolutional neural networks. In *2017 International Conference on Communication and Signal Processing (ICCSP)*, pages 0588–0592. IEEE, 2017.
- [41] Hyeon-Joong Yoo. Deep convolution neural networks in computer vision: a review. *IEIE Transactions on Smart Processing & Computing*, 4(1):35–43, 2015.
- [42] Liang Zou, Jiannan Zheng, Chunyan Miao, Martin J Mckeown, and Z Jane Wang. 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE Access*, 5:23626–23636, 2017.
- [43] Yang Yang, Xiangwei Zheng, and Feng Yuan. A study on automatic sleep stage classification based on cnn-lstm. In *Proceedings of the 3rd International Conference on Crowd Science and Engineering*, pages 1–5, 2018.
- [44] SS Hanna and N Fry. A comparison of methods of strain determination in rocks from southwest dyfed (pembrokeshire) and adjacent areas. *Journal of Structural Geology*, 1(2):155–162, 1979.
- [45] Norman Fry. Random point distributions and strain measurement in rocks. *Tectonophysics*, 60(1-2):89–105, 1979.
- [46] Mirko Ahmadfaraj, Mirsaleh Mirmohammadi, Peyman Afzal, Amir Bijan Yasrebi, and Emanuel John Carranza. Fractal modeling and fry analysis of the relationship between structures and cu mineralization in saveh region, central iran. *Ore Geology Reviews*, 107:172–185, 2019.
- [47] R Ghasemi, B Tokhmechi, and G Borg. Evaluation of effective factors in window optimization of fry analysis to identify mineralization pattern: Case study of bavanat region, iran. *Journal of Mining and Environment*, 9(1):195–208, 2018.
- [48] Vladimir Lisitsin. Spatial data analysis of mineral deposit point patterns: Applications to exploration targeting. *Ore Geology Reviews*, 71:861–881, 2015.

- [49] Brian D Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, pages 255–266, 1976.
- [50] Gabriel Rosser and T. Cheng. Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. *Applied Spatial Analysis and Policy*, 07 2016.
- [51] Mehrdad Jafari Mamaghani, Mikael Andersson, and Patrik Krieger. Spatial point pattern analysis of neurons using ripley’s k-function in 3d. *Frontiers in neuroinformatics*, 4:9, 2010.
- [52] Leonardo de Oliveira Martins, Aristófanés Corrêa Silva, Anselmo Cardoso De Paiva, and Marcelo Gattass. Detection of breast masses in mammogram images using growing neural gas algorithm and ripley’s k function. *Journal of Signal Processing Systems*, 55(1-3):77–90, 2009.
- [53] Keith T Poole. *Spatial models of parliamentary voting*. Cambridge University Press, 2005.
- [54] Michael D Ward and Kristian Skrede Gleditsch. Location, location, location: An mcmc approach to modeling the spatial context of war and peace. *Political analysis*, 10(3):244–260, 2002.
- [55] Alexander Hohl, Minrui Zheng, Wenwu Tang, Eric Delmelle, and Irene Casas. Spatiotemporal point pattern analysis using ripley’s k function. *Geospatial Data Science: Techniques and Applications*. Taylor & Francis, 2017.
- [56] Heather J Lynch and Paul R Moorcroft. A spatiotemporal ripley’s k-function to analyze interactions between spruce budworm and fire in british columbia, canada. *Canadian Journal of Forest Research*, 38(12):3112–3119, 2008.
- [57] Mahdi Hashemi and AliAsghar Alesheikh. Spatio-temporal analysis of tehran’s historical earthquakes trends. In *Advancing geoinformation science for a changing world*, pages 3–20. Springer, 2011.
- [58] Nikos Pelekis, Babis Theodoulidis, Ioannis Kopanakis, and Yannis Theodoridis. Literature review of spatio-temporal database models. *Knowledge Engineering Review*, 19(3):235–274, 2004.
- [59] Hongjie Liu, Hongzhe Xu, Yu Yan, Zaishang Cai, Tianxu Sun, and Wen Li. Bus arrival time prediction based on lstm and spatial-temporal feature vector. *IEEE Access*, 8:11917–11929, 2020.

- [60] Eric Hsueh-Chan Lu and Chen-Hao Liao. A parking occupancy prediction approach based on spatial and temporal analysis. In *Asian Conference on Intelligent Information and Database Systems*, pages 500–509. Springer, 2018.
- [61] Ben Vermeulen, Andreas Pyka, and Matthias Müller. An agent-based policy laboratory for covid-19 containment strategies, 2020.
- [62] Daniel M Altmann, Daniel C Douek, and Rosemary J Boyton. What policy makers need to know about covid-19 protective immunity. *The Lancet*, 395(10236):1527–1529, 2020.
- [63] Rintu Maria Thomas and Jisha John. A review on cell detection and segmentation in microscopic images. In *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, pages 1–5. IEEE, 2017.
- [64] Benjamin Kesler, Guoliang Li, Alexander Thiemicke, Rohit Venkat, and Gregor Neuert. Automated cell boundary and 3d nuclear segmentation of cells in suspension. *Scientific reports*, 9(1):1–9, 2019.
- [65] Ervin A Tasnadi, Timea Toth, Maria Kovacs, Akos Diosdi, Francesco Pampaloni, Jozsef Molnar, Filippo Piccinini, and Peter Horvath. 3d-cell-annotator: an open-source active surface tool for single-cell segmentation in 3d microscopy images. *Bioinformatics*, 36(9):2948–2949, 2020.
- [66] Heather S Deter, Marta Dies, Courtney C Cameron, Nicholas C Butzin, and Javier Buceta. A cell segmentation/tracking tool based on machine learning. In *Computer Optimized Microscopy*, pages 399–422. Springer, 2019.
- [67] Anindya Gupta, Philip J Harrison, Håkan Wieslander, Nicolas Pielawski, Kimmo Kartasalo, Gabriele Partel, Leslie Solorzano, Amit Suveer, Anna H Klemm, Ola Spjuth, et al. Deep learning in image cytometry: a review. *Cytometry Part A*, 95(4):366–380, 2019.
- [68] Felix Buggenthin, Florian Buettner, Philipp S Hoppe, Max Endeke, Manuel Kroiss, Michael Strasser, Michael Schwarzfischer, Dirk Loeffler, Konstantinos D Kokkaliaris, Oliver Hilsenbeck, et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nature methods*, 14(4):403–406, 2017.

- [69] Giuseppe Arbia, Giuseppe Espa, Diego Giuliani, and Maria Michela Dickson. Effects of missing data and locational errors on spatial concentration measures based on ripley’s k-function. *Spatial Economic Analysis*, 12(2-3):326–346, 2017.
- [70] Joan Negre, Facundo Muñoz, and Juan Antonio Barceló. A cost-based ripley’s k function to assess social strategies in settlement patterning. *Journal of Archaeological Method and Theory*, 25(3):777–794, 2018.
- [71] Boris Delaunay. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7(793-800):1–2, 1934.
- [72] Valerio Lupperger, Carsten Marr, and Prisca Chapouton. Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatiotemporal patterns. *PLoS biology*, 18(12):e3000708, 2020.
- [73] James P LeSage, Manfred M Fischer, and Thomas Scherngell. Knowledge spillovers across europe: Evidence from a poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86(3):393–421, 2007.
- [74] Paul Stapor, Daniel Weindl, Benjamin Ballnus, Sabine Hug, Carolin Loos, Anna Fiedler, Sabrina Krause, Sabrina Hroß, Fabian Fröhlich, and Jan Hasenauer. Pesto: parameter estimation toolbox. *Bioinformatics*, 34(4):705–707, 2018.
- [75] Jesper Møller and Rasmus P Waagepetersen. Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34(4):643–684, 2007.
- [76] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopusää, and Ludolf Erwin Meester. *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media, 2005.
- [77] Rolf Sundberg. Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scandinavian journal of statistics*, 37(4):632–643, 2010.
- [78] Benjamin Ballnus, Sabine Hug, Kathrin Hatz, Linus Görlitz, Jan Hasenauer, and Fabian J Theis. Comprehensive benchmarking of markov chain monte carlo methods for dynamical systems. *BMC systems biology*, 11(1):63, 2017.

- [79] Benjamin Ballnus, Steffen Schaper, Fabian J Theis, and Jan Hasenauer. Bayesian parameter estimation for biochemical reaction networks using region-based adaptive parallel tempering. *Bioinformatics*, 34(13):i494–i501, 2018.
- [80] John Aldrich et al. Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science*, 12(3):162–176, 1997.
- [81] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.
- [82] Carsten Marr, Michael Strasser, Michael Schwarzfischer, Timm Schroeder, and Fabian J Theis. Multi-scale modeling of gmp differentiation based on single-cell genealogies. *The FEBS journal*, 279(18):3488–3500, 2012.
- [83] Mark M Tanaka, Andrew R Francis, Fabio Luciani, and SA Sisson. Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics*, 173(3):1511–1520, 2006.
- [84] Michael GB Blum and Viet Chi Tran. Hiv with contact tracing: a case study in approximate bayesian computation. *Biostatistics*, 11(4):644–660, 2010.
- [85] Radu S Stoica, Anne Philippe, Pablo Gregori, and Jorge Mateu. Abc shadow algorithm: a tool for statistical analysis of spatial patterns. *Statistics and computing*, 27(5):1225–1238, 2017.
- [86] Xing Ju Lee, Markus Hainy, James P McKeone, Christopher C Drovandi, and Anthony N Pettitt. Abc model selection for spatial extremes models applied to south australian maximum temperature data. *Computational Statistics & Data Analysis*, 128:128–144, 2018.
- [87] Carolin Loos, Carsten Marr, Fabian J Theis, and Jan Hasenauer. Approximate bayesian computation for stochastic single-cell time-lapse data using multivariate test statistics. In *International Conference on Computational Methods in Systems Biology*, pages 52–63. Springer, 2015.

- [88] Emmanuel Klinger, Dennis Rickert, and Jan Hasenauer. pyabc: distributed, likelihood-free inference. *Bioinformatics*, 34(20):3591–3593, 2018.
- [89] Espen Bernton, Pierre E Jacob, Mathieu Gerber, and Christian P Robert. Approximate bayesian computation with the wasserstein distance. *arXiv preprint arXiv:1905.03747*, 2019.
- [90] Paul Fearnhead and Dennis Prangle. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012.
- [91] Tom Serge Weber, Irene Jaehnert, Christian Schichor, Michal Or-Guil, and Jorge Carneiro. Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLoS Comput Biol*, 10(7):e1003616, 2014.
- [92] Paul Macklin and Mary Elizabeth Edgerton. Agent-based cell modeling: application to breast cancer. In *Multiscale Modeling of Cancer: An Integrated Experimental and Mathematical Modeling Approach*, pages 206–234. Cambridge University Press, 2010.
- [93] Jörn Starruß, Walter de Back, Lutz Brusch, and Andreas Deutsch. Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. *Bioinformatics*, 30(9):1331–1332, 2014.
- [94] François Graner and James A Glazier. Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical review letters*, 69(13):2013, 1992.
- [95] Craig W Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34, 1987.
- [96] Roland Bouffanais. *Design and control of swarm dynamics*. Springer, 2016.
- [97] Jin-Hyung Jun, Dong-Wook Lee, and Kwee-Bo Sim. Realization of cooperative strategies and swarm behavior in distributed autonomous robotic systems using artificial immune system. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, volume 6, pages 614–619. IEEE, 1999.

- [98] Lawrence J Shimkets and Dale Kaiser. Induction of coordinated movement of myxococcus xanthus cells. *Journal of Bacteriology*, 152(1):451–461, 1982.
- [99] Dongsheng Jiang, Simon Christ, Donovan Correa-Gallegos, Pushkar Ramesh, Shruthi Kalgudde Gopal, Juliane Wannemacher, Christoph H Mayr, Valerio Lupperger, Qing Yu, Haifeng Ye, et al. Injury triggers fascia fibroblast collective cell migration to drive scar formation through n-cadherin. *Nature Communications*, 11(1):1–13, 2020.
- [100] Valerio Lupperger, Felix Buggenthin, Prisca Chapouton, and Carsten Marr. Image analysis of neural stem cell division patterns in the zebrafish brain. *Cytometry Part A*, 93(3):314–322, 2018.
- [101] Oriol Viader-Llargués, Valerio Lupperger, Laura Pola-Morell, Carsten Marr, and Hernán López-Schier. Live cell-lineage tracing and machine learning reveal patterns of organ regeneration. *ELife*, 7:e30823, 2018.
- [102] Dominik Jens Elias Waibel, Sayedali Shetab Boushehri, and Carsten Marr. Instantdl—an easy-to-use deep learning pipeline for image segmentation and classification. *bioRxiv*, 2020.
- [103] Viv Bewick, Liz Cheek, and Jonathan Ball. Statistics review 7: Correlation and regression. *Critical care*, 7(6):451, 2003.
- [104] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial point patterns: methodology and applications with R*. CRC press, 2015.
- [105] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- [106] Aurelie C Lozano, Hongfei Li, Alexandru Niculescu-Mizil, Yan Liu, Claudia Perlich, Jonathan Hosking, and Naoki Abe. Spatial-temporal causal modeling for climate change attribution. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 587–596, 2009.
- [107] Ammar Jalalimanesh, Hamidreza Shahabi Haghghi, Abbas Ahmadi, and Madjid Soltani. Simulation-based optimization of radiotherapy: Agent-based modeling and reinforcement learning. *Mathematics and Computers in Simulation*, 133:235–248, 2017.

- [108] Damien Arnol, Denis Schapiro, Bernd Bodenmiller, Julio Saez-Rodriguez, and Oliver Stegle. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell reports*, 29(1):202–211, 2019.
- [109] Haoyang Mi, Chang Gong, Jeremias Sulam, Elana J Fertig, Alexander S Szalay, Elizabeth M Jaffee, Vered Stearns, Leisha A Emens, Ashley M Cimino-Mathews, and Aleksander S Popel. Digital pathology analysis quantifies spatial heterogeneity of cd3, cd4, cd8, cd20, and foxp3 immune markers in triple-negative breast cancer. *Frontiers in physiology*, 11, 2020.
- [110] Muhammad Khalid Khan Niazi, Anil V Parwani, and Metin N Gurcan. Digital pathology and artificial intelligence. *The lancet oncology*, 20(5):e253–e261, 2019.
- [111] Muhammad Shaban, Ruqayya Awan, Muhammad Moazam Fraz, Ayesha Azam, Yee-Wah Tsang, David Snead, and Nasir M Rajpoot. Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE Transactions on Medical Imaging*, 2020.
- [112] Jinjin Hai, Hongna Tan, Jian Chen, Minghui Wu, Kai Qiao, Jingbo Xu, Lei Zeng, Fei Gao, Dapeng Shi, and Bin Yan. Multi-level features combined end-to-end learning for automated pathological grading of breast cancer on digital mammograms. *Computerized Medical Imaging and Graphics*, 71:58–66, 2019.
- [113] Gabriele Campanella, Vitor Werneck Krauss Silva, and Thomas J Fuchs. Terabyte-scale deep multiple instance learning for classification and localization in pathology. *arXiv preprint arXiv:1805.06983*, 2018.
- [114] Maisa Daoud and Michael Mayo. A survey of neural network-based cancer prediction models from microarray data. *Artificial intelligence in medicine*, 97:204–214, 2019.
- [115] Jérôme Galon, Anne Costes, Fatima Sanchez-Cabo, Amos Kirilovsky, Bernhard Mlecnik, Christine Lagorce-Pagès, Marie Tosolini, Matthieu Camus, Anne Berger, Philippe Wind, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, 313(5795):1960–1964, 2006.

- [116] Souptik Barua, Penny Fang, Amrish Sharma, Junya Fujimoto, Ignacio Wistuba, Arvind UK Rao, and Steven H Lin. Spatial interaction of tumor cells and regulatory t cells correlates with survival in non-small cell lung cancer. *Lung Cancer*, 117:73–79, 2018.
- [117] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.
- [118] Valentine Svensson, Sarah A Teichmann, and Oliver Stegle. Spatialde: identification of spatially variable genes. *Nature methods*, 15(5):343–346, 2018.
- [119] Eric Shifrut, Julia Carnevale, Victoria Tobin, Theodore L Roth, Jonathan M Woo, Christina T Bui, P Jonathan Li, Morgan E Diolaiti, Alan Ashworth, and Alexander Marson. Genome-wide crispr screens in primary human t cells reveal key regulators of immune function. *Cell*, 175(7):1958–1971, 2018.
- [120] Kristen E Maynard, Leonardo Collado-Torres, Lukas M Weber, Cedric Uytingco, Brianna K Barry, Stephen R Williams, Joseph L Catallini, Matthew N Tran, Zachary Besich, Madhavi Tippani, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *bioRxiv*, 2020.
- [121] Konstantin Carlberg, Marina Korotkova, Ludvig Larsson, Anca I Catrina, Patrik L Ståhl, and Vivianne Malmström. Exploring inflammatory signatures in arthritic joint biopsies with spatial transcriptomics. *Scientific reports*, 9(1):1–10, 2019.
- [122] Chiara Baccin, Jude Al-Sabah, Lars Velten, Patrick M Helbling, Florian Grünschläger, Pablo Hernández-Malmierca, César Nombela-Arrieta, Lars M Steinmetz, Andreas Trumpp, and Simon Haas. Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nature cell biology*, pages 1–11, 2019.
- [123] Ed Lein, Lars E Borm, and Sten Linnarsson. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*, 358(6359):64–69, 2017.

- [124] Charlotte Giesen, Hao AO Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J Schüffler, Daniel Grolimund, Joachim M Buhmann, Simone Brandt, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature methods*, 11(4):417–422, 2014.
- [125] Emma Lundberg and Georg HH Borner. Spatial proteomics: a powerful discovery tool for cell biology. *Nature Reviews Molecular Cell Biology*, 20(5):285–302, 2019.
- [126] Grace C Bingham, Fred Lee, Alexandra Naba, and Thomas H Barker. Spatial-omics: Novel approaches to probe cell heterogeneity and extracellular matrix biology. *Matrix Biology*, 2020.
- [127] Anna S Nam, Ronan Chaligne, and Dan A Landau. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nature Reviews Genetics*, pages 1–16, 2020.
- [128] Manish Sapkota, Xiaoshuang Shi, Fuyong Xing, and Lin Yang. Deep convolutional hashing for low-dimensional binary embedding of histopathological images. *IEEE journal of biomedical and health informatics*, 23(2):805–816, 2018.
- [129] Nick Jagiella, Dennis Rickert, Fabian J Theis, and Jan Hasenauer. Parallelization and high-performance computing enables automated statistical inference of multi-scale models. *Cell Systems*, 4(2):194–206, 2017.
- [130] Riccardo Guidotti, Anna Monreale, Stan Matwin, and Dino Pedreschi. Black box explanation by learning image exemplars in the latent feature space. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 189–205. Springer, 2019.
- [131] Philipp Eulenberg, Niklas Köhler, Thomas Blasi, Andrew Filby, Anne E Carpenter, Paul Rees, Fabian J Theis, and F Alexander Wolf. Reconstructing cell cycle and disease progression using deep learning. *Nature communications*, 8(1):1–6, 2017.
- [132] Gregory P Way and Casey S Greene. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. *BioRxiv*, page 174474, 2017.

- [133] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*, 2017.
- [134] Davide Castelvechi. Ai copernicus' discovers' that earth orbits the sun. *Nature*, 575(7782):266–268, 2019.



## Appendix A

# Image Analysis of Neural Stem Cell Division Patterns in the Zebrafish Brain.

Lupperger, V., Buggenthin, F., Chapouton, P., Marr, C. (2018). **Image analysis of neural stem cell division patterns in the zebrafish brain.** *Cytometry Part A*, 93(3), 314-322.

# Image Analysis of Neural Stem Cell Division Patterns in the Zebrafish Brain

Valerio Lupperger,<sup>1</sup> Felix Buggenthin,<sup>1</sup> Prisca Chapouton,<sup>2\*</sup> Carsten Marr<sup>1\*</sup>

<sup>1</sup>Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

<sup>2</sup>Research Unit Sensory Biology and Organogenesis, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

Received 30 May 2016; Revised 30 August 2017; Accepted 6 September 2017

\*Correspondence to: Carsten Marr; Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. E-mail: carsten.marr@helmholtz-muenchen.de or Prisca Chapouton; Research Unit Sensory Biology and Organogenesis, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. E-mail: chapouton@helmholtz-muenchen.de

Published online 10 November 2017 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.23260

© 2017 The Authors. Cytometry Part A published by Wiley Periodicals, Inc. on behalf of ISAC.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## • Abstract

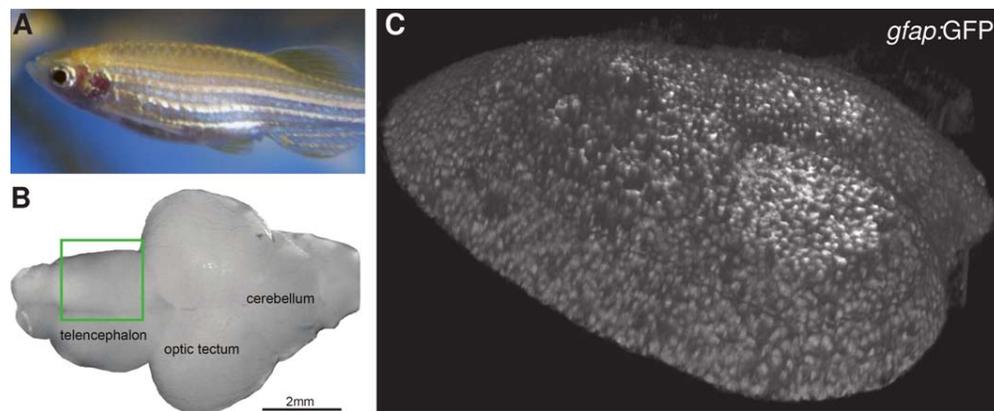
Proliferating stem cells in the adult body are the source of constant regeneration. In the brain, neural stem cells (NSCs) divide to maintain the stem cell population and generate neural progenitor cells that eventually replenish mature neurons and glial cells. How much spatial coordination of NSC division and differentiation is present in a functional brain is an open question. To quantify the patterns of stem cell divisions, one has to (i) identify the pool of NSCs that have the ability to divide, (ii) determine NSCs that divide within a given time window, and (iii) analyze the degree of spatial coordination. Here, we present a bioimage informatics pipeline that automatically identifies GFP expressing NSCs in three-dimensional image stacks of zebrafish brain from whole-mount preparations. We exploit the fact that NSCs in the zebrafish hemispheres are located on a two-dimensional surface and identify between 1,500 and 2,500 NSCs in six brain hemispheres. We then determine the position of dividing NSCs in the hemisphere by EdU incorporation into cells undergoing S-phase and calculate all pairwise NSC distances with three alternative metrics. Finally, we fit a probabilistic model to the observed spatial patterns that accounts for the non-homogeneous distribution of NSCs. We find a weak positive coordination between dividing NSCs irrespective of the metric and conclude that neither strong inhibitory nor strong attractive signals drive NSC divisions in the adult zebrafish brain. © 2017 The Authors. Cytometry Part A published by Wiley Periodicals, Inc. on behalf of ISAC.

## • Key terms

bioimage informatics; neural stem cells; zebrafish brain; cell identification; neuroscience image computing

**THE** analysis of spatial patterns is prevalent in distinct disciplines such as ecology (1,2), geostatistics (3), and developmental biology (4). Adult neural stem cells (NSCs) have been found at specific locations within the brain of adult vertebrates [reviewed in (5,6)], but the spatiotemporal regulation of NSC quiescence and proliferation is only roughly understood. To investigate dividing NSCs and the maintenance of stem cell pools, the zebrafish is an ideal model organism due to its high neurogenic activity, its accessibility for imaging, and the availability of transgenic lines.

In the adult zebrafish pallium (the dorsal telencephalon, see Fig. 1), the ventricular zone containing the NSCs is located on the surface, as a result of an eversion during development (7). The dorsal domain is a spatially clearly defined two-dimensional (2D) surface at the border of the telencephalon. NSCs within this surface repopulate mature neurons and glial cells via constant proliferation and differentiation (8,9). This situation has been utilized in *in vivo* imaging studies to follow the fate of labeled NSCs (10,11) over several weeks. To describe the spatial organization of all cycling NSCs within the pool of NSCs, we used whole mount preparations and labeled S-phase NSCs within three-dimensional (3D) images of the ventricular zone.



**Figure 1.** NSCs in the zebrafish telencephalon. (A) In all experiments, four-month-old zebrafish of a *gfap:GFP* transgenic strain were used, where green fluorescent protein (GFP) is expressed under the control of *gfap* enhancer elements. Average length of adult fish is ~3 cm (B) Top view on a zebrafish brain showing the telencephalon, optic tectum, and cerebellum. We image one hemisphere of the telencephalon (marked with a green rectangle). Scale bar: 2 mm. (C) Reconstructed 3D image stack from confocal microscopy. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Identification and segmentation of single cells in 3D image stacks is a challenging problem for quantitative bio-imaging. To identify single nuclei in 3D, several methods were recently proposed (12–17) that rely on nuclear staining. Such automated methods normally start with separation of background and foreground, followed by the identification of single objects, using, for example, k-means (12), water shedding (13), or graph-cut segmentation (14). Notably, the application of available methods to a specific data set requires adaptation and manual fine-tuning of the parameters. For neural neuronal cells, Schmitz et al. (18) recently stated that available methods fail to properly identify single cells in 3D.

Here, we present a single-cell identification pipeline (SCIP) that explicitly uses prior knowledge on the organization of NSCs in the zebrafish brain. It exploits the fact that NSCs in the zebrafish brain are located on a 2D surface to accurately identify them in 3D. A polynomial regression model as approximation to the hemisphere surface improves the identification and is used to remove imaging artifacts. We apply SCIP to six 3D image stacks of adult zebrafish hemispheres, automatically identify thousands of NSCs, and apply three different metrics to determine distances between all pairs of cells. Within the six hemispheres, we then locate stem cells in S-Phase labeled by the incorporated thymidine analogue EdU. To assess a possible interaction between the dividing cells quantitatively, we evaluate and later fit a simple interaction model and find a weak positive coordination of S-Phase NSCs.

## MATERIALS AND METHODS

### Animal Maintenance

Zebrafish (*Danio rerio*) were kept in the fish facility of the Helmholtz Zentrum München at 28°C with a light/dark cycle of 14/10 h. We used four-month old zebrafish (Fig. 1A) of the transgenic line *gfap:GFP* (19) in an AB background. Experiments were performed in accordance with the

regulations of the Regierung von Oberbayern on animal welfare (Animal protocol 55.2-1-54-2531-83-14).

### Sample Preparation and Image Acquisition

In the transgenic *gfap:GFP* zebrafish strain (19) NSCs are fluorescently labeled with GFP under the control of *gfap* enhancer. Dividing cells were labeled by intraperitoneal injection of the thymidine analogue 5-ethynyl-2-deoxyuridine (EdU, 1 mg/ml, 5  $\mu$ l/0.1 g body weight), which incorporates into replicating DNA, one hour before killing the animals and brain fixation. Zebrafish were over-anesthetized and killed in 0.1% buffered MS222, the brains dissected and fixed overnight in 4% PFA. After blocking in 10% normal goat serum (Sigma), EdU was revealed by binding to azide-Alexa Fluor 555 through a click reaction (Invitrogen). Brains were mounted in Vectashield medium (Vector Laboratories) between two coverslips separated by parafilm spacers.

An inverted confocal laser scanning microscope (Leica SP5) with a 20 $\times$  glycerol immersion objective (HC PL APO 20 $\times$ /0.70 IMM CS), which corrects for field curvature astigmatism, was used for image acquisition. The field of view covers one hemisphere of the pallium (dorsal telencephalon) nearly completely (see Figs. 1B and 1C). All images were taken with 2048  $\times$  2048 pixel in *x-y* direction with a pixel size of 0.38  $\times$  0.38  $\mu$ m. Resolution in *z* direction differed between the 6 hemispheres (Experiment 1: 49 and 62 slices with distance of 2.0  $\mu$ m, Experiment 2: 72 and 84 slices with distance of 2.2  $\mu$ m, Experiment 3: 105 and 85 slices with distance of 1.3  $\mu$ m) and was adapted to brain size. Visual inspection confirms that aberration effects are minimal and do not impinge on cell identification.

### Single-Cell Identification Pipeline

Starting from a 3D image stack (Fig. 2A) SCIP projects the maximum intensity of every *x-y* coordinate in *z*-direction to a 2D image (Fig. 2B). A Laplace of Gaussian (LoG) filter followed by a local maximum/minimum search is used to identify cell sized blobs of approximately  $b = 8 \mu$ m in

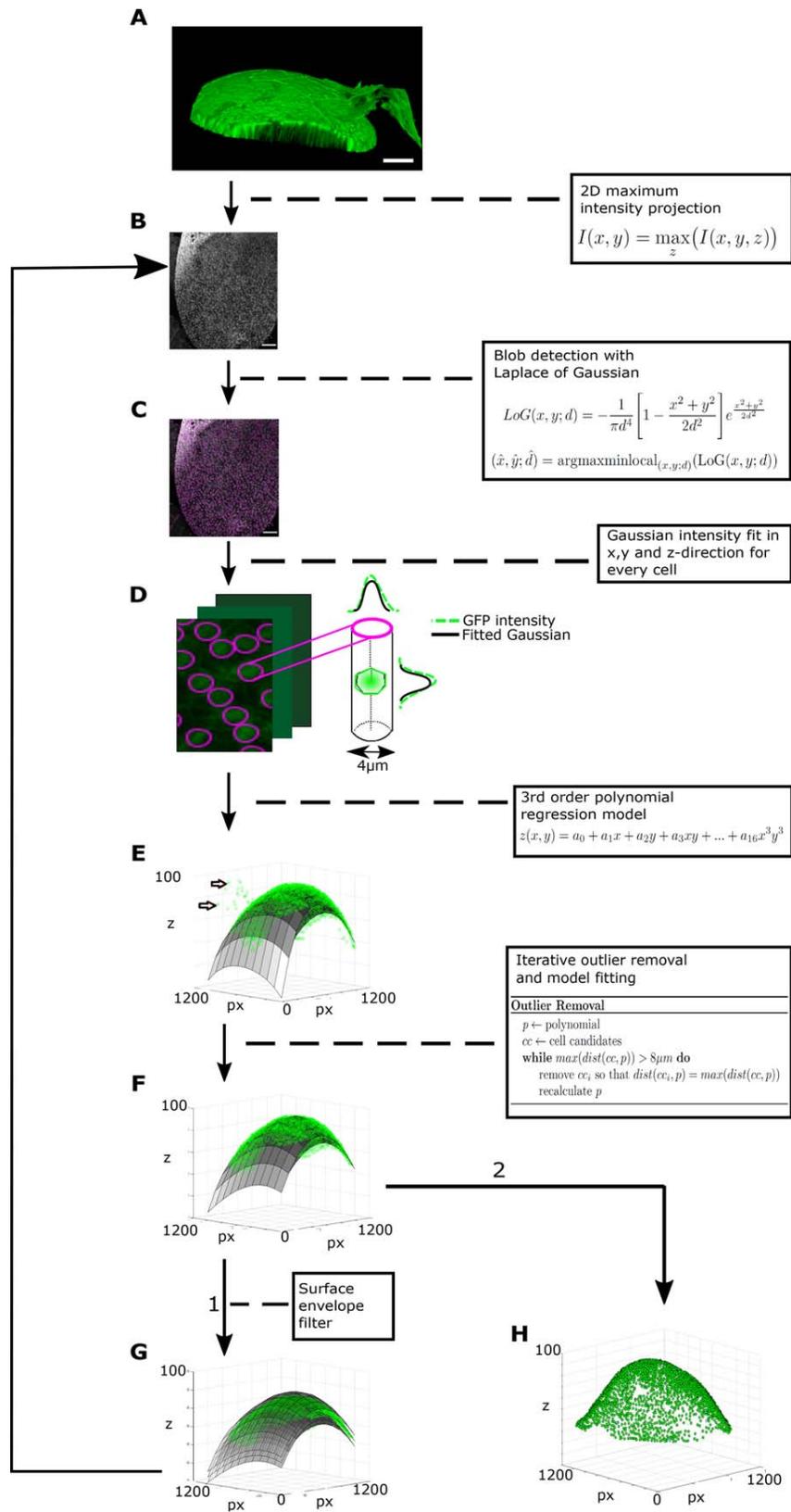
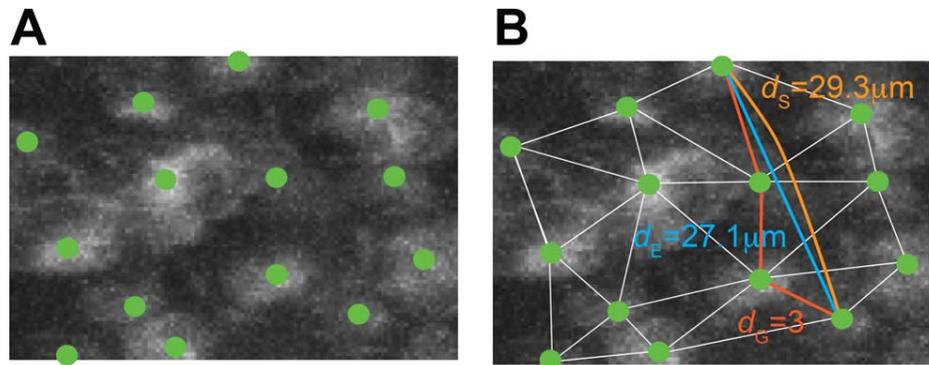


Figure 2.



**Figure 3.** Three metrics are defined on the hemispheres. NSCs are identified (A) and then used to calculate the Euclidean distance  $d_E$  (blue line), the surface distance  $d_S$  (orange line) and the graph distance  $d_G$  (red lines) between all pairs of identified cells on the 2D surface (B). For the surface distance, the shortest path between two cells in 3D is projected on the surface fitted with SCIP. For the graph distance, the shortest path in a network derived from a Delaunay triangulation is calculated. Identified NSCs are shown in green on top of the *gfap*:GFP signal. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

diameter as possible cell candidates (Fig. 2C). From every detected blob, the  $x$ - $y$  centroid is used for the determination of the  $z$ -coordinate. To discard possible signals from neighboring cells, we use pixel information inside a cylinder with a diameter of  $4 \mu\text{m}$  around the centroids (Fig. 2D) and fit a Gaussian curve to the mean intensity profile along  $z$ . This allows to filter out false positives due to contaminations during sample preparation. The argument of the maximum (arg max) of the fitted Gaussian is used as the  $z$ -coordinate of the cell centroid. Additionally, its standard deviation is used to filter real cells from image artifacts: Single pixel errors show a strong intensity and a small standard deviation. We thus filter out objects with a standard deviation in  $z$  below  $1.5 \mu\text{m}$ . Analogously, a 2D Gaussian distribution is fitted in  $x$ - $y$  direction and objects with a standard deviation below  $2 \mu\text{m}$  are excluded.

As NSCs are only found on the edge of a hemisphere we fit a third order polynomial to all 3D centroids to roughly approximate the hemisphere surface and avoid overfitting of single cells. The fitted model is used to exclude outliers by repetitively removing the most distant cell candidate (see arrows in Fig. 2E) and recalculating the polynomial model. The procedure is repeated until all cells are closer than  $16 \mu\text{m}$  ( $\sim 2$  cell diameters) to the polynomial surface (Fig. 2F).

As cells from the adjacent hemisphere, and image contaminations from the 3D image stack may overlap with real cells in the 2D projection, we introduce an envelope in  $z$ -direction  $20 \mu\text{m}$  above and below the polynomial (Fig. 2G). All pixels outside this envelope are set to background intensity, determined via the median of the eight corner pixels of the 3D stack. Steps

(B) - (F) of the pipeline are then repeated on the filtered image stack. The resulting 3D centroids (Fig. 2H) represent NSCs on the zebrafish hemisphere and are used for further analyses.

### Distance Measures

We define distances between pairs of cells using three metrics (Fig. 3): (i) The Euclidean distance  $d_E$  between two points in the 3D space in  $\mu\text{m}$ . (ii) The surface distance  $d_S$  is the projection of the shortest path between two cells on the polynomial in  $\mu\text{m}$ . (iii) The graph distance  $d_G$  between two cells is the number of edges connecting the cells in the shortest path of a graph consisting of nodes (the NSCs identified) and edges between them (see Fig. 3B) as calculated with a Delaunay triangulation (20). The shortest path is determined using the Floyd-Warshall algorithm (21,22).

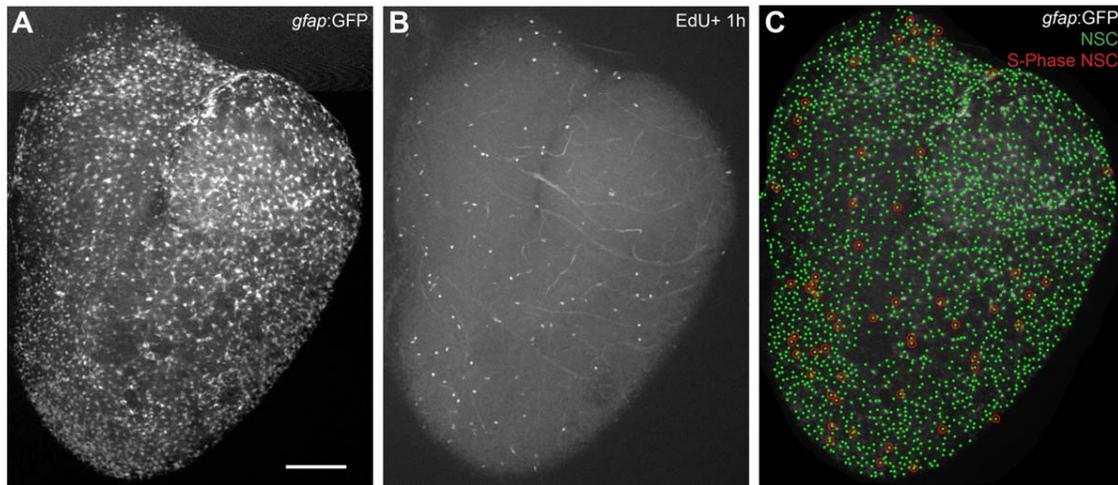
### Identification of S-Phase NSCs

We gave a pulse of EdU 1 h prior to brain dissection, fixation, fluorescent staining and imaging. Cells with a fluorescent EdU signal (see Fig. 4) were automatically identified with SCIP. For optimal data quality, we manually verified S-phase NSCs as EdU positive cells that are also positive for *gfap*:GFP.

### Influence Model

From the observed patterns of S-phase NSCs (Fig. 4), we ask whether specific rules underlie the spatial arrangement of division events. To quantify the distribution of S-phase NSCs within the set of all NSCs, we introduce  $P_i$ , the probability that an NSC  $i$  is in S-phase at a particular time point, as a function of S-phase NSCs in the neighborhood of cell  $i$ . In the case of no influence between division events, this probability is the same for all NSCs. To simulate a pattern with S-phase

**Figure 2.** SCIP for NSCs in the adult zebrafish brain. Raw 3D data (A) is transformed into 2D images (B) via 2D maximum intensity projection. Cell somata are touching each other on the surface, without intermediate space. Cell centers display a high GFP intensity and are used for identification. A blob detection using LoG identifies cell candidates (C). A Gaussian curve is fitted to the intensity profile of a cylinder with  $4 \mu\text{m}$  diameter along  $z$  of every cell candidate (D). The mean of the Gaussian is taken as the  $z$ -coordinate of cell candidate centroid. A surface based on a 3rd order polynomial regression model is fitted to all centroids (E). Cells that are further away than two cell diameters ( $\sim 16 \mu\text{m}$ ) are excluded step by step by removing iteratively the most distant outlier and recalculate the surface (F). To remove remaining image artifacts an envelope is placed in  $20 \mu\text{m}$  distance around the surface (G). All pixels outside this envelope are set to background intensity. Afterwards the pipeline starts over again at (B) using the filtered image stack without image artifacts. (H) The resulting cell centroids can now be used for further analyses. Scale bars:  $50 \mu\text{m}$ . [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**Figure 4.** Spatial pattern of NSC divisions. NSCs are identified in the *gfap:GFP* channel (A) using SCIP. Cells in S-phase were automatically identified by EdU signal and manually verified (B). S-phase NSCs are identified as cells that appear both in the *gfap:GFP* and in the EdU channel (C). Scale bar: 100  $\mu\text{m}$ . [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

NSCs, we randomly select  $S$  cells from all  $N$  NSCs, where the probability for NSC  $i$  to be selected,  $P_i$ , can be normalized to

$$P_i = 1/N. \quad (1)$$

To account for an attractive or repulsive influence between S-phase NSCs, we scale this probability with a positive parameter  $g$ . The exponent of  $g$  is the number of S-phase NSCs within a distance  $r$  (Fig. 5). The probability  $P_i$  for cell  $i$  to be selected as an S-phase NSC then becomes

$$P_i = g \sum_{j=1, j \neq i}^S I(\text{dist}(i,j) \leq r) / Z \quad (2)$$

where we evaluate the number of S-Phase NSCs within radius  $r$  using the indicator function  $I$ , and

$$Z = \sum_{k=1}^N g \sum_{j=1, j \neq k}^S I(\text{dist}(k,j) \leq r)$$

is the normalization constant. To simulate a pattern with  $S$  S-phase NSCs, we now select  $S$  cells one after the other and update the probabilities  $P_i$  after each selection.

Without any influence between S-phase NSCs,  $g = 1$  and Eq. (2) simplifies to Eq. (1). For  $g < 1$  or  $g > 1$  we induce a repulsive or attractive influence within a distance  $r$ , leading to visually distinct patterns of S-phase NSCs (Fig. 5).

### Parameter Inference

From the division probability [Eq. (2)], we can calculate the likelihood  $L$  for an observed S-phase pattern given the interaction strength  $g$  and interaction radius  $r$ . The log-likelihood for  $S$  cells in S-Phase within the set of  $N$  NSCs is

$$\log L(g, r) = \sum_{i=1}^S \log P_i. \quad (3)$$

We can calculate the sum of log-likelihoods for all observed brain hemispheres and find the maximum likelihood estimate

for our parameters  $g$  and  $r$  by optimizing Eq. (3): First, we evaluated all combinations of  $g$  and  $r$  with  $g \in \{0.1, 0.2, \dots, 4\}$  and  $r \in \{5, 6, \dots, 150\} \mu\text{m}$  for Euclidean distance  $d_E$  and surface distance  $d_S$ , and  $r \in \{1, 2, \dots, 20\}$  cells for graph distance  $d_G$ , respectively. From the parameter combination with the largest  $L$ , we then start a gradient descent optimization algorithm (using MATLAB'S `fmincon` function).

To evaluate our inference algorithm, we repeatedly generated 30 data sets per tested influence parameter combination with different influence strengths and radii and applied our inference algorithm to them (see Fig. 6). In the box plots, the boundary of the box closest to zero indicates the 25th percentile ( $q_1$ ), a black line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile ( $q_3$ ). Whiskers above and below the box include points that are not outliers. Points are considered as outliers if they are bigger than  $q_3 + 1.5(q_3 - q_1)$  or smaller than  $q_1 - 1.5(q_3 - q_1)$ .

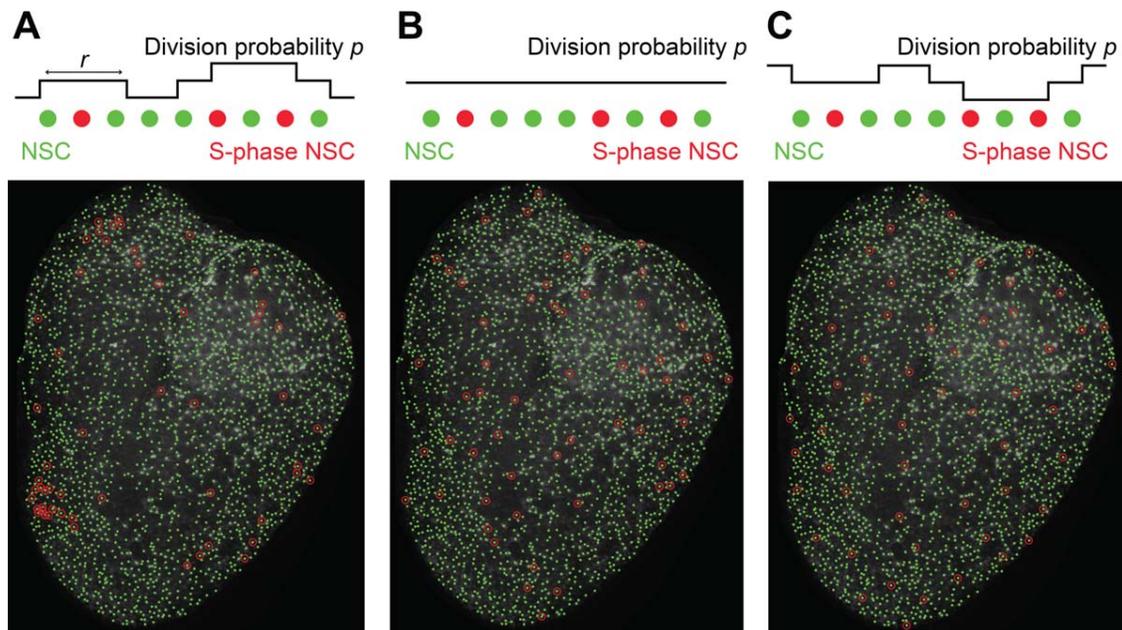
### Implementation

SCIP and the statistical analysis are implemented in MATLAB and performed on a Windows 7 machine with 4 cores (i7-5500 U, 2.4 GHz) and 16 Gigabyte DDR3 memory. The time to identify NSCs in one hemisphere took between 5

**Table 1.** Evaluation of three NSC identification methods

	SCIP	3D OBJECT COUNTER	IMARIS
Precision	<b>93 <math>\pm</math> 3%</b>	91 $\pm$ 5%	92 $\pm$ 4%
Recall	<b>91 <math>\pm</math> 6%</b>	66 $\pm$ 14%	<b>91 <math>\pm</math> 3%</b>
$F_1$ score	<b>92 <math>\pm</math> 2%</b>	76 $\pm$ 9%	91 $\pm$ 1%

NSCs have been counted manually (by P.C.) in 5 different regions in the 2D maximum intensity projection of three zebrafish hemispheres (see Fig. 7). We show precision, recall, and  $F_1$  score (mean  $\pm$  s.d.,  $n = 5$  regions) for SCIP, the 3D object counter plugin in ImageJ, and Imaris. Best mean values are indicated in bold.



**Figure 5.** Simulated attractive spatial influence (A), no spatial influence (B), and repulsive spatial influence (C) of S-phase NSCs result in visually distinct patterns. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

and 10 min. Parameter estimation for one hemisphere took between 3 and 8 min. Code available at <https://github.com/QSCD/SCIP>.

## RESULTS

### Identification of Neural Stem Cells Using SCIP

SCIP can be used to identify single NSCs on the 2D hemisphere surface. We evaluate the quality of cell identification via SCIP by comparing automatically derived cell counts of NSCs in the 2D maximum intensity projection in five different regions of three hemispheres to those generated manually by a human expert on a 2D maximum intensity projection (exemplarily shown as green circles in Fig. 7). We manually count 94, 84, 118, 136, and 160 cells in the five regions. SCIP identifies 85, 92, 123, 125, and 151 NSCs in the same five regions. Pairs of manual (“true”) and automatically identified NSCs were matched according to their  $x$ - $y$  coordinate, which allowed us to calculate precision (the fraction of true NSCs within the set of NSCs identified by SCIP, Fig. 7A), recall (the fraction of NSCs identified by SCIP within all true NSCs) and the  $F_1$  score [an accuracy measure that considers both precision and recall (23)]. We compare SCIP to two other standard approaches for cell identification and segmentation: The “3D object counter” plugin (24) in ImageJ (25) where we carefully chose the intensity threshold as 75 to optimally match our manual counts, and the commercial Imaris software (version 8.4.1, Bitplane, Zurich, Switzerland; Fig. 7C), with target cell diameter of  $5 \mu\text{m}$  and an intensity threshold of 7.14.

The 3D object counter identifies only a fraction of the manually annotated cells (recall  $66 \pm 14\%$ , mean  $\pm$  s.d.,  $n = 5$  regions). Cells that are close to each other are often identified

as large single objects. SCIP and Imaris were able to identify NSCs with the same recall but SCIP identifies less false positive cells resulting in a slightly higher precision and  $F_1$  score compared to Imaris (see Table 1). Overall, we expect that the manual counts underestimate the number of NSCs slightly due to the inherent unidentifiability of nearby cells with a small shift in  $z$  direction in the 2D maximum intensity projection.

### NSC Divisions in the Zebrafish Brain

We apply SCIP to six zebrafish hemispheres from three different adult animals. In each hemisphere, we automatically identify NSCs using SCIP. Moreover, we identify S-phase NSCs via a fluorescently labeled EdU incorporation 1 h prior to imaging (see Methods), stained in a second fluorescence channel (Fig. 4).

In six hemispheres, we identify between 1458 and 3195 NSCs (Table 2). Furthermore, we identify between 42 and 67 S-phase NSCs. The ventricular zone contains dividing NSCs [called type II progenitors (8,26)] as well as type III dividing progenitors that do not express *gfap:GFP*. We consider in this study specifically the dividing type II NSCs. The fraction of S-phase NSCs is  $2.6\% \pm 0.9\%$  (mean  $\pm$  s.d.,  $n = 6$  hemispheres). This fraction is low compared to values obtained with PCNA and MCM5 cell cycle markers (11,26) due to EdU labeling of S-phase only within the small time window of one hour. Visually, the S-phase NSCs (marked by red circles in Fig. 4C) form no obvious cluster, but regions in the brain that show a depletion of S-phase NSCs seem to appear.

### Analysis of S-Phase Patterns

To quantify a potential spatial attraction or repulsion of S-phase NSCs, we define a probabilistic spatial model, where

**Table 2.** We identify around 2.5% of S-phase NSCs in each of the six hemispheres

HEMISPHERE ID	NSCS	S-PHASE NSCS	FRACTION OF S-PHASE NSCS
Exp1L	2399	67	2.79%
Exp1R	2046	56	2.74%
Exp2L	3195	46	1.44%
Exp2R	3017	56	1.86%
Exp3L	1458	56	3.84%
Exp3R	1482	42	2.83%

NSCs have been automatically identified using SCIP, S-phase NSCs have been additionally manually verified.

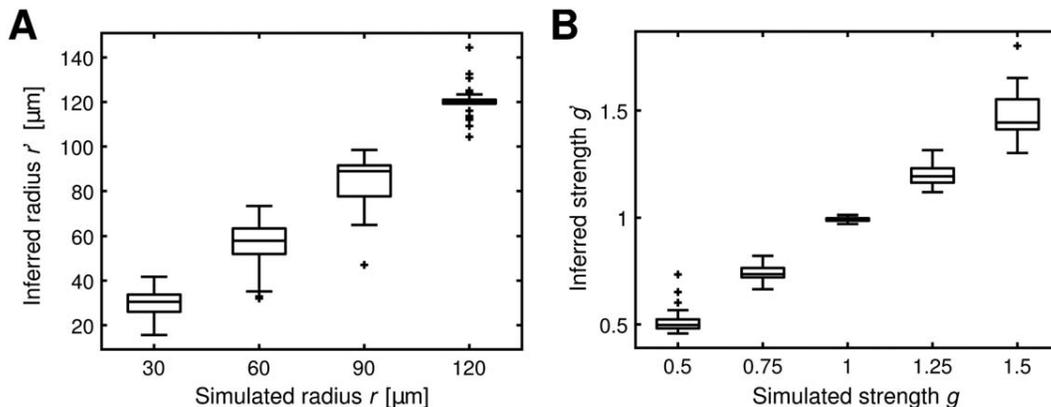
the division probability of NSCs is regulated by two parameters: an interaction radius  $r$ , and the strength  $g$  of the interaction within this radius  $r$ . For  $g = 1$ , the probability to enter S-phase is not changed, but it is increased or decreased for  $g$  above or below 1, respectively (see Fig. 5 and Methods). We simulate patterns of S-phase NSCs (same number as observed) for an attractive (with  $r = 50 \mu\text{m}$  and  $g = 1.6$ ), no influence ( $g = 1$ ), and repulsive ( $r = 50 \mu\text{m}$  and  $g = 0.1$ , see Fig. 5) model. Visually, the three different models can be discriminated clearly via clustering of events (Fig. 5A), spatially random events (Fig. 5B), and events that are evenly spaced across the surface (Fig. 5C).

Formulating the likelihood of the observed spatial pattern of S-phase NSCs [Eq. (3)] allows to infer the most likely parameters  $r$  and  $g$  for the assumed probabilistic model. For simulated data, both parameters are inferred reliably for realistic range of parameters (Fig. 6). Here, we simulated patterns by drawing the same number of S-phase NSCs as observed in the data one after the other, consecutively updating the probabilities  $P_i$  for the next draw based on the chosen parameters  $g$  and  $r$ . We determine the maximum likelihood estimate for the two parameters by maximizing the log-likelihood, which sums over all six hemispheres.

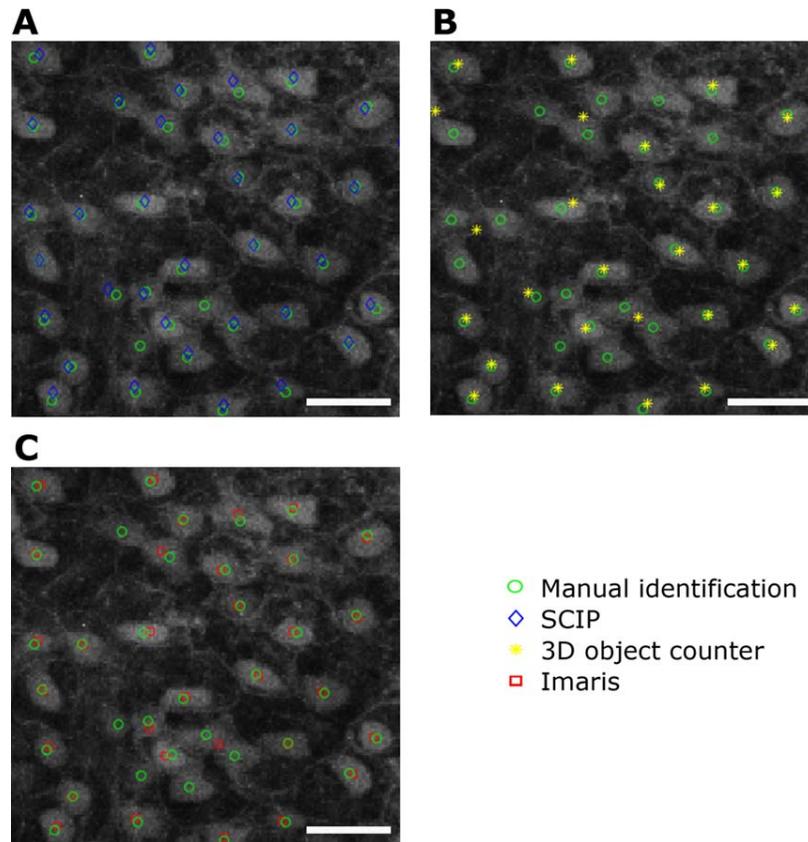
We finally apply our inference algorithm to the S-Phase NSC patterns in the six hemispheres (see Table 2 and Fig. 4). Using the surface distance  $d_s$ , we identify the most likely model with a weak positive influence ( $g = 1.15$ ) between S-phase NSCs and an interaction radius of  $100 \mu\text{m}$ . The most likely model for the Euclidean distance  $d_E$  has an interaction radius of  $89 \mu\text{m}$  with a weak positive influence ( $g = 1.16$ ). Since the surface and Euclidean distances are ignorant of the inhomogeneous density of NSCs on the brain, we also infer the parameters of our probabilistic model for the graph distance  $d_G$ . Here we identify again a weak positive influence ( $g = 1.17$ ) to be most consistent with the observed data with an influence radius of 6 cells (Table 3). Correlating graph distance  $d_G$  with surface distance  $d_s$  on the six hemispheres, we find that a  $d_G = 6$  corresponds to  $d_s = 100 \pm 44 \mu\text{m}$  (mean  $\pm$  s.d.,  $n = 100000$  pairwise distances). The log-likelihoods of these two models are comparable ( $-2467$  vs.  $-2466$ ), which does not allow for model selection or rejection on the basis of the observed data. However, the Euclidean distance model has a considerably smaller log-likelihood ( $-2502$ ) than the other two models suggesting that a metric based on surface and graph distance describes the weak interactions better.

## DISCUSSION

Analyzing the patterns of dividing NSCs in adult zebrafish brains with a probabilistic model, we observe a preference for the positive influence model with a rather large interaction radius of  $100 \mu\text{m}$  or 6 neighboring cells, respectively. The weak influence strength of 1.15 and 1.17, respectively, (a strength of  $g = 1$  corresponds to random patterns) fits to the visual impression that S-phase NSCs are neither strongly clustered, nor particularly regular spaced on the zebrafish hemispheres. Our approach quantifies this impression, by taking the heterogeneous distribution of NSCs into account. Random spatial distributions of cells entering cell cycle have been observed in other systems, such as the ear epidermis (27,28). Individual stochastic cellular behavior resulting in a controlled



**Figure 6.** Model inference on simulated data works for a variety of radii  $r$  based on the surface distance  $d_s$  (A) and strengths  $g$  (B). For each parameter set, we simulated 30 spatial pattern on each of the six hemisphere and find the maximum likelihood estimate [Eq. (3)]. Box boundaries indicate the 25th and 75th percentile, respectively, the black line within the box marks the median. Whiskers above and below the box include points that are not outliers.



**Figure 7.** Evaluation of SCIP. We use manually (by P.C.) identified NSCs in different image regions on the 2D maximum intensity projection as ground truth (green circles). We evaluate SCIP (A), the 3D object counter plugin in ImageJ (B), and Imaris (C) by comparison of the identified cells to the manually detected ones. Scale bar: 30  $\mu\text{m}$ . [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

growth of organs has been also observed in the retina (29), in the intestinal crypts (30), or in *Arabidopsis* sepals (31). The emergence of tissue with regular size and organization from unpredictable individual cell behavior, therefore, seems to be a robust strategy for tissue maintenance. However, the mechanisms regulating the amount of cycling NSCs have to be further investigated. Based on our present analysis, we detect randomness, but we cannot reject more complicated models with, for example, a large heterogeneity in cell cycle times, or a spatiotemporal dependence of NSCs entering cell cycle.

Our single-cell identification pipeline concatenates existing bioimage informatics processing steps and implicitly fits a 2D surface. While for the present study we compared our approach only to two other basic cell identification methods, recently proposed

approaches could help identify single cells in challenging *in vivo* setting [see e.g., Arteta et al. (32) or Cireşan et al. (33)]. In the future, we would like to extend our approach to the segmentation of single cells, supplemented by a morphological analysis that might also allow for functional predictions (34). It would be interesting to see if cells in S-Phase can be inferred from the morphology of the GFP signal alone, and how strong cell positions correlate with morphological features.

A more detailed analysis would require a spatiotemporal analysis of division patterns also with other spatial statistics methods like Ripley's  $K$  (35) or the pair correlation function (36) to test the existence of between-timepoint influence. Overall it will be important in the future to decipher the mechanisms of synchronization of stem cells activity, to understand how groups of cells coordinate their recruitment. In this regard, comparing patterns between young and old animals might help to understand the mechanisms of stem cell depletion during aging.

**Table 3.** Parameter inference on 6 hemispheres finds a weak positive influence between S-phase NSCs for all three models with surface, graph, and Euclidean distance

MODEL METRIC	RADIUS ( $r$ )	STRENGTH ( $g$ )	LOG-LIKELIHOOD
Graph distance ( $d_G$ )	6 (cells)	1.17	-2466
Surface distance ( $d_s$ )	100 ( $\mu\text{m}$ )	1.15	-2467
Euclidean distance ( $d_E$ )	89 ( $\mu\text{m}$ )	1.16	-2502

#### ACKNOWLEDGMENTS

The authors thank Michael Strasser, Carolin Loos, Dennis Rickert and the rest of the QSCD group for helpful comments. Also they thank Jovica Ninkovic and Tamara Durovic for software support.

## AUTHOR CONTRIBUTIONS

V.L. developed the pipeline with F.B. and performed analyses. P.C. performed all experiments. V.L. and C.M. wrote the article with P.C. C.M. and P.C. conceived and supervised the study.

## LITERATURE CITED

- Clark PJ, Evans FC. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 1954;35:445–453.
- Tarnita CE, Bonachela JA, Sheffer E, Guyton JA, Coverdale TC, Long RA, Pringle RM. A theoretical foundation for multi-scale regular vegetation patterns. *Nature* 2017;541:398–401.
- Krige DG. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J South Afr Inst Min Metall* 1951;52:119–139.
- Watanabe M, Kondo S. Is pigment patterning in fish skin determined by the Turing mechanism? *Trends Genet* 2015;31:88–96.
- Chapouton P, Jagasia R, Bally-Cuif L. Adult neurogenesis in non-mammalian vertebrates. *Bioessays* 2007;29:745–757.
- Bond AM, Ming G-L, Song H. Adult mammalian neural stem cells and neurogenesis: Five decades later. *Cell Stem Cell* 2015;17:385–395.
- Mueller T, Wullimann MF. An evolutionary interpretation of teleostean forebrain anatomy. *Brain Behav Evol* 2009;74:30–42.
- März M, Chapouton P, Diotel N, Vaillant C, Hesel B, Takamiya M, Lam CS, Kah O, Bally-Cuif L, Strähle U. Heterogeneity in progenitor cell subtypes in the ventricular zone of the zebrafish adult telencephalon. *Glia* 2010;58:870–888.
- Than-Trong E, Bally-Cuif L. Radial glia and neural progenitors in the adult zebrafish central nervous system. *Glia* 2015;63:1406–1428.
- Barbosa JS, Sanchez-Gonzalez R, Di Giaimo R, Baumgart EV, Theis FJ, Götz M, Ninkovic J. Neurodevelopment. Live imaging of adult neural stem cell behavior in the intact and injured zebrafish brain. *Science* 2015;348:789–793.
- Dray N, Bedu S, Vuillemin N, Alunni A, Coolen M, Krecsmerik M, Supatto W, Beaupaire E, Bally-Cuif L. Large-scale live imaging of adult neural stem cells in their endogenous niche. *Development* 2015;142:3592–3600.
- Chinta R, Wasser M. Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live *Drosophila* embryos. *Cytometry Part A* 2012;81A:52–64.
- Wählby C, Sintorn I-M, Erlandsson F, Borgefors G, Bengtsson E. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *J Microsc* 2004;215:67–76.
- Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng* 2010;57:841–852.
- Mathew B, Schmitz A, Muñoz-Descalzo S, Ansari N, Pampaloni F, Stelzer EHK, Fischer SC. Robust and automated three-dimensional segmentation of densely packed cell nuclei in different biological specimens with Lines-of-Sight decomposition. *BMC Bioinf* 2015;16:187.
- Faure E, Savy T, Rizzi B, Melani C, Stašová O, Fabrèges D, Špir R, Hammons M, Cünderlik R, Recher G, et al. A workflow to process 3D+time microscopy images of developing organisms and reconstruct their cell lineage. *Nat Commun* 2016;7:8674.
- Schmitz A, Fischer SC, Mattheyer C, Pampaloni F, Stelzer EHK. Multiscale image analysis reveals structural heterogeneity of the cell microenvironment in homotypic spheroids. *Sci Rep* 2017;7:43693.
- Schmitz C, Eastwood BS, Tappan SJ, Glaser JR, Peterson DA, Hof PR. Current automated 3D cell detection methods are not a suitable replacement for manual stereologic cell counting. *Front Neuroanat* 2014;8:27.
- Bernardos RL, Raymond PA. GFAP transgenic zebrafish. *Gene Express Patterns* 2006;6:1007–1013.
- Delaunay B. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk.* 1934;7:1–2.
- Floyd RW. Algorithm 97: Shortest path. *Commun ACM* 1962;5:345.
- Warshall S. A Theorem on Boolean matrices. *J ACM* 1962;9:11–12.
- Lever J, Krzywinski M, Altman N. Points of significance: Classification evaluation. *Nat Methods* 2016;13:603–604.
- Bolte S, Cordelières FP. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* 2006;224:213–232.
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671–675.
- Chapouton P, Skupien P, Hesel B, Coolen M, Moore JC, Madelaine R, Kremmer E, Faus-Kessler T, Blader P, Lawson ND, et al. Notch activity levels control the balance between quiescence and recruitment of adult neural stem cells. *J Neurosci* 2010;30:7961–7974.
- Doupé DP, Klein AM, Simons BD, Jones PH. The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Dev Cell* 2010;18:317–323.
- Mesa KR, Kawaguchi K, Gonzalez DG, Cockburn K, Boucher J, Xin T, Klein AM, Greco V. Epidermal stem cells self-renew upon neighboring differentiation. 2017. Available at: <https://doi.org/10.1101/155408>
- He J, Zhang G, Almeida AD, Cayouette M, Simons BD, Harris WA. How variable clones build an invariant retina. *Neuron* 2012;75:786–798.
- Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C, Barker N, Klein AM, van Rheenen J, Simons BD, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* 2010;143:134–144.
- Hong L, Dumond M, Tsugawa S, Sapala A, Routier-Kierzkowska A-L, Zhou Y, Chen C, Kiss A, Zhu M, Hamant O, et al. Variable cell growth yields reproducible organ development through spatiotemporal averaging. *Dev Cell* 2016;38:15–32.
- Arteta C, Lempitsky V, Noble JA, Zisserman A. Learning to detect cells using non-overlapping extremal regions. In: Ayache N, Delingette H, Golland P, Mori K, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*. Vol 7510. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. pp 348–356.
- Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. *Med Image Comput Comput Assist Interv* 2013;16:411–418.
- Buggenthin F, Buettner F, Hoppe PS, Ende M, Kroiss M, Strasser M, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nat Methods* 2017;14:403–406.
- Ripley BD. Modelling spatial patterns. *J R Stat Soc Series B Stat Methodol* 1977;39:172–212.
- Wiegand T, Moloney KA. *Handbook of Spatial Point-Pattern Analysis in Ecology*. Boca Raton, Florida: CRC Press; 2013. 538 p.

## Appendix B

# Live cell-lineage tracing and machine learning reveal patterns of organ regeneration.

Viader-Llargués, O., **Lupperger, V.**, Pola-Morell, L., Marr, C., López-Schier, H. (2018). **Live cell-lineage tracing and machine learning reveal patterns of organ regeneration.** *Elife*, 7, e30823.

This eLife article is distributed under the terms of a Creative Commons Attribution License that permits unrestricted use and redistribution.

# Live cell-lineage tracing and machine learning reveal patterns of organ regeneration

Oriol Viader-Llargués<sup>1,2</sup>, Valerio Lupperger<sup>3</sup>, Laura Pola-Morell<sup>1,2</sup>, Carsten Marr<sup>3\*</sup>, Hernán López-Schier<sup>1,2\*</sup>

<sup>1</sup>Unit Sensory Biology & Organogenesis, Helmholtz Zentrum München, Neuherberg, Germany; <sup>2</sup>Laboratory of Sensory Cell Biology, Centre for Genomic Regulation, Barcelona, Spain; <sup>3</sup>Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany

**Abstract** Despite the intrinsically stochastic nature of damage, sensory organs recapitulate normal architecture during repair to maintain function. Here we present a quantitative approach that combines live cell-lineage tracing and multifactorial classification by machine learning to reveal how cell identity and localization are coordinated during organ regeneration. We use the superficial neuromasts in larval zebrafish, which contain three cell classes organized in radial symmetry and a single planar-polarity axis. Visualization of cell-fate transitions at high temporal resolution shows that neuromasts regenerate isotropically to recover geometric order, proportions and polarity with exceptional accuracy. We identify mediolateral position within the growing tissue as the best predictor of cell-fate acquisition. We propose a self-regulatory mechanism that guides the regenerative process to identical outcome with minimal extrinsic information. The integrated approach that we have developed is simple and broadly applicable, and should help define predictive signatures of cellular behavior during the construction of complex tissues.

DOI: <https://doi.org/10.7554/eLife.30823.001>

**\*For correspondence:**

carsten.marr@helmholtz-muenchen.de (CM);

hernan.lopez-schier@helmholtz-muenchen.de (HL-S)

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

**Funding:** See page 20

**Received:** 27 July 2017

**Accepted:** 28 March 2018

**Published:** 29 March 2018

**Reviewing editor:** Tanya T Whitfield, University of Sheffield, United Kingdom

© Copyright Viader-Llargués et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

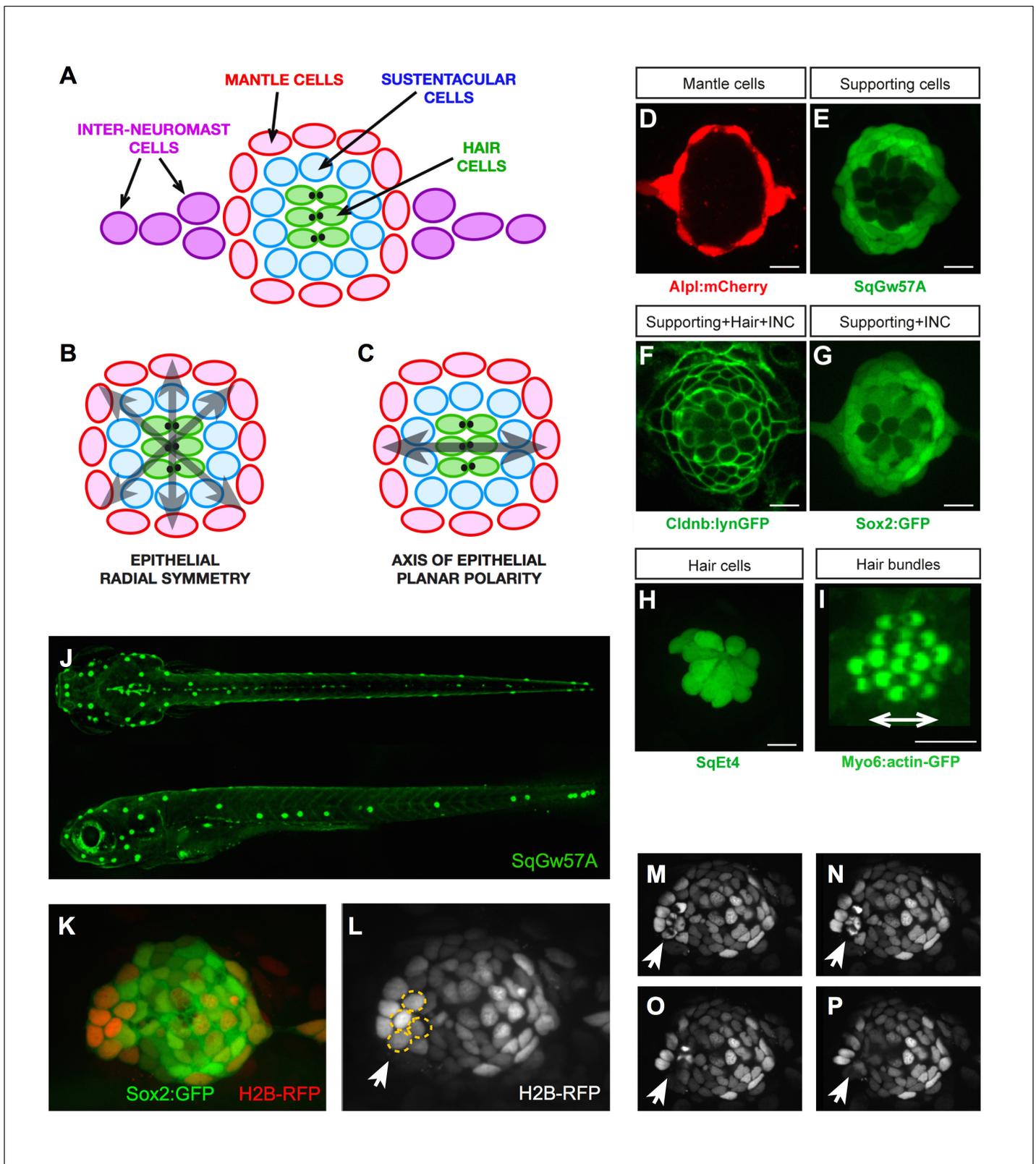
Understanding organogenesis, organ morphostasis and regeneration is crucial to many areas of biology and medicine, including controlled organ engineering for clinical applications ([Lancaster et al., 2013](#); [Boj et al., 2015](#); [Sato and Clevers, 2015](#); [Willyard, 2015](#)). External tissues sustain continuous injury and must recurrently repair to maintain physiological function during the life of the organism ([Levin, 2009](#)). Structural reproducibility depends on the re-establishment of cell identity, number, localization and polarization. Two aspects of organ regeneration are the current focus of intense attention. First, how multiple cells interact to recapitulate organ architecture. Second, what is the mechanism that controls the correct reproduction of cell number and localization. Here we use the neuromasts of the superficial lateral line in larval zebrafish to gain a global perspective on sensory-organ regeneration. The neuromasts are ideally suited for this purpose because they are small and external, facilitating physical access and three-dimensional high-resolution videomicroscopy of every cell during extended periods. We have combined live single-cell tracking, cell-lineage tracing, pharmacological and microsurgical manipulations, and multidimensional data analysis by machine learning to identify features that predict cell-fate decisions during neuromast repair. Our comprehensive approach is simple and model independent, which should facilitate its application to other organs or experimental systems that are accessible to videomicroscopy. It should help reveal the basic rules that underlie how complex structures emerge from the collective behavior of cells.

## Results

### Complete neuromast ablation is irreversible in larval zebrafish

The neuromasts of the superficial lateral line in zebrafish are formed by a circular cuboidal epithelium of 60–70 cells (López-Schier and Hudspeth, 2006; Ghysen and Dambly-Chaudière, 2007; Norden, 2017). Mechanoreceptive hair cells occupy the center of the organ, whereas non-sensory sustentacular supporting cells are found around and between the hair cells (Figure 1A). A second class of supporting cell called mantle cells forms the outer rim of the organ. The invariant spatial distribution of these three cell classes generates a radial symmetry (Figure 1B) (Pinto-Teixeira et al., 2015). Neuromasts also have an axis of planar polarity defined by the orientation of the hair-cells' apical hair bundle (Figure 1C) (Ghysen and Dambly-Chaudière, 2007; Wibowo et al., 2011). In addition to this geometric organization, cell-class number and proportions are largely constant, with around 40 sustentacular, 8–10 mantle, and 14–16 hair cells. Non-sensory cells can proliferate, whereas the sensory hair cells are postmitotic (López-Schier and Hudspeth, 2006; Ma et al., 2008; Cruz et al., 2015; Pinto-Teixeira et al., 2015). Finally, a string of interneuromast cells connects each neuromast along the entire lateral-line system (Figure 1A) (Ghysen and Dambly-Chaudière, 2007). Previous studies have extensively characterized the regeneration of the mechanosensory hair cells (Williams and Holder, 2000; Harris et al., 2003; López-Schier and Hudspeth, 2006; Hernández et al., 2006; Ma et al., 2008; Behra et al., 2009; Faucherre et al., 2009; Wibowo et al., 2011; Namdaran et al., 2012; Steiner et al., 2014; Jiang et al., 2014). However, the regeneration of non-sensory cells remains largely unexplored. To obtain quantitative data of whole sensory-organ regeneration we developed an experimental assay that combines controllable neuromast damage, long-term videomicroscopy at cellular resolution, and live cell-lineage tracing. We used combinations of transgenic lines expressing genetically encoded fluorescent proteins that allow the precise quantification and localization of each cell class in neuromasts, and which also serve as a direct and dynamic readout of tissue organization. This is important because it enables the visualization of cell-fate transitions in living specimens within the growing tissue at high temporal resolution. Specifically, the *Tg[alpl:mCherry]* line expresses cytosolic mCherry in the mantle and interneuromast cells (Figure 1D). The *Et(krt4:EGFP)sqgw57A* (hereafter SqGw57A) expresses cytosolic GFP in sustentacular cells (Figure 1E). The *Tg[-8.0cldnb:LY-EGFP]* (Cldnb:lynGFP) express a plasma-membrane targeted EGFP in the entire neuromast epithelium and in the interneuromast cells (Figure 1F), and the *Tg[Sox2-2a-sfGFP]* (Sox2:GFP) expresses cytosolic GFP in all the supporting cells and the interneuromast cells (Figure 1G). For hair cells, we use *Et(krt4:EGFP)sqet4* (SqEt4) that expresses cytosolic GFP (Figure 1H), or the *Tg(myo6b:actb1-EGFP)* (Myo6b:actin-GFP) that labels filamentous actin (Figure 1I). These transgenic lines have been previously published, but are reproduced here for clarity and self-containment of this work (López-Schier and Hudspeth, 2006; Kondrychyn et al., 2011; Kindt et al., 2012; Shin et al., 2014; Steiner et al., 2014; Pinto-Teixeira et al., 2015).

To induce tissue damage in a controllable and reproducible manner, we used a nanosecond ultraviolet laser beam that was delivered to individual cells through a high numerical-aperture objective, which was also used for imaging. The stereotypic localization of the neuromasts along the zebrafish larva varies only marginally between individuals and during larval growth (Figure 1J) (Ledent, 2002; López-Schier et al., 2004). This permits the unambiguous identification of the manipulated neuromast throughout the experiment, and the comparison between corresponding organs in different animals. Using Sox2:GFP 5 day-old zebrafish larvae that ubiquitously express a nucleus-targeted red-fluorescent protein (H2B-RFP) (Figure 1K–L), we certified that laser-targeted cells are rapidly eliminated from the neuromast epithelium with no detectable collateral damage (Figure 1M–P and Video 1). Having established a well-controlled injury protocol, we decided to probe the limits of neuromast regeneration. We first used specimens co-expressing Alpl:mCherry and Cldnb:lynGFP, which reveal all neuromast cells in green and the mantle cells in red (Figure 2A). We began by ablating entire neuromasts and assessed regeneration for 7 days (Figure 2B–E). Specifically, we looked at the response of flanking interneuromast cells because it has been demonstrated that they can proliferate and generate additional neuromasts, particularly upon loss of ErbB2 signaling (López-Schier and Hudspeth, 2005; Grant et al., 2005; Sánchez et al., 2016). Four hours post-injury (4 hpi) a wound remains evident at the target area (Figure 2B). One day post-injury (1 dpi), the damaged area was occupied by a thread of Alpl:mCherry(+) cells, which based on marker expression are



**Figure 1.** Geometric organization of the neuromast. (A–C) Schematic representation of a neuromast depicting (A) cell classes identifiable by expression of transgenic markers. Grey arrows indicate, respectively, (B) radial symmetry and (C) epithelial planar polarity. (D–I) Confocal images of cell-specific transgenic markers. (D) *Alpl:mCherry* marks mantle and interneuromast cells, (E) *SqGw57A* shows all supporting cells, (F) *Cldnb:lynGFP* marks all neuromast cells, (G) *Sox2:GFP* marks supporting and interneuromast cells, (H) *SqEt4* labels hair cells, and (I) *Myo6:actin-GFP* highlights the planar

*Figure 1 continued on next page*

Figure 1 continued

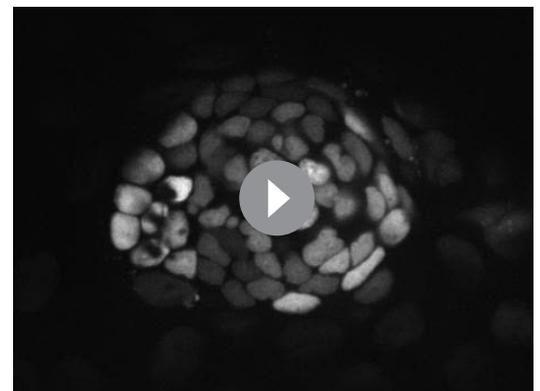
polarization of the hair cells by decorating their apical stereocilia. Scale bars: 10  $\mu\text{m}$ . (J) Images of dorsal (top) and lateral (bottom) views of a SqGw57A transgenic zebrafish larva, revealing the full complement of superficial neuromasts and their stereotypic position. (K) A single confocal section of the lateral view of a neuromast expressing GFP in supporting cells (Sox2-GFP) and a RFP in all nuclei (H2B-RFP). (L) Same neuromast in K showing RFP-marked nuclei. The white arrow indicates 4 cells (circled), which are target of the laser beam for ablation. (M–P) Four still images of the neuromast in L over a period of five minutes, in which the laser-targeted cells are eliminated from the epithelium (white arrow).

DOI: <https://doi.org/10.7554/eLife.30823.002>

likely interneuromast cells (Figure 2C). None of the removed neuromasts regenerated after 7 days ( $n = 22$ ) (Figure 2D–E). We obtained an identical outcome using the independent pan-supporting cell marker Sox2:GFP ( $n = 9$ ) (Figure 2F–J). Finally, incubation of Alpl:mCherry specimens with Bromodeoxy-Uridine (BrdU) to reveal the DNA synthesis that occurs prior to mitosis showed that interneuromast cells do not proliferate after neuromast ablation (Figure 2K–N) (Gratzner, 1982). These data indicate that in contrast to what occurs in embryos (Sánchez et al., 2016), the complete elimination of a neuromast is irreversible in larval zebrafish.

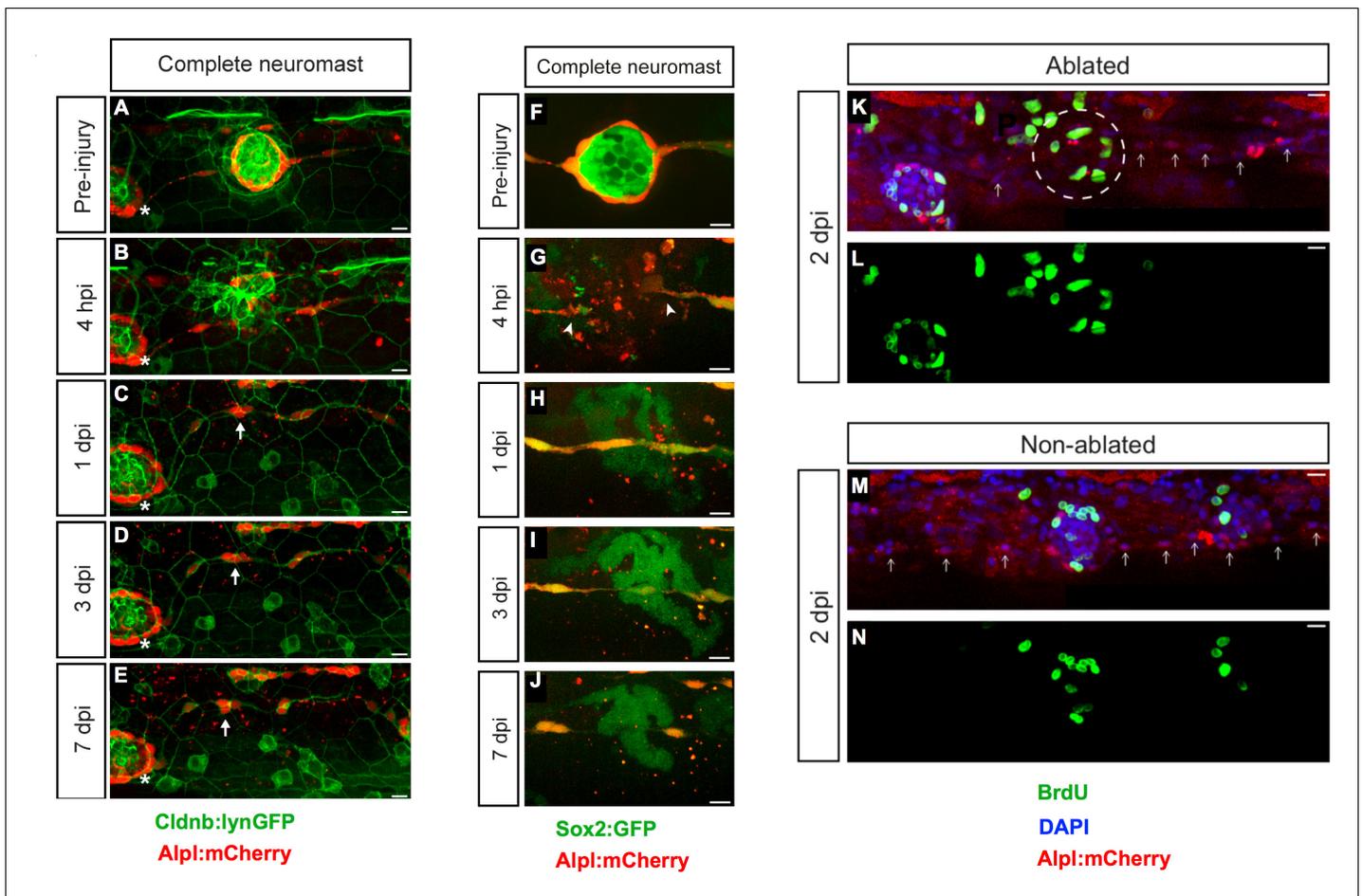
### Neuromasts have isotropic regenerative capacity

To further explore neuromast repair we decided to use milder injury regimes. We systematically produced controlled damage of well-defined scale and location in double transgenic specimens that combine the supporting cell marker Cldnb:lynGFP and the mantle-cell marker Alpl:mCherry (Figure 3A–O). We found that ablation of the posterior half of the neuromast was followed by closure of the wound within 24 hr (Figure 3A–C). At 3 dpi, target neuromasts regained normal cell-class spatial distribution ( $n = 6$ ) (Figure 3D). At 7 dpi, neuromasts recovered approximately 70% of the normal cell number (Figure 3E,Z). We found no difference in speed and extent of regeneration after concurrently ablating the posterior half of neuromasts and flanking interneuromast cells ( $n = 5$ ) (Figure 3F–J,Z). The ablation of the posterior or the dorsal half of the epithelium resulted in identical outcome, suggesting that neuromasts are symmetric in their regenerative capacity ( $n = 6$ ) (Figure 3K–O,Z). Next, we assessed mantle-cell regeneration using a double transgenic line expressing Sox2:GFP and Alpl:mCherry, which reveal mantle cells in red and sustentacular cells in green (Figure 3P–Y). The complete elimination of mantle cells was followed by their re-emergence 3 dpi (Figure 3Q–S), and the reconstitution of the outer rim of the neuromast 7 dpi ( $n = 15$ ) (Figure 3T,Z). The simultaneous ablation of the mantle cells and the adjacent interneuromast cells led to identical outcome ( $n = 6$ ) (Figure 3U–Z). The ablation of the interneuromast cells in fish co-expressing Sox2:GFP and Alpl:mCherry on one side of a neuromast ( $n = 12$ ), or between two adjacent organs ( $n = 8$ ) did not trigger the proliferation of the remaining interneuromast cells over a period of 7 days (Figure 3—figure supplement 1A–J). Because the complete ablation of mantle cells leaves intact the sustentacular-cell population, and the hair cells are postmitotic, these results yield three important and novel findings: (1) interneuromast cells are not essential for neuromast regeneration in larval zebrafish, although they may contribute to mantle cell regeneration; (2) neuromasts have isotropic regenerative capacity; (3) sustentacular cells are tri-potent progenitors able to self-renew and to generate mantle and hair cells.



**Video 1.** A 20 min videomicroscopic recording of a neuromast after laser-mediated ablation of supporting cells. Four laser-targeted cells (showing a dark spot in the nuclei from focal fluorescent-protein bleaching) are eliminated from the epithelium, which closes the wound. There is no noticeable collateral damage. Time resolution is one image per 30 s.

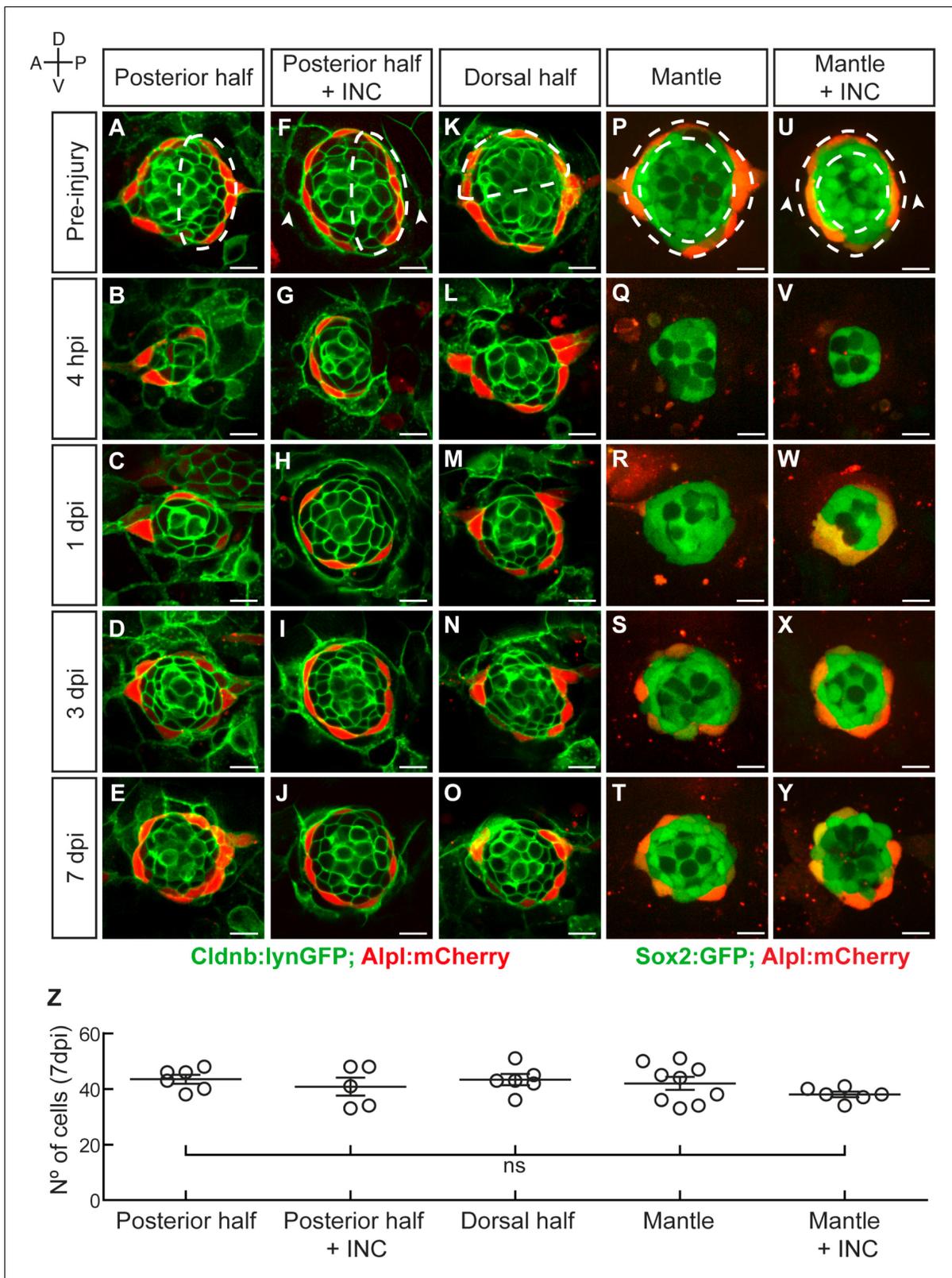
DOI: <https://doi.org/10.7554/eLife.30823.004>



**Figure 2.** Zebrafish larvæ do not regenerate completely-ablated neuromasts. (A–E) Confocal images of a 7 day follow-up of the complete ablation of a neuromast in the double transgenic line *Tg[Cldnb:lynGFP; Alpl:mCherry]*. (A) The site of damage was identified over subsequent days by the position of an intact reference neuromast (white asterisk). (B) Laser-mediated cell ablation produced a wound 4 hours-post-injury (hpi). (C–E) This wound was replaced by a thread of mCherry(+) cells (white arrow) 1 day-post-injury (dpi), which did not change over the subsequent 6 days. (F–J) Confocal images over a 7 day time course after the ablation of a neuromast in the double transgenic line *Tg[Sox2:GFP; Alpl:mCherry]*. Identically to A–E, the complete ablation of the target neuromast results in a thin trail of interneuromast cells (white arrowheads) covering the damaged area (K–N). Scale bars: 10  $\mu$ m. DOI: <https://doi.org/10.7554/eLife.30823.003>

### Neuromast architecture recovers after severe loss of tissue integrity

To test the limits of neuromast regeneration we systematically ablated increasing numbers of cells. Extreme injuries that eliminated all except 1 to 3 cells almost always led to neuromast loss (not shown), whereas ablations that left between 4 and 10 cells, reducing the organ to a combination of 2–3 mantle and 2–7 sustentacular cells, allowed regeneration (**Figure 4A–E,K**). We found that after losing over 95% of their cellular content, neuromasts recover an average of 45 cells at 7 dpi (or approximately 70% of the normal cell count), with exceptional cases reaching 60 cells (equivalent to over 90% of a normal organ) ( $n = 15$ ) (**Figure 4K**). Regenerating neuromasts became radial-symmetric as early as 3 dpi (**Figure 4D**), and had normal cell-class composition and proportions 7 dpi (**Figure 4L–M**). Next, we concurrently ablated 95% of the neuromast and the flanking interneuromast cells (**Figure 4F–G**). This intervention was followed by a similar regeneration process, but led to smaller organs ( $n = 6$ ) (**Figure 4H–J,N–P**). These observations reinforce our previous suggestion that interneuromast cells have a non-essential, yet appreciable contribution to regeneration. Timed quantification of cell-class number and localization showed a reproducible pattern of tissue growth and morphogenesis. During the first 24 hpi, the intact cells rebuilt a circular epithelium (**Figure 4B**). From 1 dpi to 3 dpi, cell number increases rapidly and proportion is restored (**Figure 4C,K–M**). After



**Figure 3.** Neuromasts have isotropic regenerative capacity. (A) Ablation of the posterior half of a neuromast. (B–C) The damage is resolved by cellular movement from the undamaged site 1 dpi. (D) Neuromasts recover geometric order after 3 days and (J) return to homeostasis by 7 dpi. Dashed lines in A, F, K, P, U delineate the ablated area. (F–J) Simultaneous ablation of the posterior half of a neuromast and the interneuromast cells flanking its anterior and posterior sides (n = 5) led to a regeneration outcome identical to that of the experiment in (A–E). Arrowheads in (F) point the location normally

Figure 3 continued on next page

Figure 3 continued

occupied by the interneuromast cells. (K–O) Neuromasts depleted from their dorsal half ( $n = 6$ ) also recover epithelial size, proportions and geometry in a manner indistinguishable from equatorial-side ablation after 7 days. (P–T) 7 days after their complete laser-mediated ablation, mantle cells regenerated for neuromasts to recover the mantle. (U–Y) The ablation of interneuromast cells flanking both sides of neuromasts that were depleted of mantle cells resulted in the same outcome ( $n = 6$ ). (Z) Quantification of the number of cells in regenerated neuromasts at 7 dpi. Number of neuromast cells was no statistically significant between groups of different damage regimes as determined by one-way ANOVA ( $F(4,27)=1.013$ ,  $p=0.4183$ ). Scatter plot shows mean  $\pm$  s.e.m. ns: non-significant. Scale bars: 10  $\mu$ m.

DOI: <https://doi.org/10.7554/eLife.30823.005>

The following figure supplement is available for figure 3:

**Figure supplement 1.** Interneuromast cells do not regenerate.

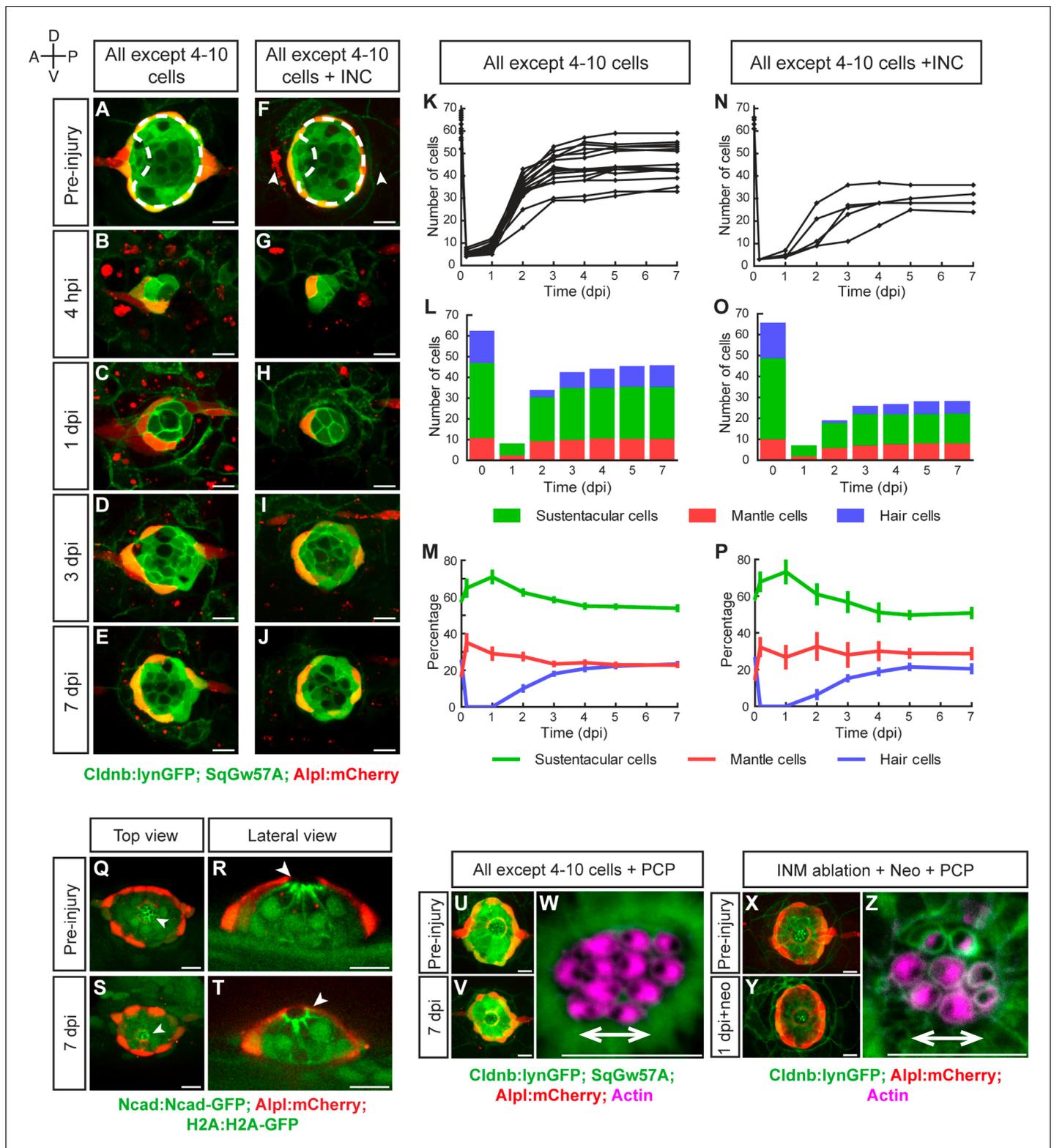
DOI: <https://doi.org/10.7554/eLife.30823.006>

3 dpi, cell number increases at a slower pace (Figure 4K–M). Importantly, each cell class assumes an appropriate position despite a much reduced cell number (Figure 4E, J, L–P).

Next, we examined if the orthogonal polarity axes of the epithelium are re-established after the severest of injuries. To assess tissue apicobasal polarity we used a combination of transgenic lines that allows the observation of the invariant basal position of the nucleus and the apical adherens junctions (Figure 4Q–R) (Ernst et al., 2012; Harding and Nechiporuk, 2012; Hava et al., 2009). We found correct positioning of these markers in the regenerated epithelium ( $n = 4$ ), including the typical apicobasal constriction of the hair cells (Figure 4S–T). To assess epithelial planar polarity, we looked at hair-bundle orientation using fluorescent phalloidin, which revealed that 7 dpi the regenerated neuromasts were plane-polarized in a manner indistinguishable from unperturbed organs, with half of the hair cells coherently oriented in opposition to the other half ( $n = 10$ ) (Figure 4U–W). To test if plane-polarizing cues derive from an isotropic forces exerted by the interneuromast cells that are always aligned to the axis of planar polarity of the neuromast epithelium, we ablated these cells flanking an identified neuromast, and concurrently killed the hair cells with the antibiotic neomycin (Figure 4X–Y). In the absence of interneuromast cells regenerating hair cells recovered normal coherent planar polarity ( $n = 16$ ), suggesting the existence of alternative sources of polarizing cues (Figure 4Z). Collectively, these findings reveal that as few as four supporting cells can initiate and sustain integral organ regeneration.

### Sustentacular and mantle cells have different regenerative potential

Injury in the wild is intrinsically stochastic. Thus, we hypothesized that the regenerative response must vary according to damage severity and location, but progress in a predictable manner. To test this assumption and unveil the underlying cellular mechanism, we systematically quantified the behavior of individual cells by high-resolution videomicroscopy. We conducted 15 independent three-dimensional time-lapse recordings of the regenerative process using a triple-transgenic line co-expressing *Cldnb:lynGFP*, *SqGw57A* and *Alpl:mCherry* (Figure 5A–B), ranging from 65 to 100 hr of continuous imaging (each time point 15 min apart). Starting immediately after the ablation of all except 4–10 cells, we tracked every intact original cell (called founder cell) and their progeny (cellular clones) (Figure 5A and Video 2). We followed a total 106 founder cells (76 sustentacular cells and 30 mantle cells). We tracked individual cells manually in space and time, recording divisions and identity until the end of the recording, resulting in 763 tracks and 104,863 spatiotemporal cell coordinates (Figure 5A–B). Each clone was represented as a tree to visualize the contribution of each founder cell to the resulting clones (Figure 5C). We found that the majority of the founder sustentacular cells underwent three divisions and that some divided up to five times (Figure 5D). 14 out of 30 founder mantle cells did not divide at all, and the rest divided once or, rarely, twice. Founder sustentacular cells required on average  $19 \pm 6$  hr (mean  $\pm$  s.d.,  $n = 76$ ) to divide, whereas the founder mantle cells that divided required on average  $27 \pm 5$  hr, (mean  $\pm$  s.d.,  $n = 30$ ) (Figure 5E). Clones from founder sustentacular and founder mantle cells were markedly different: founder sustentacular cells produced all three cell classes (sustentacular, mantle and hair cells), whereas founder mantle cells produced clones containing only mantle cells (Figure 5F). We categorized all cell divisions according to the fate of the two daughter cells at the time of the following division, or at the end of the time-lapse recording (Figure 5G). This analysis revealed that 97% of the sustentacular-cell divisions were symmetric: 78% produced two sustentacular cells (SS), 16% produced a pair of hair cells (HH), and



**Figure 4.** Recovery of organ architecture after loss of tissue integrity. (A–E) Confocal images of a neuromast regenerating from 4 to 10 cells during a period of 7 days. Neuromasts recover radial symmetry 3 dpi (D), and original organ proportions at 7 dpi (E). (F–J) Neuromasts reduced to 4–10 cells that were previously deprived from adjacent interneuromast cells (INCs) (arrowheads in F), regenerated and reformed radial symmetry (H–I) and proportions 7 dpi, despite maintaining a reduced size (J). Dashed circles in (A,F) illustrate damaged areas. Scale bars: 10  $\mu$ m. (K,N) Total cell numbers in regenerating neuromasts over 7 days in the two conditions depicted in (A–J). (L,O) In the first 2 dpi neuromast consist almost exclusively of

Figure 4 continued

supporting cells (green and red). Hair cells (blue) begin to appear between at 2dpi. (M,P) Percentages of cell classes during a 7 day regeneration period. Right after damage, neuromast experience an imbalance of cell proportions that is re-established over the course of 3 days. Afterwards the neuromasts continues to slowly increase total cell number at similar rates. The final proportion of cell classes recapitulates that of the starting condition. Time points show mean  $\pm$  s.e.m. [All except 4–10 cells] n = 15, [All except 4–10 cells + INC] n = 6. (Q) Top and (R) lateral views of a triple-transgenic *Tg* [*Ncad: Ncad-EGFP*; *Alp:mCherry*; *H2A:H2A-EGFP*] neuromast before injury. (S) Top and (T) lateral views of a regenerated neuromast 7 days post injury (n = 4). Basal location of nuclei and apical N-cadherin enrichment evidence the apicobasal polarization of the organ. The accumulation of N-cadherin (white arrowheads) in the regenerated neuromast shows that apical constrictions are properly re-established during the process. (U–V) Maximal intensity projection of a neuromast in the triple transgenic line *Tg* [*Clidnb:lynGFP*; *SqGw57A*; *Alp:mCherry*] prior to injury that eliminates all except 4 to 10 cells (U), and the same neuromast 7 days after damage (V). (W) Hair-bundle staining with rhodamine-phalloidin (colored in pink) reveals the coherent planar polarization of the hair cells in the regenerated neuromast shown in (V). (X) Confocal projection of a neuromast before the removal of flanking interneuromast cells. (Y) Maximal projection of a neuromast 48 hr after interneuromast-cell ablation and 24 hr after neomycin treatment. (Z) Phalloidin staining of hair bundles of hair cells regenerated in the absence of interneuromast cells, showing recovery of coherent epithelial planar polarity. Scale bars: 10  $\mu$ m.

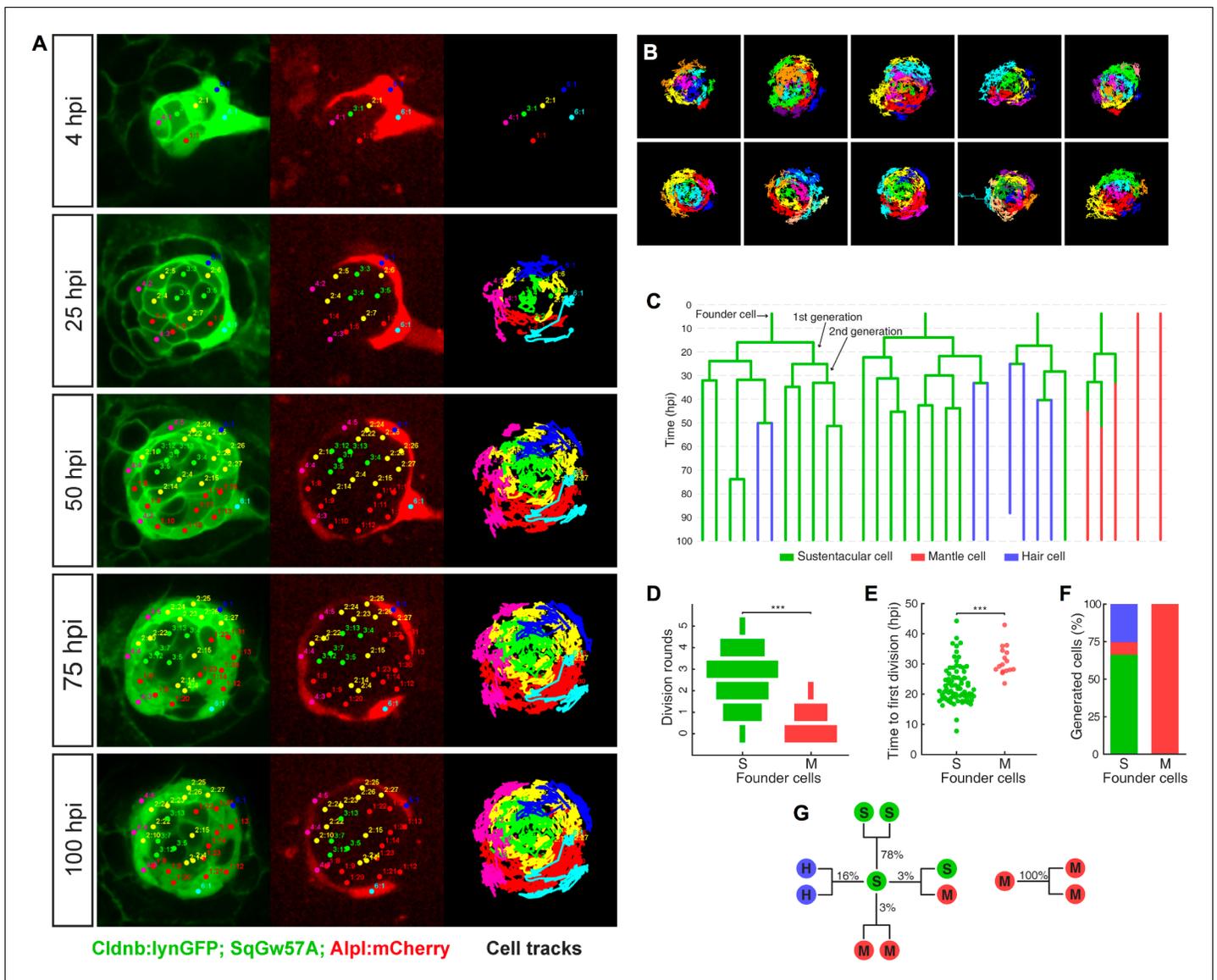
DOI: <https://doi.org/10.7554/eLife.30823.007>

3% generated two mantle cells (MM). Only 3% of the divisions were asymmetric, generating one sustentacular and one mantle cell (SM) (n = 307). All mantle-cell divisions were symmetric (MM) (n = 20). These observations further support the conclusion that sustentacular cells are tri-potent progenitors.

Previous studies have firmly established that hair-cell regeneration is strongly anisotropic because hair-cell progenitors develop almost exclusively in the polar areas of horizontal neuromasts, elongating the macula in the dorsoventral direction (Wibowo et al., 2011; Romero-Carvajal et al., 2015). Although our static images suggest that neuromasts have isotropic regenerative capacity, we nevertheless wondered whether regeneration of non-sensory cells is directional. To this end, we fractionated the epithelium of horizontal neuromasts in four quarters of equal dimension (dorsal, ventral, anterior and posterior) (Figure 6A–B), which reflects the known functional territorialization of the neuromast epithelium based on the expression of transgenic markers and Notch signaling (Ma et al., 2008; Wibowo et al., 2011). We first assessed the spatial distribution of cell divisions during the first 60 hr of regeneration and found no pattern that would suggest regeneration anisotropy (Figure 6A). However, 60 hpi, most divisions (74%) took place in the dorsal and ventral (polar) quarters (Figure 6B). This is expected because later divisions mainly produce hair cells from polar progenitors (Figure 4L, M). Thus, the regenerating epithelium is initially homogeneous and becomes territorialized 60 hpi. We reasoned that epithelial territorialization could occur either by the migration of similar cells that are scattered throughout the tissue, or by position-adaptive differentiation of an initially equivalent population of cells. To test these possibilities, we generated a virtual Cartesian coordinate system at the center of the neuromast to fit all founder cells at the beginning of regeneration (4hpi). Next, we analyzed the localization of their progeny 60 hpi (Figure 6C–H). We found that 60% of the progeny of anterior-localized founder cells were located in the anterior side of the resulting epithelium, whereas 64% of the progeny of posterior-located founder cells were found in the posterior side (Figure 6C–E). We also found that 72% of cells derived from dorsal founder cells and 74% of cells from ventral founder cells were located on the same side of the virtual dorsal/ventral midline (Figure 6F–H). Therefore, most of the clones remain ipsilateral to the founder cell. These results indicate that neuromasts have isotropic regenerative capacity and their territorialization occurs by location-adaptive cellular differentiation.

### The sustentacular-cell population is tri-potent and plastic

To answer the long-standing question of whether the sustentacular-cell population is homogeneous and approach the problem of what determines symmetric versus asymmetric modes of division, we characterized the composition of all 72 clones from founder sustentacular cells. We found four types of clones: containing only sustentacular cells (S), sustentacular and mantle cells (SM), sustentacular and hair cells (SH), and all three cell classes (SHM) (Figure 6I). Of note, founder mantle cells produced clones containing only mantle cells (M) (Figures 5G and 6I). We observed that 37/72 of the clones from founder sustentacular cells were SH, 21/72 were S, 12/72 were SM, and 2/72 were SHM (Figure 6I–K). The proportion of each clone type suggests that either the sustentacular-cell



**Figure 5.** Long-term whole-organ single-cell tracking reveals cell-clone formation during neuromast regeneration. (A) Still images showing a representative 100 hr time-lapse recording of a regenerating neuromast in *Tg[Cldnb:lynGFP; SqGw57A; Alpl:mCherry]* larva (left and middle panels). Cellular clones that share a common founder cell are clustered and color-coded. Cell trajectories reveal a concentric growth pattern (right panel). (B) Cell trackings at the last recorded timepoints for 10 out of the total of 15 regenerated neuromasts. (C) Cell-lineage tracing from time-lapse movie shown in (A). Branching points symbolize cell divisions. The division of a founder cell generates two cells of the 1st generation. Subsequent divisions produce cells of the 2nd, 3rd and 4th generation. Cell classes are indicated with green (sustentacular), blue (hair) and red (mantle) colors. (D) Sustentacular founder cells undergo significantly more ( $p=3.59e-06$ , Mann-Whitney test) division rounds than mantle founder cells during 100 hr of neuromast regeneration. (E) The first division of sustentacular founder cells ( $n = 76$ ) occurs significantly earlier ( $p=1.13e-5$ , Mann-Whitney test) than that of mantle founder cells ( $n = 16$ ). (F) Sustentacular founder cells ( $n = 76$ ) generate all three neuromast cell classes whereas mantle founder cells ( $n = 30$ ) produce only mantle cells. (G) Out of 307 sustentacular cell divisions, 78% were self-renewing, 16% produced a pair of hair cells, 3% produced sustentacular cells that both became mantle cells within the next generation and 3% generated two sustentacular cells of which only one transitioned to mantle cell fate within the next generation. All 20 observed mantle cell divisions were self-renewing.

DOI: <https://doi.org/10.7554/eLife.30823.008>

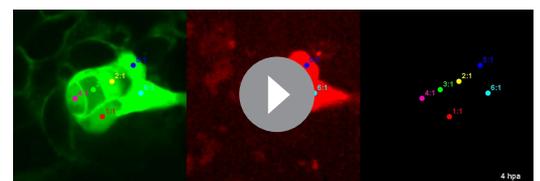
population is heterogeneous, or that it is homogeneous but plastic. In searching for potential sources of clone heterogeneity, we noted that in some developmental contexts cell-cycle length or proliferative potential can determine the fate of the daughter cells (Calegari et al., 2005; Rossi et al., 2017). Therefore, we quantified the kinetics of proliferation of founder sustentacular cells and of

their daughters and compared them to clone composition. We found three clear waves of cell divisions, each spaced by 8–10 hr (**Figure 7A**), respectively peaking at 20 hr, 28 hr and 38 hr (**Figure 7B–C**), suggesting that cell-cycle length is strictly regulated. Cell-cycle length in the 1st generation peaks around 10 hr ( $9.8 \pm 3.3$  hr, median  $\pm$ interquartile range (iqr)) (**Figure 7C**), but it begins to increase and to vary in the 2nd generation ( $11.5 \pm 7.3$  hr, median  $\pm$ iqr), and more so in the 3rd generation ( $18.8 \pm 20.3$  hr, median  $\pm$ iqr). To identify transition points in cycle lengths, we tested the goodness of fit of a two-segment regression model with variable change points. We found that the length of cell cycles is initially around  $11 \pm 3$  hr (mean  $\pm$ s.d.) and slowly increases up to 47 hpi. Afterwards, cell-cycle length increases more rapidly and is more variable (**Figure 7D**). To test if cell number influences cell-cycle length we used a similar two-segment regression model to define when cell-cycle length loosens, and discovered that the vast majority of the cell cycles (76%) span 7–13 hr below a threshold of 24 cells (**Figure 7E**). Above this threshold, cell-cycles lengths show large variation. With these data, we plotted proliferation kinetics against clone type, and found no significant difference between clones (**Figure 7F–G**). Thus, the length of the cell cycle or the proliferative potential of founder sustentacular cells cannot explain clone composition.

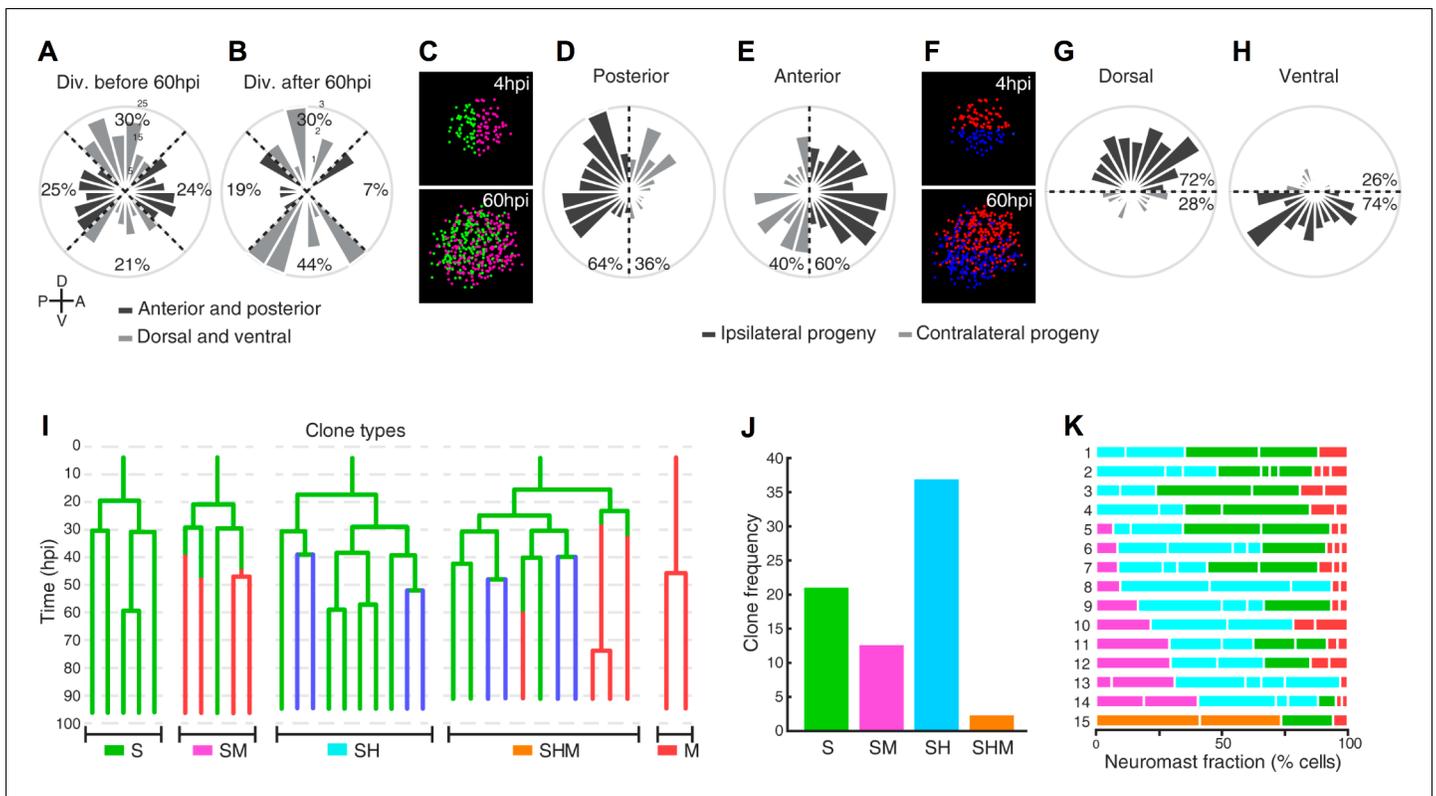
### Machine learning identifies predictive features for cell-fate acquisition

Multiple extrinsic factors that vary in space and time could determine cell-fate choices. Because manual analysis of such multidimensional data might be biased or neglect certain factors we implemented a quantitative and unbiased computational approach based on machine learning to identify variables (features) that correlate with clone composition. The first step of the workflow is the extraction of spatiotemporal coordinates and cell-lineage information from the manual tracks of the video-microscopic data sets ( $n = 15$ ) (**Figure 8A**). For each cell-track coordinate, we extracted 32 quantifiable features (**Table 1**), which were used to train the machine-learning algorithm. In a pre-analysis, we compared the performance of 20 algorithms (support vector machines, decision trees and nearest neighbor classifiers) in terms of accuracy and area under the curve (AUC) and chose the ensemble bagged tree random forest algorithm (**Breiman, 2001**) as the best performing method (**Figure 8—figure supplement 1**). To avoid overfitting, we trained the random forest using 14 samples to predict clone composition in the remaining sample in a round robin fashion. We evaluated the quality of predictions using Matthews correlation coefficient (MCC) to compensate for imbalances of clone frequencies (**Figure 6K**)

Using machine learning, we were able to predict the occurrence of SH vs. SM clones from a founder sustentacular cell with high accuracy (42 out of 49 correctly predicted clones,  $MCC = 0.63 \pm 0.09$ , mean  $\pm$  s.d.,  $n = 15$  bootstrapped samples), while neither SH nor SM clones could be discriminated when compared to S clones (**Figure 8B**). Of the 32 features that we used, those that best discriminated SH vs SM clones were the sustentacular cells' distance to the center of the epithelium, and the distance to the mantle cells (**Figure 8C** and **Figure 8—figure supplement 2**). Next, we focused on the decision-making process of individual sustentacular cells at the time of their division. We trained a random forest to discriminate between SS, HH and SM/MM divisions in a pairwise fashion. The HH and SM/MM divisions were highly predictable (63 out of 66 divisions correctly predicted,  $MCC = 0.91 \pm 0.07$ , mean  $\pm$  s.d.,  $n = 15$  bootstrapped samples), while the discrimination between SS and HH or SM/MM divisions was much less accurate (**Figure 8D**). Again, the most informative features were the distance to the neuromast center and the distance to the mantle cells (**Figure 8E**, **Figure 8—figure supplement 3**). SM/MM divisions occur consistently at the outer perimeter of the neuromast (**Figure 8F**), whereas HH divisions take place near the center. Self-renewing SS divisions occupy the area between HH and SM/MM divisions. Interestingly, SM/MM divisions were never seen in the anterior-most region of the organ, suggesting that progenitor sustentacular



**Video 2.** 100 hr time-lapse recording of a regenerating neuromast after severe ablation. A neuromast regenerates its original architecture from as few as six founder cells. Founder cells are identified by 1–6 ( $n$ ) and their daughter cells receive  $2n$  and  $2n + 1$  identities. Recording starts 4 hr post injury (hpi) and shows single focal planes. Time is in hours post injury. DOI: <https://doi.org/10.7554/eLife.30823.009>



**Figure 6.** Neuromast regeneration is not stereotypic and reveals different clone type compositions. (A) Proliferation is markedly isotropic during the first 60 hr of neuromast regeneration ( $n = 348$ ). (B) Homeostatic, dorso-ventral (DV) proliferative territories are restored after 60hpi ( $n = 27$ ). (C) 40% and 36% of the progeny from anterior ( $n = 202$ ) and posterior ( $n = 173$ ) founder cells crossed to the contralateral side (light grey) after 60 hr of regeneration. (D) Only 28% and 26% of the progeny from dorsal ( $n = 199$ ) and ventral ( $n = 176$ ) founder cells crossed to the contralateral side (light grey) during the same period of time. (E) Representative examples of different clone types extracted from time-lapse data. Sustentacular cells give rise to S, SM, SH, and SHM clones (color coded respectively with green, pink, cyan and orange) whereas mantle cells produce only pure mantle cell clones. (F) The clone composition of the 15 regenerated neuromasts is not stereotypic. The length of each bar represents the proportion of neuromast cells that belong to each clone. Neuromast eight has been shown in **Figure 5A,B**. (G) The most frequent clones contain sustentacular and hair cells (SH,  $n = 37$  clones), followed by those with only sustentacular cells (S,  $n = 21$  clones). The third most frequent are composed by sustentacular and mantle cells (SM,  $n = 12$  clones). Clones containing all three cell classes were rare (SHM,  $n = 2$  clones).

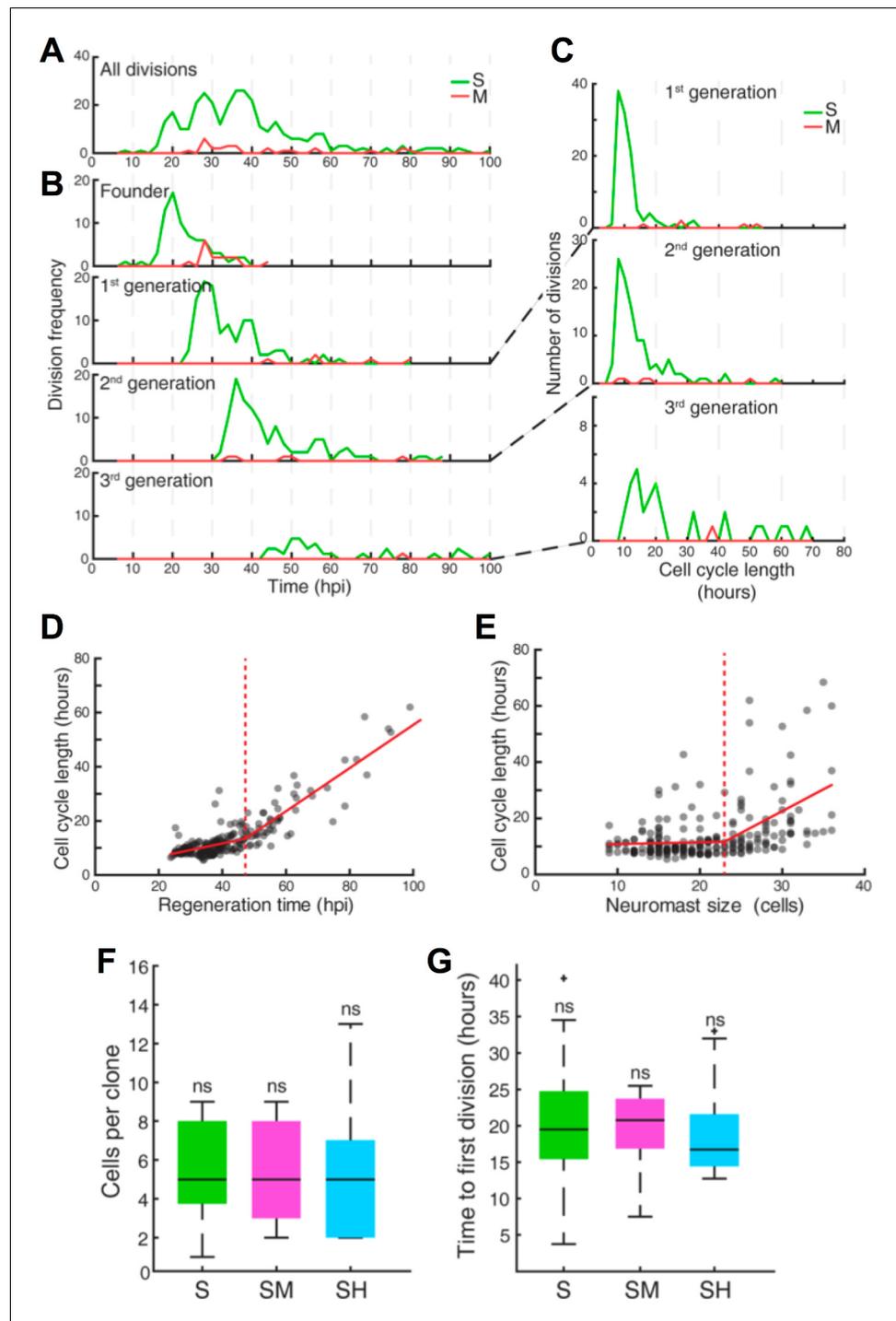
DOI: <https://doi.org/10.7554/eLife.30823.010>

cells are routed into generating mantle cells specifically in the perimetral areas that lack mantle cells but not elsewhere. Therefore, regenerating neuromasts appear to sense cell-class composition and route cellular differentiation in a spatially regulated manner to regain cell-class proportion and distribution.

## Discussion

One long-standing goal of biological research is to understand the regeneration of tissues that are exposed to persistent environmental abrasion. Here we address this problem by developing a quantitative approach based on videomicroscopic cell tracking, cell-lineage tracing, and machine learning to identify features that predict cell-fate choices during organ regeneration. Using the superficial neuromasts in zebrafish, we demonstrate that a remarkably small group of resident cells suffices to rebuild a functional organ following severe disruption of tissue integrity. Our findings reveal that the sustentacular-cell population is tri-potent, and suggest that integral organ recovery emerges from multicellular organization employing minimal extrinsic information. Below, we discuss the evidence that supports these conclusions.

By systematically analyzing cellular behavior, we reveal a hierarchical regenerative process that begins immediately after injury. First, surviving founder cells reconstitute an epithelium. Second,



**Figure 7.** Quantification of cell divisions during neuromast regeneration. (A–B) Equally spaced waves of coordinated sustentacular cell divisions (green) underlie the recovery of neuromast cell size. Mantle cell divisions (red) occur occasionally and do not follow the pattern of sustentacular cells. Proliferative waves correspond to the coordinated divisions of cells from independent generations. (C) Cells from the 1<sup>st</sup> and 2<sup>nd</sup> generation divide on average after cell cycles of  $11 \pm 5$  and  $14 \pm 9$  hr respectively (mean  $\pm$  s.d.). Coordination is lost at the 3<sup>rd</sup> generation when cell cycles start to lengthen ( $26 \pm 18$  hr, mean  $\pm$  s.d.). (D) Cell cycle length ( $11 \pm 3$  hr, mean  $\pm$  s.d.) is marginally influenced by regeneration time until 47 hr after injury, when cycle length starts increasing proportionally with regeneration time. (E) Cell cycle lengths ( $12 \pm 6$  hr, mean  $\pm$  s.d.) do not correlate directly with neuromast size until 24 neuromast cells. (F) S, SM and SH clones produce similar number of cells ( $p=0.68$ , Kruskal Wallis test). In the box plots, the boundary of the box indicates the 25<sup>th</sup> and 75<sup>th</sup> percentile, respectively the black horizontal line indicates the median. (G) Time to first division is similar for S, SM and SH clones. *Figure 7 continued on next page*

Figure 7 continued

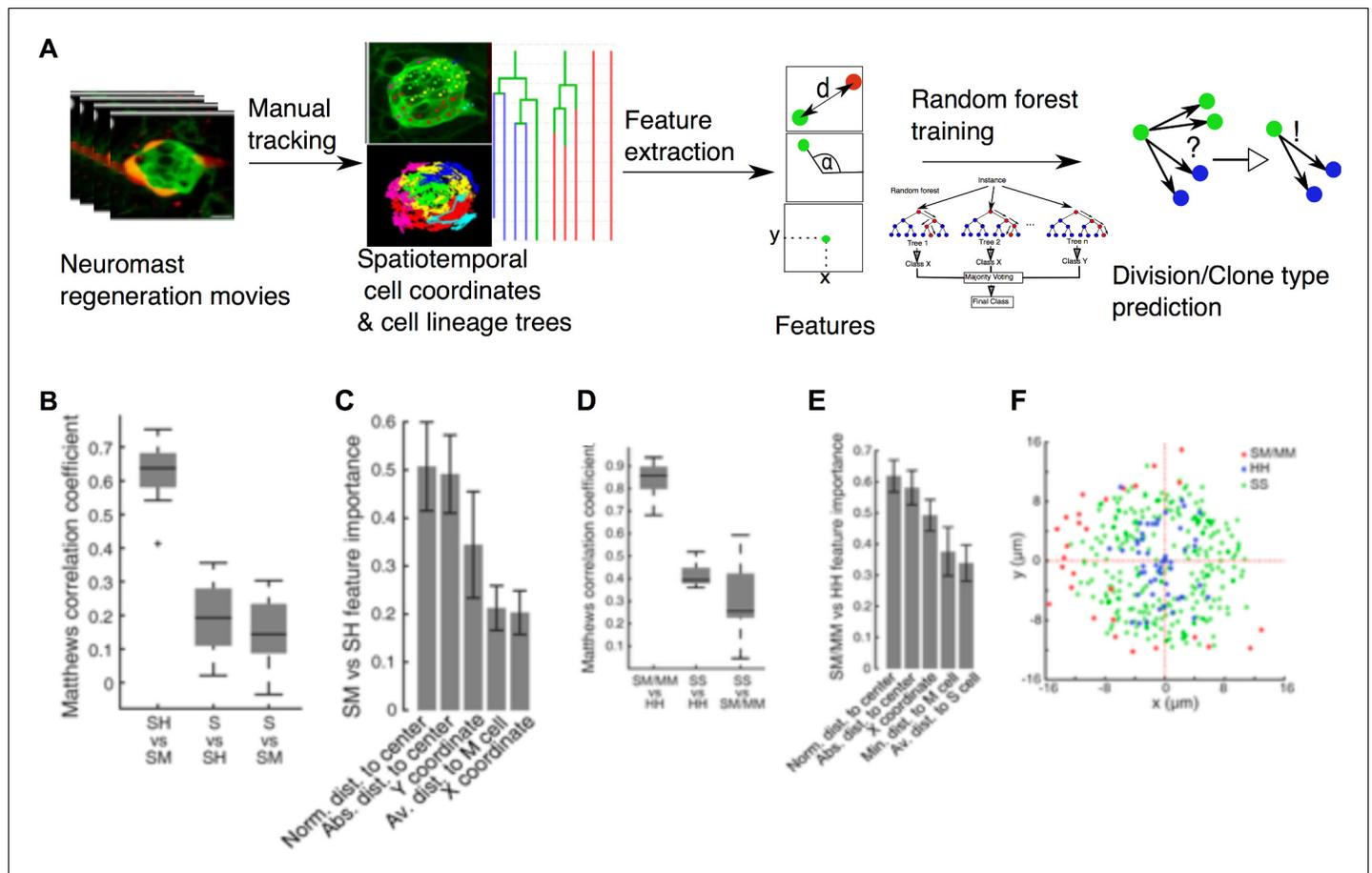
line within the box marks the median. Whiskers above and below the box include points that are not outliers. (G) Sustentacular founder cells of S, SM, and SH clones divide similarly early ( $p=0.42$ , Kruskal Wallis test) after approximately 18 hr after neuromast injury. (H) Sustentacular founder cells that produce SH (cyan) and S clones (green) are distributed similarly around the center of the organ (at  $x = y = 0$ ). Those that generate SM clones (pink) are localized further away from the center and are biased towards the posterior side.

DOI: <https://doi.org/10.7554/eLife.30823.011>

sustentacular cells become proliferative and restore organ size. Cellular intercalation is rare. Third, daughter cells differentiate in a position-appropriate manner to recreate cell-class proportions and organ geometric order. Fourth, the epithelium returns to a homeostatic state that is characterized by low mitotic rate. The milder damage regimes that eliminated one half of the epithelium show that neuromasts are symmetric in their regenerative capacity, and that they preferentially regenerate the cells that have been eliminated. Importantly, these findings, which rely on the quantitative spatio-temporal analysis of regeneration data, could not have been predicted from previous studies using static and largely qualitative information (Williams and Holder, 2000; López-Schier and Hudspeth, 2005; Dufourcq et al., 2006; López-Schier and Hudspeth, 2006; Ma et al., 2008; Wibowo et al., 2011; Wada et al., 2013; Steiner et al., 2014; Romero-Carvajal et al., 2015; Cruz et al., 2015; Pinto-Teixeira et al., 2015). An important corollary of these results is that neuromasts do not contain specialized cells that contribute dominantly to repair. We propose that progenitor behavior is a facultative status that every sustentacular cell can acquire or abandon during regeneration. We did not observe regenerative overshoot of any cell class (Agarwala et al., 2015), suggesting the existence of a mechanism that senses the total number of cells and the cell-class balance during tissue repair (Simon et al., 2009). Together with previous work, our results support the possibility that such mechanism is based on the interplay between Fgf, Notch and Wnt signaling (Ma et al., 2008; Wibowo et al., 2011; Wada et al., 2013; Romero-Carvajal et al., 2015; Dalle Nogare and Chitnis, 2017). Our combination of machine learning and quantitative videomicroscopy shows clear differences between sustentacular and mantle cells, but does not indicate heterogeneity within the sustentacular-cell population. However, further application of this integrated approach and new transgenic markers may reveal uncharacterized cells in the neuromast. This may be expected given recent work that showed the existence of a new cell class in neuromasts of medaka fish (Seleit et al., 2017). It is technically challenging to consistently maintain fewer than 4 cells *in toto* without eliminating the entire neuromast. Thus, we cannot rule out the possibility that a single founder cell may be able to regenerate a neuromast. We show that the complete elimination of a neuromast is irreversible in larval zebrafish. However, Sánchez and colleagues have previously reported that interneuromast cells can generate new neuromasts (Sánchez et al., 2016). By assaying DNA synthesis prior to mitosis, we show that interneuromast cells do not proliferate after neuromast ablation. These differences may be explained by differences in ablation protocols (electroablation versus laser-mediated cell killing), the age of the specimens (embryos versus early larva) or the markers used to assess cellular elimination.

We find that interneuromast cells are not essential for neuromast regeneration because severely damaged organs recover all cell classes in the appropriate localization in the absence of interneuromast cells. However, we systematically observed smaller organs when interneuromast cells were ablated. These observations suggest that these peripheral cells may yet help regeneration, either directly by contributing progeny, or by producing mitogenic signals to neuromast-resident cells.

The behavior of the mantle cells is especially intriguing. Complete elimination of parts of the lateral line by tail-fin amputation have revealed that mantle cells are able to proliferate and generate a new primordium that migrates into the regenerated fin to produce new neuromasts (Dufourcq et al., 2006). This observation can be interpreted as suggesting that under some injury conditions, mantle cells are capable of producing all the cell classes of a neuromast. Transcriptomic profiling of mantle cells following neuromast injury revealed that these cells up-regulate the expression of multiple genes (Steiner et al., 2014). Furthermore, a recent study has revealed that mantle cells constitute a quiescent pool of cells that re-enters cell cycle only in response to severe depletion of sustentacular cells (Romero-Carvajal et al., 2015), suggesting that these cells may conform a stem-cell niche for proliferation of sustentacular cells. Thus, the collective evidence indicates that the



**Figure 8.** Implementation of predictive machine-learning analysis. (A) Overview from experiments to prediction. Movies of neuromast regeneration allow us to track every single cell over 100hpi and to generate a cell lineage from these track points. Information covered in all tracks and lineages can be extracted as features with which we train our random forest machine-learning classifier to predict division or cell lineage fate. (B) Sustentacular founder cell choices between SH vs. SM clones can be predicted with high accuracy (MCC =  $0.63 \pm 0.09$ , mean  $\pm$  s.d.,  $n = 15$  bootstrapped samples) whilst choices between S and SH or SM clones are highly inaccurate (MCC =  $0.19 \pm 0.11$  and  $0.15 \pm 0.10$ , mean  $\pm$  s.d., respectively,  $n = 15$  bootstrapped samples), based on 32 calculated features. (C) Features relative to the position of the founder cells and their nearest cellular environment can discriminate between SM and SH clone types. (D) Choices between SM/MM and HH divisions can be predicted with high accuracy (MCC =  $0.91 \pm 0.07$ , mean  $\pm$  s.d.,  $n = 15$  bootstrapped samples) while those between SS and HH or SM/MM have low accuracy (MCC =  $0.50 \pm 0.05$  and  $0.38 \pm 0.15$ , respectively, mean  $\pm$  s.d.,  $n = 15$  bootstrapped samples) (E) Features describing the cell's position in relation to the neuromast center and their proximity to other mantle cells have the highest influence on the cell fate choices of a sustentacular cell. (F) SM/MM divisions (red) appear predominantly at the periphery of the organ whereas HH divisions (blue) appear proximal to the center. Sustentacular cell self-renewing divisions (SS, green) occur mostly around the neuromast center, generating a ring-like pattern.

DOI: <https://doi.org/10.7554/eLife.30823.013>

The following figure supplements are available for figure 8:

**Figure supplement 1.** Comparison of different classification methods.

DOI: <https://doi.org/10.7554/eLife.30823.014>

**Figure supplement 2.** Features used to predict SM vs SH clones sorted by predictive importance.

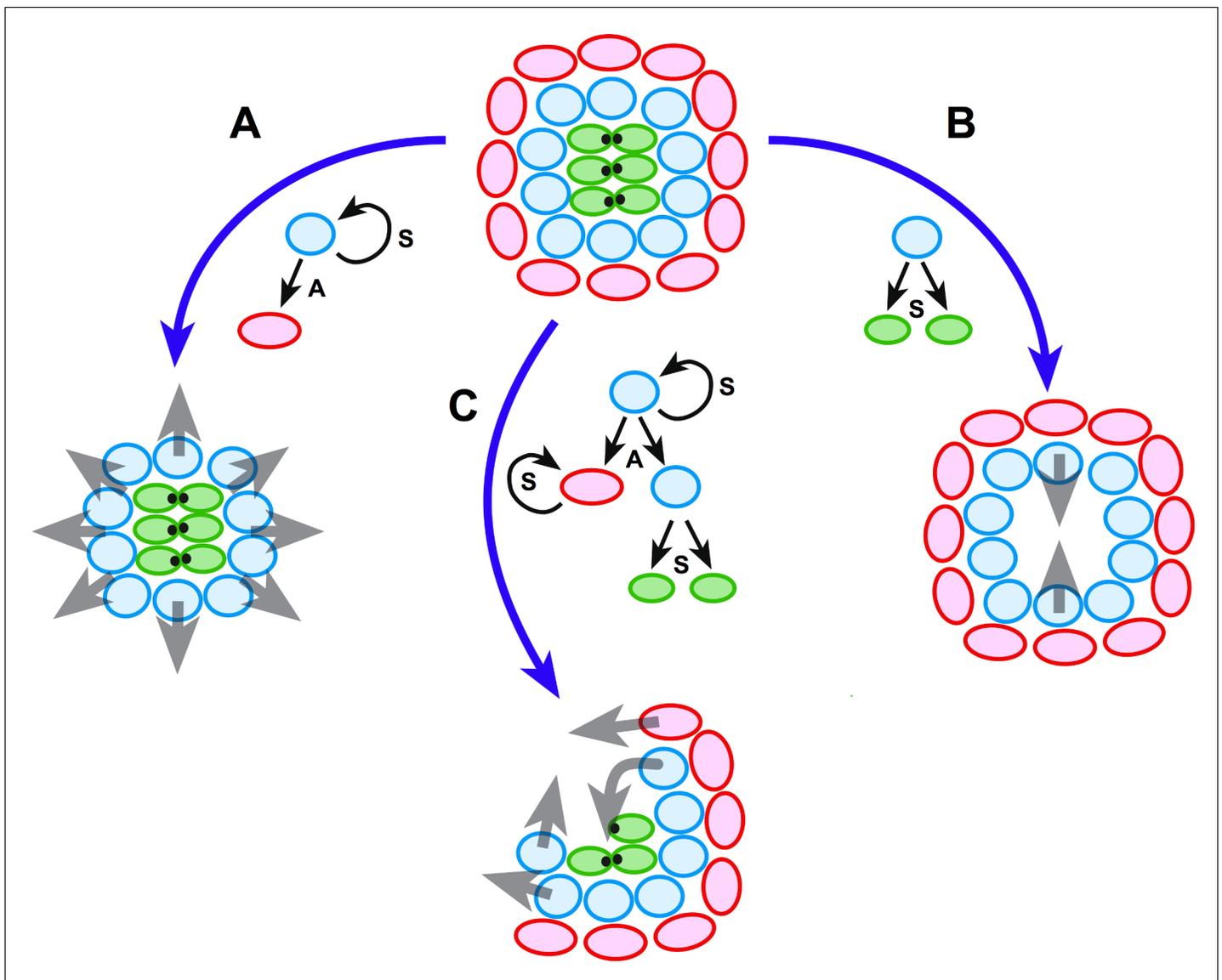
DOI: <https://doi.org/10.7554/eLife.30823.015>

**Figure supplement 3.** All features used to predict SM/MM vs HH divisions sorted by predictive importance.

DOI: <https://doi.org/10.7554/eLife.30823.016>

mantle cells respond to damage and contribute to the regenerative processes, and may drive the regeneration of an entire organ if every other cell class is lost.

One outstanding question is how regeneration is controlled spatially. The epithelium may respond to damage via dynamic formation of an injured-intact axis at the onset of repair. Our results support this scenario by unveiling the adaptability of the neuromast epithelium to the localization



**Figure 9.** Schematic model of neuromast regeneration. The top diagram exemplifies the architecture of an intact neuromast. A, B and C indicate three types of injury: A when mantle cells are lost, B when hair cells are ablated, and C when a localized combination of all three cell classes is lost. Under the model that we present, radial symmetry serves to localize damage and canalize regeneration spatially. If central hair cells are lost (**A**), radial symmetry is maintained for sustentacular progenitors to regenerate hair cells centripetally (grey arrows in **A**). If outer cells are lost (**B**), radial symmetry is also maintained for the generation of progeny that will acquire mantle fate and propagate centrifugally to reform the outer rim of the neuromast (grey arrows in **B**). Upon asymmetric damage, however, the radial symmetry is partially broken (**C**). The neuroepithelium repolarizes along an injured-intact axis, which canalizes regeneration towards the damaged areas (grey arrows in **C**). Individual cells are color-coded (mantle cells in red, sustentacular cells in light blue, and hair cells in green), and in each case we indicate the type of division that the intact cells undergo: symmetric (**S**) when they produce two equivalent cells or self-renew, and asymmetric (**A**) when their division generates sibling cells that differentiate into different classes.

DOI: <https://doi.org/10.7554/eLife.30823.017>

and scale of damage. We suggest a model in which the invariant radial symmetry of the neuromast serves as a rheostat to identify the site of damage to guide regeneration spatially (**Figure 9**). A polarized axis along structurally intact and injured areas underlies this process. However, the complete reconstruction of a neuromast by as few as 4 cells suggests that a partial maintenance of radial symmetry is not essential for organ regeneration. Therefore, radial-symmetry maintenance cannot have a deterministic impact on the recovery of geometric order. Yet, partial structural maintenance and polarized tissue responses may optimize repair, respectively, by preventing superfluous cellular

production in undamaged areas and by biasing the production of lost cells in the damaged areas. For organs that have evolved under the pressure of persistent damage, compliance to the extent of the injury may be an advantage because the regenerative responses can be scalable and localized, allowing faster and more economical regeneration.

After the severest of injuries, regenerated neuromasts were plane polarized in a manner indistinguishable from unperturbed organs. This startling result indicates that as few as four founder supporting cells can re-organize the local coherent planar polarity of the epithelium during neuromast repair. An alternative explanation is that founder cells have access to external polarizing cues. One source of this information is an isotropic mechanical forces exerted by the interneuromast cells that flank a neuromast. This is possible because interneuromast cells are always aligned to the neuromast's axis of planar polarity. Yet, the concurrent ablation of resident hair cells and the interneuromast cells around an identified neuromast led to regenerated hair cells whose local orientation was coherent. Interestingly, recent studies have identified a transcription factor called *Emx2* that regulates the orientation of hair cells in neuromasts of the zebrafish (*Jiang et al., 2017*). *Emx2* is expressed in one half of the hair cells of the neuromast (those oriented towards the tail) and absent in the other half (which are coherently oriented towards the head). Loss- and gain-of-function of *Emx2* alter planar cell polarity in a predictable manner: loss of *Emx2* leads to neuromasts with every hair cells pointing towards the head of the animal, and *Emx2* broad expression orients hair cells towards its tail. Because the coherent local axis of polarity is not affected by these genetic perturbations, *Emx2* may act in hair cells as a decoder of global polarity cues. This evidence, together with our results, suggests that during neuromast regeneration founder cells autonomously organize the variegated expression of *Emx2* in the regrowing epithelium with consequent recovery of a coherent axis of planar polarity and with one half of the hair cells pointing opposite to the other half. The future development of live markers of *Emx2* expression will be able to test this prediction. We would like to highlight that we do not currently understand the global polarization of the neuromast epithelium relative to the main body axes of the animal. External sources of polarity may impinge in the recovery of these global axes during neuromast regeneration. Previous work has demonstrated that local and global polarization occur independently of innervation (*López-Schier and Hudspeth, 2006*), but other potential polarizing cues remain untested. Therefore, at present we can only support the notion that local coherent polarity is self-organizing, whereas global orientation may be controlled externally.

Our results beg the question of whether neuromast cells self-organize. Our operational definition of self-organization is an 'autonomous increase in order by the sole interaction of the elements of the system' (*Haken, 1983*), implying that a cellular collective organizes a complex structure without the influence of external morphogenetic landmarks, patterning cues, or pre-existent differential gene-expression profiles. If these conditions are not met, cellular groups may nevertheless form a complex structure through a process of 'self-assembly' (*Sasai, 2013; Turner et al., 2016*). The reduction of neuromasts to around 5% of their original size shows that intact resident cells can rapidly recreate their original microenvironment to rebuild a neuromast with normal organization, proportions and polarity. Although these observations suggest autonomy, extrinsic sources of information including the extracellular matrix that remains intact after cell loss may serve as a blueprint for epithelial organization. Yet, unless such patterns are rebuilt together with the organ, neuromasts architecture and proportions would depend on the area occupied by the regrowing epithelium. In other words, cell-fate acquisition and cell-class distribution must be tissue-size dependent. However, we show that neuromast regain geometric order as early as 2 days after injury, when their cellular content is less than 60% of the original. Although our results do not irrefutably demonstrate self-organization during neuromast regeneration, they strongly support this idea. We argue that self-organization is an optimal morphogenetic process to govern organ repair because (i) it requires the least amount of previous information and (ii) it is robust to run-off signals that could lead to catastrophic failure.

## Conclusions

Understanding how tissues respond to the inherently random nature of injury to recapitulate their architecture requires the identification of cues and signals that determine cell-fate acquisition, localization and three-dimensional organization. Here we reveal an archetypal sensory organ endowed with isotropic regenerative ability and responses that comply to damage severity, nature and localization. An important corollary of our findings is that progenitor behavior is a facultative status that every

**Table 1.** List of prediction features with description.

We used 32 mainly spatial and neighborhood specific features for the classification. Features are explained in the description column.

Feature name	Description
Absolut time	Hours post induction (hpi)
Absolute distance to center	Euclidean distance to the neuromast center
Average distance to H cell	-
Average distance to M cell	-
Average distance to S cell	-
Cell generation	Number of divisions that the cell has undergone
Founder Cell Type	-
Minimum distance to H cell	-
Minimum distance to M cell	-
Minimum distance to S cell	-
Movement angle to last division	Angle between current cell location, neuromast center and location of last cell division (or start of the movie in case of founder cell division)
Movement direction compared to center	Radial distance between current cell location and location of last cell division (or start of the movie in case of founder cell division). If the current location is nearer to the center the value is (+) in case it is further away the value is (-)
Movement distance since last division	Euclidean distance between current cell location and last cell division (or start of the movie in case of founder cell division)
Normalized distance to center	Radial distance of current cell location divided by the radial distance of the current furthest cell (to approximate the neuromast size)
Number of founder cells	-
Number of H cells	-
Number of H cells in 10 $\mu\text{m}$ radius	-
Number of H cells in 20 $\mu\text{m}$ radius	-
Number of H cells in 30 $\mu\text{m}$ radius	-
Number of M cells	-
Number of M cells in 10 $\mu\text{m}$ radius	-
Number of M cells in 20 $\mu\text{m}$ radius	-
Number of M cells in 30 $\mu\text{m}$ radius	-
Number of S cells	-
Number of S cells in 10 $\mu\text{m}$ radius	-
Number of S cells in 20 $\mu\text{m}$ radius	-
Number of S cells in 30 $\mu\text{m}$ radius	-
Number of total cells	-
Polar angle	Polar angle is the counterclockwise angle between the x-axis, the neuromast center and the current cell location
Time to last division	Time to last division (or start of the movie in case of founder cell division)
X coordinate	-
Y coordinate	-

DOI: <https://doi.org/10.7554/eLife.30823.012>

sustentacular cell can acquire or abandon during regeneration (*Blanpain and Fuchs, 2014; Wymeersch et al., 2016*). Importantly, we illustrate a machine learning implementation to identify features that predict cell-fate choices during tissue growth and morphogenesis. This quantitative approach is simple and model-independent, which facilitates its application to other organs or experimental systems to understand how multiple cells interact dynamically during organogenesis and organ regeneration in the natural context of the whole animal, and to identify how divergences from the normal regenerative processes lead to failed tissue repair.

## Materials and methods

### Zebrafish strains and husbandry

Zebrafish were maintained under standard conditions, and experiments were performed in accordance with protocols approved by the PRBB Ethical Committee of Animal Experimentation of the PRBB Barcelona, Spain. Eggs were collected from natural spawning and maintained at 28.5°C in Petri dishes at a density of up to 50 per dish. Transgenic lines used were *ET(krt4:EGFP)SqGw57A* (referred to in the text as SqGw57A) (*Kondrychyn et al., 2011*), *ET(krt4:EGFP)SqET4* (SqET4) (*Parinov et al., 2004*), *Tg[Myo6b:actb1-EGFP]* (*Kindt et al., 2012*), *Tg[-8.0cldnb:Lyn-EGFP]* (Cldnb:lynGFP) (*Haas and Gilmour, 2006*), *Tg[Alpl:mCherry]* (*Steiner et al., 2014*), *Tg[Sox2-2a-sfGFPstl84]* (referred to as Sox2:GFP) (*Shin et al., 2014*). To label cell nuclei, in vitro transcribed capped RNA coding for histone 2B-mCherry was injected in 1–4 cell embryos at a concentration of 100 ng/μl (*Rosen et al., 2009*). Throughout the study, zebrafish larvae were anesthetized with a 610 μM solution of the anesthetic 3-aminobenzoic acid ethyl ester (MS-222).

### Laser-mediated cell ablations

For *in toto* cell ablation, we used the iLasPulse laser system (Roper Scientific SAS, Evry, France) mounted on a Zeiss Axio Observer inverted microscope equipped with a 63X water-immersion objective (N.A. = 1.2) (*Xiao et al., 2015*). The same ablation protocol was used for all experiments using five dpf larvae. Briefly, zebrafish larvae were anesthetized, mounted on a glass-bottom dish and embedded in 1% low-melting point agarose. Three laser pulses (355 nm, 400 ps/2.5 μJ per pulse) were applied to each target cell. After beam delivery, larvae were removed from the agarose and placed in anesthesia-free embryo medium. All ablations were systematically performed on the L2 or L3 posterior lateral-line neuromasts, except for those in **Figure 6F**, for which we targeted the LII.2 neuromast.

### Phalloidin staining

Samples were fixed in 4% PFA overnight at 4°C, washed several times in 0.1% PBSTw and incubated in phalloidin-Alexa 568 or Alexa 488 (Invitrogen) diluted 1:20 in 0.1% PBSTw overnight at 4°C. Samples were washed several times in 0.1% PBSTw and mounted in 0.1% PBSTw with Vectashield (1/100, Vector Labs, Burlingame, CA, USA).

### Regeneration analysis and quantification

For quantification of cell numbers during neuromast regeneration, *Tg[Cldnb:lynGFP; SqGw57A; Alpl:mCherry]* zebrafish larvae were anesthetized, mounted on a glass-bottom dish and embedded in 1% low-melting point agarose. All samples were imaged before injury, 4 hpi and every 24 hr up to 7 dpi with an inverted spinning-disc confocal microscope (Zeiss by Visitron), under a 63X water-immersion objective. After imaging, larvae were quickly transferred to anesthetic-free medium. Cells were manually counted using the FIJI multi-point tool by scrolling throughout the entire volume of the neuromast. Cell classes were identified by the following criteria: Interneuromast cells: Cldnb:lynGFP(+), SqGw57A(-), Alpl:mCherry(+). Mantle cells: Cldnb:lynGFP(+), SqGw57A(+), Alpl:mCherry(+). Sustentacular cells: Cldnb:lynGFP(+), SqGw57A(+), Alpl:mCherry(-). Hair cells: Cldnb:lynGFP(+), SqGw57A(-), Alpl:mCherry(-). Hair cell identity was verified by the concomitant observation of the correct transgene expression pattern, central-apical location and the presence of a hair-cell bundle. Data was processed and analyzed using GraphPad Prism version 6.04 for Windows (GraphPad Software, La Jolla, CA, USA, [www.graphpad.com](http://www.graphpad.com)). In the box plots, the boundary of the box closest to zero indicates the 25th percentile (q1), a black line within the box marks the median, and the boundary of the

box farthest from zero indicates the 75th percentile (q3). Whiskers above and below the box include points that are not outliers. Points are considered as outliers if they are bigger than  $q3 + 1.5(q3 - q1)$  or smaller than  $q1 - 1.5(q3 - q1)$ .

## Videomicroscopy, cell tracking and lineage tracing

Larvæ were anesthetized, mounted onto a glass-bottom 3 cm Petri dish (MatTek) and covered with 1% low-melting point agarose with diluted anesthetic. Z-stack series were acquired every 15 min at 28.5°C using a 63X water-immersion objective. Cells were tracked overtime using volumetric Z-stack images with FIJI plugin MTrackJ (Meijering et al., 2012). Movies were registered two times for image stabilization and centered upon the centroid of the surviving group of cells and the subsequent regenerating organs. Founder cells are identified from 1 to 6 (*n*) and their daughter cells receive  $2n$  and  $2n + 1$  identities. All images were processed with the FIJI software package.

## Pharmacology

All pharmacological treatments were performed as described previously (López-Schier and Hudspeth, 2006; Wibowo et al., 2011; Pinto-Teixeira et al., 2015). Briefly, the following concentrations and timings used were: Neomycin sulfate (Sigma, St. Louis, MO) 250  $\mu$ M for 45 min; N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine-t-butyl ester (DAPT) (Sigma) 100  $\mu$ M for 24–48 hr. Equal amounts of DMSO were diluted in embryo medium for control specimens.

## Random forest prediction

Random forest algorithms use the majority vote of numerous decision trees based on selected features to predict choices between given outcomes (Murphy, 2012). We used a list of spatial, movement and neighborhood features (see Suppl. Table 1) to perform the random forest prediction of fate choice. We trained the random forest on 14 experiments and tested our prediction on one left-out experiment in a round robin fashion, leading to 15 test sets overall. To evaluate our prediction, we calculated Matthews correlation coefficient (MCC) (Matthews, 1975), which accounts for imbalance in our data (e.g. 78% of all divisions are SS divisions). The MCC is calculated by:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP denotes true positive, TN true negative, FP true positive and FN false negative predictions. The MCC can have values between  $-1$  and  $+1$ , where  $-1$  is a completely incorrect,  $0$  a random and  $+1$  a perfect prediction. To evaluate the variance of the MCC on the 15 test sets we used a bootstrapping approach, where we draw 15 samples from all test sets with replacement 15 times. From this resampled data we calculated the mean MCC and the standard deviation as shown in Figure 8B and D. All machine-learning analyses were performed using MATLAB (Version 2015b on a Windows 7 machine)

## Acknowledgements

We thank A Steiner, T Nicolson and L Solnica-Krezel for transgenic zebrafish, the animal facility personnel at the CRG of Barcelona and the HMGU for animal care, and Kirill Smirnov for statistical support. Funding was provided by the European Research Council Grant 2007\_205095, by the ESF Research Networking Programme 'QuanTissue', and by the AGAUR Grant 2009-SGR-305 of Spain.

## Additional information

### Funding

Funder	Grant reference number	Author
European Research Council	2007_205095	Hernán López-Schier
Agència de Gestió d'Ajuts Universitaris i de Recerca	2009-SGR-305	Hernán López-Schier

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

#### Author contributions

Oriol Viader-Llargués, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review and editing; Valerio Lupperger, Resources, Software, Methodology, Writing—review and editing; Laura Pola-Morell, Resources, Investigation, Writing—review and editing; Carsten Marr, Software, Formal analysis, Supervision, Methodology, Writing—review and editing; Hernán López-Schier, Conceptualization, Supervision, Funding acquisition, Visualization, Writing—original draft, Project administration, Writing—review and editing

#### Author ORCIDs

Hernán López-Schier  <https://orcid.org/0000-0001-7925-7439>

#### Ethics

Animal experimentation: Zebrafish were maintained under standard conditions. Experiments with wild-type, mutant and transgenic embryos of undetermined sex were conducted in accordance with institutional guidelines and under a protocol approved by the Ethical Committee of Animal Experimentation of the Parc de Recerca Biomedica de Barcelona, Spain, and protocol number Gz.:55.2-1-54-2532-202-2014 by the "Regierung von Oberbayern", Germany.

#### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.30823.020>

Author response <https://doi.org/10.7554/eLife.30823.021>

## Additional files

#### Supplementary files

- Transparent reporting form

DOI: <https://doi.org/10.7554/eLife.30823.018>

## References

- Agarwala S, Duquesne S, Liu K, Boehm A, Grimm L, Link S, König S, Eimer S, Ronneberger O, Lecaudey V. 2015. Amotl2a interacts with the Hippo effector Yap1 and the Wnt/ $\beta$ -catenin effector Lef1 to control tissue size in zebrafish. *eLife* **4**:e08201. DOI: <https://doi.org/10.7554/eLife.08201>, PMID: 26335201
- Behra M, Bradsher J, Sougrat R, Gallardo V, Allende ML, Burgess SM. 2009. Phoenix is required for mechanosensory hair cell regeneration in the zebrafish lateral line. *PLoS Genetics* **5**:e1000455. DOI: <https://doi.org/10.1371/journal.pgen.1000455>, PMID: 19381250
- Blanpain C, Fuchs E. 2014. Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. *Science* **344**:1242281. DOI: <https://doi.org/10.1126/science.1242281>, PMID: 24926024
- Boj SF, Hwang CI, Baker LA, Chio II, Engle DD, Corbo V, Jager M, Ponz-Sarvisé M, Tiriác H, Spector MS, Gracanin A, Oni T, Yu KH, van Boxel R, Huch M, Rivera KD, Wilson JP, Feigin ME, Öhlund D, Handly-Santana A, et al. 2015. Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**:324–338. DOI: <https://doi.org/10.1016/j.cell.2014.12.021>, PMID: 25557080
- Breiman L. 2001. Random Forests. *Machine Learning* **45**:5–32. DOI: <https://doi.org/10.1023/A:1010933404324>
- Calegari F, Haubensak W, Haffner C, Huttner WB. 2005. Selective lengthening of the cell cycle in the neurogenic subpopulation of neural progenitor cells during mouse brain development. *Journal of Neuroscience* **25**:6533–6538. DOI: <https://doi.org/10.1523/JNEUROSCI.0778-05.2005>, PMID: 16014714
- Cruz IA, Kappedal R, Mackenzie SM, Hailey DW, Hoffman TL, Schilling TF, Raible DW. 2015. Robust regeneration of adult zebrafish lateral line hair cells reflects continued precursor pool maintenance. *Developmental Biology* **402**:229–238. DOI: <https://doi.org/10.1016/j.ydbio.2015.03.019>, PMID: 25869855
- Dalle Nogare D, Chitnis AB. 2017. A framework for understanding morphogenesis and migration of the zebrafish posterior Lateral Line primordium. *Mechanisms of Development* **148**:69–78. DOI: <https://doi.org/10.1016/j.mod.2017.04.005>, PMID: 28460893

- Dufourcq P**, Roussigné M, Blader P, Rosa F, Peyrieras N, Vriz S. 2006. Mechano-sensory organ regeneration in adults: the zebrafish lateral line as a model. *Molecular and Cellular Neuroscience* **33**:180–187. DOI: <https://doi.org/10.1016/j.mcn.2006.07.005>, PMID: 16949838
- Ernst S**, Liu K, Agarwala S, Moratscheck N, Avci ME, Dalle Nogare D, Chitnis AB, Ronneberger O, Lecaudey V. 2012. Shroom3 is required downstream of FGF signalling to mediate proneuromast assembly in zebrafish. *Development* **139**:4571–4581. DOI: <https://doi.org/10.1242/dev.083253>, PMID: 23136387
- Faucherre A**, Pujol-Martí J, Kawakami K, López-Schier H. 2009. Afferent neurons of the zebrafish lateral line are strict selectors of hair-cell orientation. *PLoS One* **4**:e4477. DOI: <https://doi.org/10.1371/journal.pone.0004477>, PMID: 19223970
- Ghysen A**, Dambly-Chaudière C. 2007. The lateral line microcosmos. *Genes & Development* **21**:2118–2130. DOI: <https://doi.org/10.1101/gad.1568407>, PMID: 17785522
- Grant KA**, Raible DW, Piotrowski T. 2005. Regulation of latent sensory hair cell precursors by glia in the zebrafish lateral line. *Neuron* **45**:69–80. DOI: <https://doi.org/10.1016/j.neuron.2004.12.020>, PMID: 15629703
- Gratzner HG**. 1982. Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: A new reagent for detection of DNA replication. *Science* **218**:474–475. DOI: <https://doi.org/10.1126/science.7123245>, PMID: 7123245
- Haas P**, Gilmour D. 2006. Chemokine signaling mediates self-organizing tissue migration in the zebrafish lateral line. *Developmental Cell* **10**:673–680. DOI: <https://doi.org/10.1016/j.devcel.2006.02.019>, PMID: 16678780
- Haken H**. 1983. *Synergetics: Non-Equilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology*. 3rd ed. Berlin, Germany: Springer. DOI: <https://doi.org/10.1007/978-3-642-88338-5>
- Harding MJ**, Nechiporuk AV. 2012. Fgfr-Ras-MAPK signaling is required for apical constriction via apical positioning of Rho-associated kinase during mechanosensory organ formation. *Development* **139**:3130–3135. DOI: <https://doi.org/10.1242/dev.082271>, PMID: 22833124
- Harris JA**, Cheng AG, Cunningham LL, MacDonald G, Raible DW, Rubel EW. 2003. Neomycin-induced hair cell death and rapid regeneration in the lateral line of zebrafish (*Danio rerio*). *JARO - Journal of the Association for Research in Otolaryngology* **4**:219–234. DOI: <https://doi.org/10.1007/s10162-002-3022-x>, PMID: 12943374
- Hava D**, Forster U, Matsuda M, Cui S, Link BA, Eichhorst J, Wiesner B, Chitnis A, Abdelilah-Seyfried S. 2009. Apical membrane maturation and cellular rosette formation during morphogenesis of the zebrafish lateral line. *Journal of Cell Science* **122**:687–695. DOI: <https://doi.org/10.1242/jcs.032102>, PMID: 19208766
- Hernández PP**, Moreno V, Olivari FA, Allende ML. 2006. Sub-lethal concentrations of waterborne copper are toxic to lateral line neuromasts in zebrafish (*Danio rerio*). *Hearing Research* **213**:1–10. DOI: <https://doi.org/10.1016/j.heares.2005.10.015>, PMID: 16386394
- Jiang L**, Romero-Carvajal A, Haug JS, Seidel CW, Piotrowski T. 2014. Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. *Proceedings of the National Academy of Sciences* **111**:E1383–E1392. DOI: <https://doi.org/10.1073/pnas.1402898111>, PMID: 24706903
- Jiang T**, Kindt K, Wu DK. 2017. Transcription factor Emx2 controls stereociliary bundle orientation of sensory hair cells. *eLife* **6**:e23661. DOI: <https://doi.org/10.7554/eLife.23661>, PMID: 28266911
- Kindt KS**, Finch G, Nicolson T. 2012. Kinocilia mediate mechanosensitivity in developing zebrafish hair cells. *Developmental Cell* **23**:329–341. DOI: <https://doi.org/10.1016/j.devcel.2012.05.022>, PMID: 22898777
- Kondrychyn I**, Teh C, Garcia-Lecea M, Guan Y, Kang A, Korzh V. 2011. Zebrafish Enhancer TRAP transgenic line database ZETRAP 2.0. *Zebrafish* **8**:181–182. DOI: <https://doi.org/10.1089/zeb.2011.0718>, PMID: 22181660
- Lancaster MA**, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA. 2013. Cerebral organoids model human brain development and microcephaly. *Nature* **501**:373–379. DOI: <https://doi.org/10.1038/nature12517>, PMID: 23995685
- Ledent V**. 2002. Postembryonic development of the posterior lateral line in zebrafish. *Development* **129**:597–604. DOI: <https://doi.org/10.1111/j.1525-142X.2009.00346.x>, PMID: 11830561
- Levin M**. 2009. Errors of geometry: regeneration in a broader perspective. *Seminars in Cell & Developmental Biology* **20**:643–645. DOI: <https://doi.org/10.1016/j.semcdb.2009.05.006>, PMID: 19490948
- López-Schier H**, Hudspeth AJ. 2005. Supernumerary neuromasts in the posterior lateral line of zebrafish lacking peripheral glia. *PNAS* **102**:1496–1501. DOI: <https://doi.org/10.1073/pnas.0409361102>, PMID: 15677337
- López-Schier H**, Hudspeth AJ. 2006. A two-step mechanism underlies the planar polarization of regenerating sensory hair cells. *PNAS* **103**:18615–18620. DOI: <https://doi.org/10.1073/pnas.0608536103>, PMID: 17124170
- López-Schier H**, Starr CJ, Kappler JA, Kollmar R, Hudspeth AJ. 2004. Directional cell migration establishes the axes of planar polarity in the posterior lateral-line organ of the zebrafish. *Developmental Cell* **7**:401–412. DOI: <https://doi.org/10.1016/j.devcel.2004.07.018>, PMID: 15363414
- Ma EY**, Rubel EW, Raible DW. 2008. Notch signaling regulates the extent of hair cell regeneration in the zebrafish lateral line. *Journal of Neuroscience* **28**:2261–2273. DOI: <https://doi.org/10.1523/JNEUROSCI.4372-07.2008>, PMID: 18305259
- Matthews BW**. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **405**:442–451. DOI: [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9), PMID: 1180967
- Meijering E**, Dzyubachyk O, Smal I. 2012. Methods for cell and particle tracking. *Methods in enzymology* **504**:183–200. DOI: <https://doi.org/10.1016/B978-0-12-391857-4.00009-4>, PMID: 22264535
- Murphy KP**. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT press.
- Namdaran P**, Reinhart KE, Owens KN, Raible DW, Rubel EW. 2012. Identification of modulators of hair cell regeneration in the zebrafish lateral line. *Journal of Neuroscience* **32**:3516–3528. DOI: <https://doi.org/10.1523/JNEUROSCI.3905-11.2012>, PMID: 22399774

- Norden C.** 2017. Pseudostratified epithelia - cell biology, diversity and roles in organ formation at a glance. *Journal of Cell Science* **130**:1859–1863. DOI: <https://doi.org/10.1242/jcs.192997>, PMID: 28455413
- Parinov S**, Kondrichin I, Korzh V, Emelyanov A. 2004. Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes *in vivo*. *Developmental Dynamics* **231**:449–459. DOI: <https://doi.org/10.1002/dvdy.20157>, PMID: 15366023
- Pinto-Teixeira F**, Viader-Llargués O, Torres-Mejía E, Turan M, González-Gualda E, Pola-Morell L, López-Schier H. 2015. Inexhaustible hair-cell regeneration in young and aged zebrafish. *Biology Open* **4**:903–909. DOI: <https://doi.org/10.1242/bio.012112>, PMID: 26002931
- Romero-Carvajal A**, Navajas Acedo J, Jiang L, Kozlovskaja-Gumbrienė A, Alexander R, Li H, Piotrowski T. 2015. Regeneration of Sensory Hair Cells Requires Localized Interactions between the Notch and Wnt Pathways. *Developmental Cell* **34**:267–282. DOI: <https://doi.org/10.1016/j.devcel.2015.05.025>, PMID: 26190147
- Rosen JN**, Sweeney MF, Mably JD. 2009. Microinjection of Zebrafish Embryos to Analyze Gene Function. *Journal of Visualized Experiments* **25**:1115. DOI: <https://doi.org/10.3791/1115>
- Rossi AM**, Fernandes VM, Desplan C. 2017. Timing temporal transitions during brain development. *Current Opinion in Neurobiology* **42**:84–92. DOI: <https://doi.org/10.1016/j.conb.2016.11.010>, PMID: 27984764
- Sasai Y.** 2013. Cytosystems dynamics in self-organization of tissue architecture. *Nature* **493**:318–326. DOI: <https://doi.org/10.1038/nature11859>, PMID: 23325214
- Sato T**, Clevers H. 2015. SnapShot: Growing Organoids from Stem Cells. *Cell* **161**:1700–1700.e1. DOI: <https://doi.org/10.1016/j.cell.2015.06.028>, PMID: 26091044
- Sánchez M**, Ceci ML, Gutiérrez D, Anguita-Salinas C, Allende ML. 2016. Mechanosensory organ regeneration in zebrafish depends on a population of multipotent progenitor cells kept latent by Schwann cells. *BMC Biology* **14**:27. DOI: <https://doi.org/10.1186/s12915-016-0249-2>, PMID: 27055439
- Seleit A**, Krämer I, Riebesehl BF, Ambrosio EM, Stolper JS, Lischik CQ, Dross N, Centanin L. 2017. Neural stem cells induce the formation of their physical niche during organogenesis. *eLife* **6**:e29173. DOI: <https://doi.org/10.7554/eLife.29173>, PMID: 28950935
- Shin J**, Chen J, Solnica-Krezel L. 2014. Efficient homologous recombination-mediated genome engineering in zebrafish using TALE nucleases. *Development* **141**:3807–3818. DOI: <https://doi.org/10.1242/dev.108019>, PMID: 25249466
- Simon A**, Berg D, Kirkham M. 2009. Not lost in translation Sensing the loss and filling the gap during regeneration. *Seminars in Cell & Developmental Biology* **20**:691–696. DOI: <https://doi.org/10.1016/j.semcdb.2009.04.015>, PMID: 19463964
- Steiner AB**, Kim T, Cabot V, Hudspeth AJ. 2014. Dynamic gene expression by putative hair-cell progenitors during regeneration in the zebrafish lateral line. *PNAS* **111**:E1393–E1401. DOI: <https://doi.org/10.1073/pnas.1318692111>, PMID: 24706895
- Turner DA**, Baillie-Johnson P, Martinez Arias A. 2016. Organoids and the genetically encoded self-assembly of embryonic stem cells. *BioEssays* **38**:181–191. DOI: <https://doi.org/10.1002/bies.201500111>, PMID: 26666846
- Wada H**, Ghysen A, Asakawa K, Abe G, Ishitani T, Kawakami K. 2013. Wnt/Dkk negative feedback regulates sensory organ size in zebrafish. *Current Biology* **23**:1559–1565. DOI: <https://doi.org/10.1016/j.cub.2013.06.035>, PMID: 23891113
- Wibowo I**, Pinto-Teixeira F, Satou C, Higashijima S, López-Schier H. 2011. Compartmentalized Notch signaling sustains epithelial mirror symmetry. *Development* **138**:1143–1152. DOI: <https://doi.org/10.1242/dev.060566>, PMID: 21343366
- Williams JA**, Holder N. 2000. Cell turnover in neuromasts of zebrafish larvae. *Hearing Research* **143**:171–181. DOI: [https://doi.org/10.1016/S0378-5955\(00\)00039-3](https://doi.org/10.1016/S0378-5955(00)00039-3), PMID: 10771194
- Willyard C.** 2015. The boom in mini stomachs, brains, breasts, kidneys and more. *Nature* **523**:520–522. DOI: <https://doi.org/10.1038/523520a>, PMID: 26223610
- Wymeersch FJ**, Huang Y, Blin G, Cambray N, Wilkie R, Wong FC, Wilson V. 2016. Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* **5**:e14830. DOI: <https://doi.org/10.7554/eLife.10042>, PMID: 26780186
- Xiao Y**, Faucher A, Pola-Morell L, Heddleston JM, Liu TL, Chew TL, Sato F, Sehara-Fujisawa A, Kawakami K, López-Schier H. 2015. High-resolution live imaging reveals axon-glia interactions during peripheral nerve injury and repair in zebrafish. *Disease Models & Mechanisms* **8**:553–564. DOI: <https://doi.org/10.1242/dmm.018184>, PMID: 26035865

## Appendix C

Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatio-temporal patterns.

Lupperger, V., Marr, C., Chapouton, P. (2020). Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatio-temporal patterns. *PLoS Biology*, 18(12), e3000708..

## SHORT REPORTS

# Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatiotemporal patterns

Valerio Lupperger<sup>1</sup>, Carsten Marr<sup>1\*</sup>, Prisca Chapouton<sup>2,3‡\*</sup>

**1** Helmholtz Zentrum München—German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany, **2** Helmholtz Zentrum München—German Research Center for Environmental Health, Unit Sensory Biology & Organogenesis, Neuherberg, Germany, **3** Helmholtz Zentrum München—German Research Center for Environmental Health, Institute of Stem Cell Research, Biomedical Center, Faculty of Medicine, LMU Munich, Planegg, Germany

‡ Lead contact

\* [carsten.marr@helmholtz-muenchen.de](mailto:carsten.marr@helmholtz-muenchen.de) (CM); [chapouton@helmholtz-muenchen.de](mailto:chapouton@helmholtz-muenchen.de) (PC)



## OPEN ACCESS

**Citation:** Lupperger V, Marr C, Chapouton P (2020) Reoccurring neural stem cell divisions in the adult zebrafish telencephalon are sufficient for the emergence of aggregated spatiotemporal patterns. *PLoS Biol* 18(12): e3000708. <https://doi.org/10.1371/journal.pbio.3000708>

**Academic Editor:** Catarina C. Homem, Universidade Nova de Lisboa Faculdade de Ciencias Medicas, PORTUGAL

**Received:** February 28, 2020

**Accepted:** November 17, 2020

**Published:** December 8, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3000708>

**Copyright:** © 2020 Lupperger et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** We provide access to raw image data, processed single cell information, code for analysis, parameter inference and result

## Abstract

Regulation of quiescence and cell cycle entry is pivotal for the maintenance of stem cell populations. Regulatory mechanisms, however, are poorly understood. In particular, it is unclear how the activity of single stem cells is coordinated within the population or if cells divide in a purely random fashion. We addressed this issue by analyzing division events in an adult neural stem cell (NSC) population of the zebrafish telencephalon. Spatial statistics and mathematical modeling of over 80,000 NSCs in 36 brain hemispheres revealed weakly aggregated, nonrandom division patterns in space and time. Analyzing divisions at 2 time points allowed us to infer cell cycle and S-phase lengths computationally. Interestingly, we observed rapid cell cycle reentries in roughly 15% of newly born NSCs. In agent-based simulations of NSC populations, this redividing activity sufficed to induce aggregated spatiotemporal division patterns that matched the ones observed experimentally. In contrast, omitting redivisions leads to a random spatiotemporal distribution of dividing cells. Spatiotemporal aggregation of dividing stem cells can thus emerge solely from the cells' history.

## Introduction

Somatic stem and progenitor cells, basic units of tissue maintenance and growth, can be found in distinct states, either dividing or quiescent. The duration of the quiescence state is not predictable [1], and its regulation has a profound impact on healthy tissue maintenance. Therefore, understanding the mechanisms of adult stem cell cycle regulation is crucial. Stem cells, when compartmentalized in areas where neighboring cells fulfill structural and niche functions, can be regulated locally, such as the hair follicle in the skin [2]. However, in a homogeneous population of equipotent stem cells that reside under the same conditions, it is unclear what drives distinct behaviors of quiescence or cell cycle entry. This raises the question of how

figures at <https://github.com/QSCD/spatiotemporalAnalyses>. Raw image data can be accessed from <https://hmgubox2.helmholtz-muenchen.de/index.php/s/MBrLMra7wpj2qKn>.

**Funding:** V.L. was funded by the Bundesministerium für Bildung und Forschung (BMBF, <https://www.gesundheitsforschung-bmbf.de/de/micmode-i2t-modulare-bildanalyseplattform-zur-integration-von-mikroskopischen-bildbasierten-7611.php>), grant 01ZX1710A-F (Micmode-I2T). C.M. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 866411). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** ABC, approximate Bayesian computation; BrdU, 5-bromo-2'-deoxyuridine; CDK, cyclin-dependent kinase; CI, confidence interval; DLS, double-labeled S-phase; EdU, 5-Ethynyl-2'-deoxyuridine; EGFP, enhanced green fluorescent protein; GFP, green fluorescent protein; NSC, neural stem cell; SCIP, Single Cell Identification Pipeline; SD, standard deviation.

stem cells make the decision to remain quiescent or enter the cell cycle individually and collectively.

Several determinants have been proposed as regulators of quiescence or division of stem and progenitor cells. Feedback mechanisms between stem cells and the environment have been found in several systems: In the adult mouse forebrain, neuronal activity in mossy cells or in granule cells regulates the activation of radial neural stem cells (NSCs) of the dentate gyrus [3,4]. In the subventricular zone, NSC cycling is inhibited by direct contacts with endothelial cells [5] or by the release of miR-204 from the choroid plexus [6]. In mouse and zebrafish adult neurogenic zones, Notch activity maintains NSCs in quiescence [7–9], in particular in the immediate neighborhood of dividing cells [7]. In other adult stem cell populations such as the epidermis, cell divisions are instructed by neighboring differentiating progeny [10,11]. Besides signals from the environment, cell-intrinsic modulations have been shown to impact on the proliferative activity, for instance, the metabolic control of lipogenesis that can induce proliferation [12,13]. Conversely, the expression of miR-9 [14] and the degradation of *Ascl1* via the ubiquitin ligase *Huwe1* [15] are factors inducing quiescence. While a combination of signals received from the environment and intrinsic to the cells seems to influence proliferating activity, it remains to be precisely understood whether and how cells coordinate their activity with their neighbors.

Toward this end, we examined the distribution of cells in S-phase in the pallial (dorsal) neurogenic niche of the adult zebrafish telencephalon, which is located on the outer surface of the brain [16,17]. The adult brain of the zebrafish and the telencephalon in particular is growing steadily [18] but at a slower pace with increasing age [19]. Radial glia constitute the entire ventricular surface of the brain with a single layer of cell somata, extending filopodial extensions tangentially [20] and long ramified processes through the parenchyme. They express—among others—Notch ligands, the transcription factors *Her4* and *Fezf2*, the fatty acid binding protein *BLBP*, the enzyme *Aromatase B*, the intermediate filaments *GFAP* and *vimentin* [21–25] and a small percentage of them is dividing at any time point of observation [7]. The behavior of radial glia in steady state or injury conditions in this area of the brain hints to their function as NSCs giving rise to intermediate dividing progenitors and directly to neurons [23,26–28]. We previously found that cell cycle entries in NSCs occur with aggregated spatial patterns [29]. Here, we show that NSCs may undergo successive S-phases with short time windows, and modeling their activity creates similar spatiotemporal patterns as the ones observed experimentally.

## Results

### Neural stem cells in S-phase reveal aggregated spatial patterns on the dorsal ventricular zone

To understand cell cycle regulation in a population of mainly quiescent stem cells, we studied the spatial distribution of cells in S-phase (see [Box 1](#)) in the intact dorsal telencephalon (pallium) in whole mount preparations. We used adult *gfap:GFP* transgenic zebrafish, where NSCs express enhanced green fluorescent protein (EGFP) [30]. In order to assess which subset of cells is concomitantly in S-phase at a specific time point, we marked S-phases by the incorporation of the thymidine analog EdU 1 h before fixation ([Fig 1A](#)) and detected them by staining in whole mount preparations, followed by confocal microscopy (3D reconstruction shown in [Fig 1B](#)). We identified the 3D coordinates of *gfap:GFP+* NSCs ([Fig 1C, 1D and 1D'](#)) and of EdU+ nuclei ([Fig 1C' and 1D](#)) automatically. To analyze the spatial pattern of S-phases, we used the coordinates of all *gfap:GFP+* NSCs ([Fig 1D'](#)) as a reference grid (see [S1A Fig](#)). Via manual inspection, we then discriminated between EdU+*gfap:GFP+* cells, representing NSCs

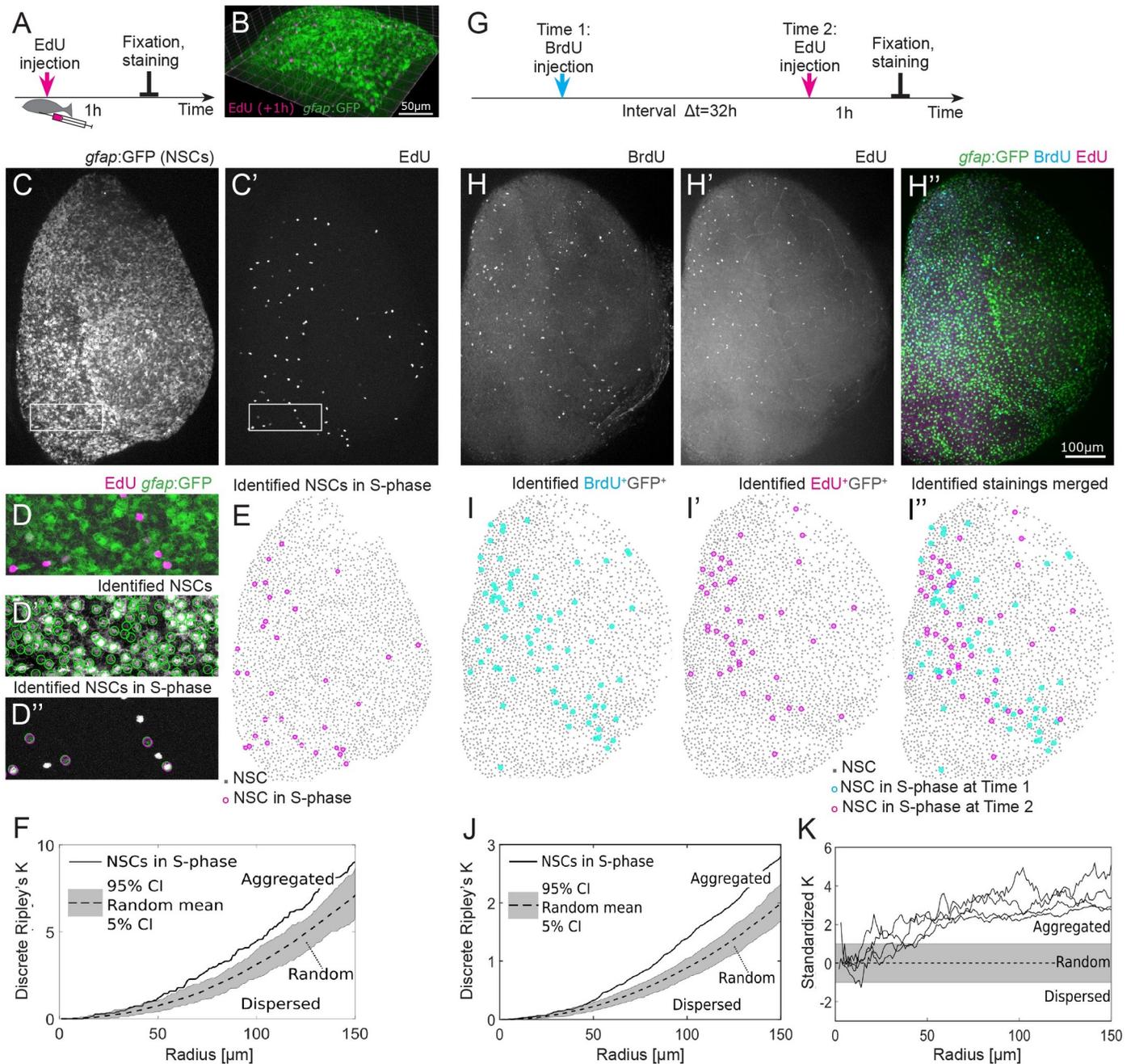
### Box 1. Statistical analysis of spatiotemporal point patterns

When events can happen at any point in a 2D Euclidean space, a **spatial Poisson process** leads to a point pattern that is completely defined by a single parameter, the density of points. This **null model of complete spatial randomness** has been used for diverse analyses, from forest structures [35] over accessibility of pediatric care [36] to road accident prevention [37]. In our case, the space analyzed is not Euclidean, but **discrete**, since NSCs in S-phase (the “events”) only appear where NSC are already present in the zebrafish hemisphere. The corresponding null model is thus not a spatial Poisson process, but represented by the distribution of randomly sampled NSCs. Accordingly, we call a pattern of NSC in S-phase **random** if the distances between the events are not significantly different from randomly sampled NSCs in S-phase. We evaluate this with an adapted version of a spatial statistics measure called **Ripley’s K** [31]. **Nonrandom patterns** can be classified into 2 types: When events are closer to each other than expected from the random null model, we call the pattern **aggregated**. If events are further apart from each other, we call the pattern **dispersed**. A simulated random division pattern is displayed in [S1B Fig](#), and a simulated dispersed division pattern on the same hemisphere is shown in [S1D Fig](#). The corresponding observed aggregated division pattern (with the same number of events) is shown in [Fig 1E](#). We follow in our notation and nomenclature the book from Baddeley et al. [38], which provides an overview and examples of diverse spatial statistics methods.

in S-phase, and EdU+*gfap*:GFP<sup>-</sup> cells, representing intermediate progenitors in S-phase [23] ([Fig 1D” and 1E](#)). We quantified the distribution of NSCs in S-phase ([Fig 1F](#)) with an adapted discrete variant of Ripley’s K [31] that measures the number of neighboring NSCs in S-phase observed in a particular radius, accounting for a possibly nonhomogeneous distribution of nondividing NSCs and edge effects (see [Methods](#)). In other biological contexts, Ripley’s K analysis has been applied to analyze single-molecule localization, reviewed in [32], or to analyze single cells behavior [33]. According to this spatial statistics measure, NSCs in S-phase reveal an aggregated pattern ([Fig 1F](#)), which is significantly deviating from a random ([S1B and S1C Fig](#)) and dispersed pattern ([S1D and S1E Fig](#)) for radii >50  $\mu\text{m}$ . Aggregated patterns of NSCs in S-phase were reproducibly found in approximately 70% of all hemispheres analyzed ([S1F Fig](#)), suggesting that S-phase entry happens in a spatially nonrandom manner. In the remaining hemispheres (30%), the observed patterns are random ([S1G Fig](#)). Following distinctions in the division activity described previously in Dray and colleagues [34], we compared division proportions and spatial patterns in 3 different hemisphere domains (lateral, medial, and anterior) and found that these domains also show aggregated and random, but no dispersed patterns. Since separate domains contain in approximately 40% of all cases less than 10 NSCs in S-phase and patterns often span across domains, we refrain from analyzing them separately.

### Neural stem cells in S-phase reveal aggregated spatiotemporal patterns

To investigate whether the spatial patterns of S-phases are influenced by previous cell cycle activity, we made use of a second thymidine analog, 5-bromo-2'-deoxyuridine (BrdU), and observed consecutive S-phases taking place in vivo. The 2 labels BrdU and EdU were administered by intraperitoneal injections separated by a labeling interval  $\Delta t$  of 32 h ([Fig 1G](#)). Fish were humanely killed, their brain dissected and fixed 1 h after the second administration



**Fig 1. S-phases occurring in the NSC population are spatiotemporally aggregated.** (A) Experimental setup: EdU is injected intraperitoneally 1 h prior to humanely killing the fish and fixation of the brain. (B) Part of the telencephalic hemisphere as a 3D reconstruction. The *gfap:GFP* transgene highlights cell bodies of NSCs, which are arranged on a 2D layer on the telencephalon surface, while their radial processes project deep in the parenchyma. (C, C') *gfap:GFP* in the whole hemisphere shown as a maximum intensity z-stack projection, anterior to the top, lateral to the right, and medial to the left. Boxed areas are shown as higher magnifications in (D–D''). (C'') EdU coupled to Azyde-Alexa 647 highlights cells in S-phase and reveals their spatial distribution. (D) Merged GFP and EdU channel to identify specifically the NSCs in S-phase. (D'') Automatically identified NSCs surrounded by green circles. (D''') EdU+ cells were subdivided into EdU+*gfap:GFP*<sup>−</sup> and EdU+*gfap:GFP*<sup>+</sup>, the latter representing NSCs in S-phase. (E) The 33 NSCs in S-phase (pink circles) exhibit a nonrandom spatial pattern on top of all 2678 NSCs (gray dots). (F) Discrete Ripley's K quantification of the pattern shown in (E) reveals that NSCs in S-phase (solid line) are aggregated, i.e., closer to each other than expected from random (dotted line with 90% CI in gray) and dispersed patterns. (G) Cells in S-phase are labeled with BrdU and EdU with an interval of 32 h. Fish are humanely killed 1 h after the EdU injection, and the brains are imaged after fixation and staining. (H–H'') Example of 1 telencephalic hemisphere, oriented as in (C), as a maximum intensity z-stack projection, in 3 different channels: BrdU (H), EdU (H'), and *gfap:GFP* transgene highlighting NSC bodies merged with the EdU and BrdU staining (H''). Scale bar: 100 μm. (I–I'') Identified NSCs in S-phase at 2 different time points exhibit aggregated spatiotemporal patterns. (J) Discrete Ripley's K reveals more EdU+ NSCs around BrdU+ NSCs as expected from a random process. (K) We find spatiotemporally aggregated patterns with radii above 50 μm in all 4 hemispheres where S-phases have been labeled with a labeling interval of 32 h. BrdU, 5-bromo-2'-deoxyuridine; CI, confidence interval; EdU, 5-Ethynyl-2'-deoxyuridine; GFP, green fluorescent protein; NSC, neural stem cell.

<https://doi.org/10.1371/journal.pbio.3000708.g001>

(Fig 1G). EdU+ and BrdU+ cells were observed as clearly distinct sets (Fig 1H–1H’). When administered simultaneously as a control, EdU and BrdU reliably labeled the same set of cells (S1H Fig). We identified the 3D coordinates of the BrdU+ and EdU+ NSCs (Fig 1I–1I’’) and observed that some areas of the ventricular zone remain devoid of S-phase NSCs at those 2 time points. This observation was confirmed by a spatiotemporal Ripley’s K analysis: From a radius  $>50\ \mu\text{m}$ , the density of EdU+ NSCs around BrdU+ NSCs was higher than expected from random (Fig 1J). Thus, NSCs in S-phase aggregate also spatiotemporally, a trend that was observed in all 4 hemispheres with a labeling interval of 32 h (Fig 1K). These results suggest that subsequent S-phases are not randomly distributed and that the spatial organization observed is linked to the past activity in the population.

We then broadened the range of labeling intervals from 9 h to 72 h (S2A Fig). Using the BrdU–EdU double-labeling approach, we processed in total 36 hemispheres and identified NSCs and NSCs in S-phase at both time points (S1 Table). Visually, the observed patterns vary from homogeneous point clouds to strongly confined regions (Fig 2A). Quantitatively, we identify in 22 out of 36 hemispheres spatiotemporally aggregated patterns of divisions between 2 labeling time points, while the remaining 14 patterns are classified as random (see S2B Fig).

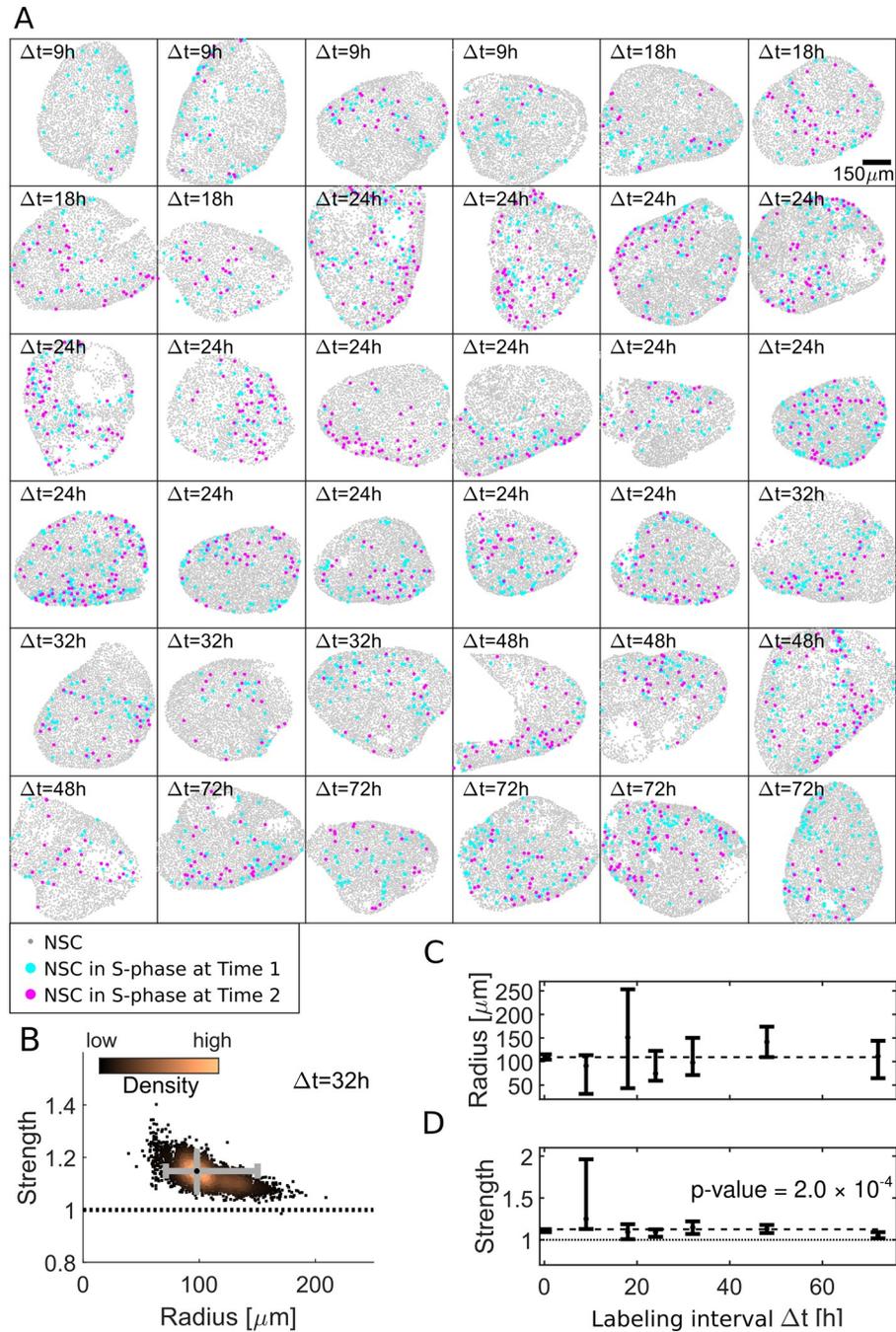
### An interaction model fits the observed spatiotemporal patterns

Ripley’s K statistics is limited: It does not allow integrating different datasets, and it cannot quantitatively infer the strength and range of an observed pattern. To remedy these aspects, we use the temporal extension of a spatial model [29] that allows inferring the most likely parameters for interaction strength and interaction radius for an arbitrary number of datasets. Aggregated patterns emerge for an interaction strength  $>1$ , random patterns for strength = 1, and dispersed patterns for strength  $<1$  (see Methods).

Applied to the 4  $\Delta t = 32\ \text{h}$  patterns shown in Fig 2A, we find that a model with an interaction radius of approximately  $100\ \mu\text{m}$  and an interaction strength  $>1$  describes the data best (Fig 2B). On average, we find  $202 \pm 63$  NSCs in total (mean  $\pm$  standard deviation (SD) from  $n = 36$  brains) and  $6 \pm 4$  NSCs in S-phase in a  $100\text{-}\mu\text{m}$  radius around an NSC in S-phase. Posterior sampling reveals a 90% confidence interval (CI) from 71 to  $150\ \mu\text{m}$ , a maximum likelihood interaction radius at  $98\ \mu\text{m}$ , and a 90% CI from 1.07 to 1.22 with a maximum likelihood interaction strength of 1.15. Applied to all 36 hemispheres with coordinates of 87,807 NSCs at distinct labeling intervals  $\Delta t$  (see S2C Fig), we find the most likely interaction radii to be around  $100\ \mu\text{m}$  (Fig 2C). The interaction strength (Fig 2D) is robustly above 1, indicating a significant ( $p$ -value = 0.0002 for a linear regression model) spatiotemporal aggregation of S-phase NSCs for all labeling intervals  $\Delta t$ .

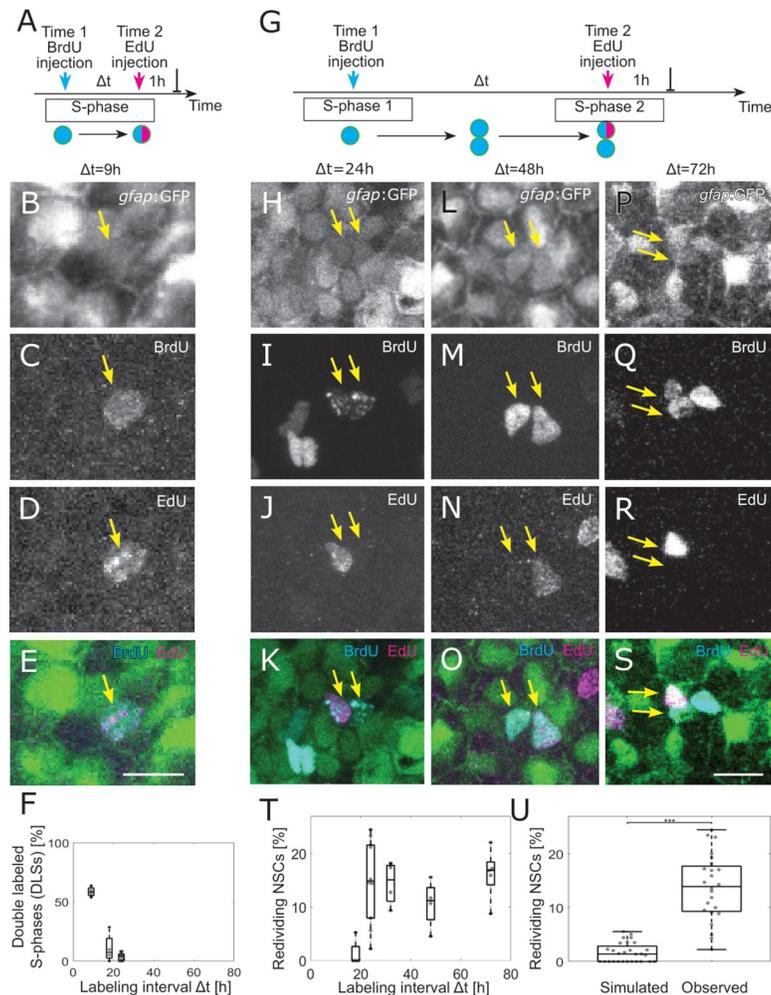
### Division of NSC daughter cells reoccur within 24 to 72 h

Our analysis revealed a weak aggregation of successive NSCs in S-phase. Aggregated spatiotemporal patterns can emerge out of different mechanisms: Signaling waves that stimulate cells to divide in a particular region, activation via cell–cell contacts, or a behavior specified by the particular state of a cell. We thus inquired whether an NSC’s history would be of relevance for new S-phase entries and made use of the double-labeling approach introduced above. Short labeling intervals up to  $\Delta t = 24\ \text{h}$  revealed cells in S-phase incorporating both labels (Fig 3A–3F), hence denoted as double-labeled S-phases (DLSs). Remarkably, we also noticed cells with both S-phase markers that, unlike DLS, were arranged as doublets, i.e., 2 daughter cells close to each other. Such doublets allow us to extrapolate from the incorporation of S-phase labels to actual cell division events. One cell of these doublets entered a second S-phase, and these observations, denoted as redivisions, occurred with labeling intervals of  $\Delta t = 24\ \text{h}$ ,  $32\ \text{h}$ ,



**Fig 2. Computational approach identifies an approximately 100- $\mu\text{m}$  aggregation radius of NSCs in S-phase.** (A) Our dataset comprises 36 hemispheres with labeling intervals from  $\Delta t = 9\text{ h}$  to  $72\text{ h}$ . (B) Posterior sampling identifies the most likely interaction strength of 1.15 and most likely interaction radius of  $98\ \mu\text{m}$  for 4  $\Delta t = 32\text{ h}$  hemispheres. Whiskers (gray) cover the 95% CIs for strength and radius. Sampling point density is visualized from copper (high) to black (low). (C) Applied to all 36 hemispheres posterior sampling reveals an interaction radius around  $100\ \mu\text{m}$ . (D) The interaction strength is significantly above 1 for all labeling intervals  $\Delta t$  ( $p$ -value = 0.0002 for constant fit to most likely values) thus inducing aggregated patterns. CI, confidence interval; NSC, neural stem cell.

<https://doi.org/10.1371/journal.pbio.3000708.g002>



**Fig 3. A large proportion of NSCs redivide within 24 to 72 h.** (A) The 2 injections label the same cells in S-phase for small-labeling intervals, leading to NSCs that are both EdU and BrdU positive, denoted as double-labeled S-phase (DLS). (B–E) Example DLS (yellow arrow) for a labeling interval  $\Delta t = 9$  h. (F) The DLS proportion is high for  $\Delta t = 9$  h and decreases rapidly with increasing  $\Delta t$ . Each dot represents the value for 1 brain hemisphere. (G) After a division, 1 of the daughter cells already labeled by the Time 1 label can enter in a new S-phase and incorporate a second label. This cell thereby redivides. (H–S) Three examples of redividing NSCs with labeling intervals of 24 h (H–K), 48 h (L–O), and 72 h (P–S). Scale bar: 10  $\mu$ m. (T) The proportions of redividing NSCs within the dividing NSCs at Time 1 remain high from  $\Delta t = 24$  h to  $\Delta t = 72$  h labeling intervals. Each dot represents 1 brain hemisphere. (U) Only  $1.9 \pm 1.7\%$  of randomly drawn divisions (same amount as observed per hemisphere) from all NSCs would be redrawn at random, while  $14 \pm 8\%$  of observed NSCs in S-phase reenter S-phase ( $p = 9.4 \times 10^{-10}$ , 2-sample Kolmogorov–Smirnov test). Box plots range from the 25th to the 75th percentile, and the central mark indicates the median and whiskers include points that are not more than 1.5 times the interquartile range away from the top or bottom of the box. BrdU, 5-bromo-2'-deoxyuridine; DLS, double-labeled S-phase; EdU, 5-Ethynyl-2'-deoxyuridine; GFP, green fluorescent protein; NSC, neural stem cell.

<https://doi.org/10.1371/journal.pbio.3000708.g003>

48 h, and 72 h (Fig 3G–3T). The observed redivision frequency of NSCs is significantly higher than expected from random: While  $14 \pm 8\%$  of NSCs in S-phase reenter S-phase, only  $1.9 \pm 1.7\%$  of randomly drawn divisions (same amount as observed per hemisphere) would be redrawn at random again ( $p = 9.4 \times 10^{-10}$ , 2-sample Kolmogorov–Smirnov test, Fig 3U). We are confident that we are observing this phenomenon in stem cells, as we could clearly distinguish between the events taking place in *gfap:GFP*<sup>+</sup> and *gfap:GFP*<sup>−</sup> progenitors (S3A Fig). Reoccurring divisions in NSCs happened in similar proportions as in the *gfap:GFP*

– progenitors, the latter being considered so far as transit amplifying progenitors (S3B and S3C Fig and S1 Data, sheet S3C Fig). Notably, the *gfap*:GFP marker was the original label of the NSCs, without additional immunostaining for GFP, thus truly representing NSCs. Hence, S-phase NSCs display a high likelihood of undergoing another division within the following days.

To support these results, we performed independent experiments with a third time point of observation. We injected first BrdU, then EdU 24 h later, and dissected the fish another 24 h later, adding a PCNA immunostaining as a third cell cycle marker (S3B–S3Q Fig). In those brains, we found all combinations of NSC redivisions: BrdU+ daughter cells that incorporated EdU, BrdU+ daughter cells that expressed PCNA, and EdU+ daughter cells that expressed PCNA (S3K Fig and S1 Data, sheet S3K Fig for statistics). EdU+ doublets, which were also PCNA+ might represent cells that reached the end of a cell cycle without necessarily reentering a second division round. However, we also observed EdU+ doublets in which only 1 daughter cell was PCNA+, signifying a specific cell cycle reentry of this daughter cell.

We could not assess how many rounds of reoccurring divisions NSCs may undergo maximally, as an increasing number of BrdU-labeled cells with an increasing chase time renders a discrimination between neighboring clones impossible. However, experiments carried out with 72-h labeling interval revealed the presence of triplets of BrdU-labeled *gfap*:GFP+ cells (Fig 3Q), indicating that at least 2 rounds of divisions have been taking place within this time window. This implies that the distribution of cell cycle entries in the NSC population is also a result of the recent history of individual cells.

To assess whether cellular niches might be involved in the observed reoccurring divisions, e.g., via the formation of groups of NSCs with distinctive volumes, we segmented single NSCs in 3D whole mount brain images of 4 hemispheres and measure their volumes (see Methods and S3L–S3P Fig). We considered 3 groups of NSCs: (i) PCNA+ dividing NSCs; (ii) PCNA+BrdU+ and PCNA+EdU+ redividing NSCs; and (iii) nondividing NSCs without any marker. Measuring the volumes of all these NSCs and all immediately touching neighbors did not reveal any significant difference between the 3 groups (Kruskal–Wallis test,  $p$ -values = 0.1 and 0.37, respectively; S3Q Fig). Hence, the NSC volume does not reveal any distinctive organization around actively dividing NSCs, arguing against a model where cell density dependent niches would be associated with NSC activity.

### An agent-based model with redividing NSCs recreates aggregated spatiotemporal patterns observed experimentally

To quantitatively evaluate if redivisions suffice to induce the observed aggregated patterns, we simulated dividing NSCs with an agent-based spatiotemporal model. Such a model simulates the actions and interactions of autonomous agents, in our case NSCs. Here, every single cell can be modeled at every time point, while in other non-agent-based approaches, one only obtains summary statistics or averages per time point. However, fitting a spatiotemporal model to data is extremely challenging, since parameter estimation relies on repeated, computationally expensive simulations [39]. We thus split our simulation approach into 2 steps: First we fit a simple, nonspatial cell division model to the observed DLS and redividing NSCs for different labeling intervals  $\Delta t$  (Fig 3F and 3T) to derive the length and variability of cell cycle and S-phase. Second, we simulate an agent-based spatiotemporal model with the inferred parameters, analyze the simulated patterns statistically, and compare the results with experimental data.

To infer cell cycle kinetics from our data, we implemented a cell division model of dividing NSCs with 5 parameters (see Methods): the minimal cell cycle length  $d_{cc}$  and minimal S-phase length  $d_{sp}$ , their variability  $\beta_{cc}$  and  $\beta_{sp}$  parametrizing a lag-exponential [40] distribution, and

the redivision probability  $p_{\text{rediv}}$  (see [S4A Fig](#)). From this cell division model, we simulate dividing NSCs and a BrdU-EdU-labeling experiment ([S4B Fig](#)) and evaluate the percentage of redividing cells and DLS. We optimized the 5 parameters to the observed frequencies (see [Methods](#)) using approximate Bayesian computation (ABC) [41]. Our cell division model fitted the data, in particular the plateau of redividing cells ([S4C Fig](#)), and the sharp decrease of DLS cells after 9 h ([S4D Fig](#)). It estimated a minimal cell cycle time of 22.2 h with a mean cell cycle time of 107.5 h, a minimal S-phase length of 16.6 h with a mean S-phase length of 18.2, and a redivision probability of 0.38 ([S4E and S4F Fig](#)). Note that while the redivision probability is a parameter of our cell division model, the redivision fraction is an experimentally observable variable, which depends on the measurement method. Using snapshot measurements, we found a redivision fraction of 15%, which is considerably lower than the redivision probability of 38% (see [S4B Fig](#) for a detailed explanation).

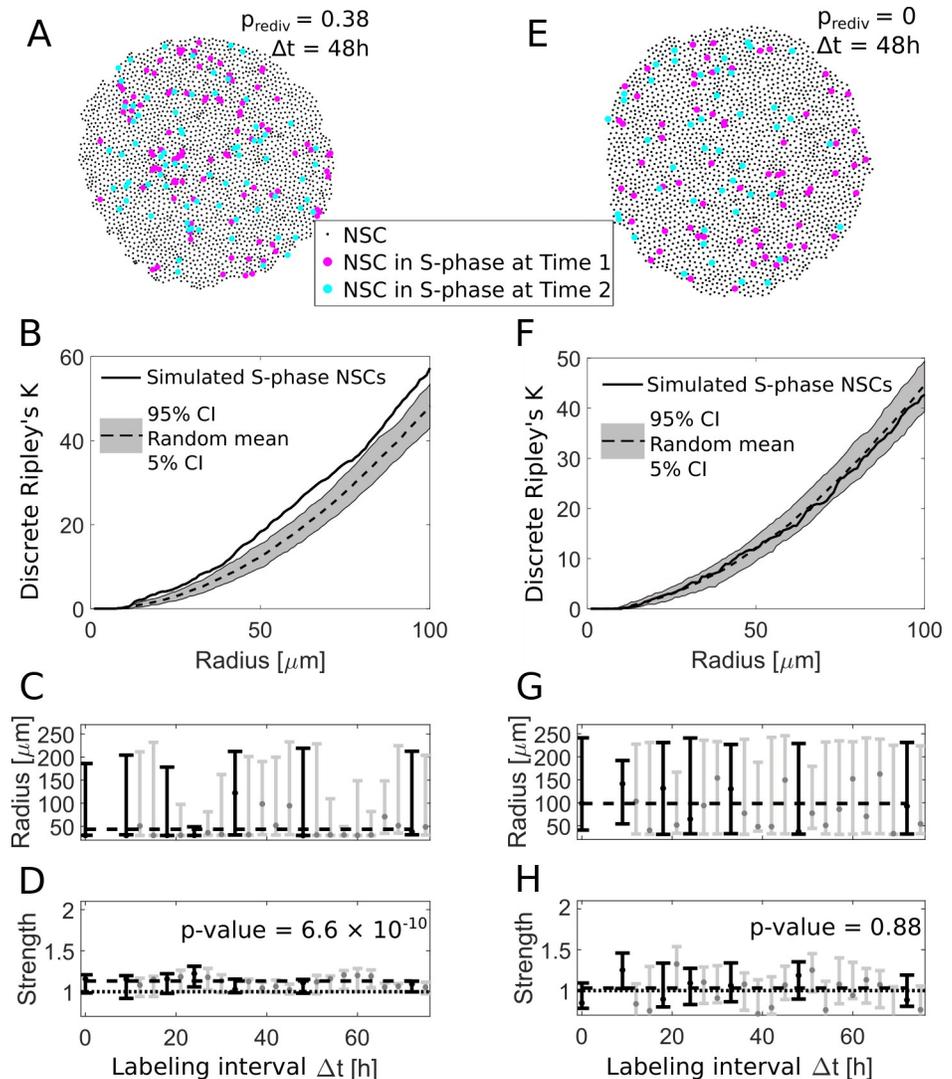
We now fed an agent-based model with the inferred parameters to generate spatiotemporal patterns (see [Methods](#) and [S4G and S4H Fig](#)). To model cell shape kinetics, cellular potts models [42], vertex models [43], and cellular automata [44] have been used. We chose a cellular potts model, since it allows for arbitrary shapes and stochasticity in cell movement and is easily applicable as the default implementation in the Morpheus environment [45]. After a transient phase, we simulated a first measurement by marking all cells in S-phase at that time point. We simulated NSC kinetics further for different S-phase labeling intervals and simulated a second measurement, analogously to the first. We then created a distance matrix for S-phase cells at the 2 time points (exemplary patterns shown in [Fig 4A](#) for  $\Delta t = 48$  h), calculated the discrete Ripley's K (see [Fig 4B](#)), and inferred interaction radius and strength, analogously to experimental data processing. Analogously to our experimental data, we found an interaction strength significantly  $>1$  ( $p$ -value =  $6.6 \times 10^{-10}$ ) and an interaction radius around  $45 \mu\text{m}$  ([Fig 4C and 4D](#)), which was a bit smaller than experimentally observed. We could also reproduce the observed heterogeneity in the emergence of the spatiotemporal patterns: In 36 simulated hemispheres, we found  $58 \pm 8\%$  aggregated patterns ( $n = 10$  independent simulations, mean  $\pm$  SD) in accordance with our experimental observations (61%, see [S2B Fig](#)). Also in the simulations, the remaining patterns are classified as random throughout.

In control simulations, we omitted redivisions. In that case, all NSCs entered division with a rate of  $1 \times 10^{-3}$  per hour (see [Methods](#)). In the corresponding simulated measurements ([Fig 4E](#)), neither the discrete Ripley's K (see [Fig 4F](#)) nor interaction parameters hint to an aggregation of divisions. While the interaction radius strongly varied around a mean value of  $89 \pm 42 \mu\text{m}$  ([Fig 4G](#)), the interaction strength is  $0.97 \pm 0.2$ , indicating random division patterns ([Fig 4H](#)). Thus, with redivisions and no additional assumptions, our agent-based simulations created aggregated spatiotemporal patterns of NSCs in S-phase as observed experimentally.

Finally, we analyze whether the agent-based simulations are able to reproduce the input parameters, i.e., observed DLS and redivision fractions, cell cycle length, and redivision probability. Redivision fractions and DLS fractions from simulations are similar to the ones observed in experimental data ([S4I and S4J Fig](#)), and cell cycle lengths observed in the simulations fit nicely to the probability density function of the lag-exponential distribution ([S4K Fig](#)). S-phase lengths are correct by default as we draw them after simulation for post-simulation analysis, from the respective lag-exponential distribution. The proportion of redividing cells in the simulation is confirmed with  $38 \pm 0.5\%$  ([S4L Fig](#)).

## Discussion

In this study, we detected nonrandom, aggregated spatiotemporal patterns of successively dividing NSCs in the zebrafish brain using an experimental double-labeling approach and



**Fig 4. An agent based redivision model can explain spatial aggregation of NSCs in S-phase.** (A) We use an agent-based model to simulate NSC divisions with a redivision probability of  $p_{\text{rediv}} = 0.38$  and perform virtual measurements with labeling intervals  $\Delta t$  between 9 h and 72 h (here shown for  $\Delta t = 48$  h). (B) The simulated NSCs in S-phase in (A) exhibit an aggregated spatiotemporal pattern according to the discrete Ripley's K curve (solid line) which is above the 90% CI of randomly sampled patterns (gray area), similar to experimentally observed patterns. (C) The fitted radii for simulations with redividing NSCs for different labeling intervals  $\Delta t$  are variable with a maximum likelihood value of 50  $\mu\text{m}$ . Labeling intervals that are also available from experimental data (see Fig 2C) are shown in black, all others in gray. (D) The respective fitted strengths values are above 1 indicating aggregated patterns. Fitting a constant model to the most likely values with the same  $\Delta t$  as experimentally observed (black bars) reveals a significant shift ( $p$ -value =  $6.6 \times 10^{-10}$ ) from a strength of 1. Labeling intervals that are also available from experimental data (see Fig 2D) are shown in black, all others in gray. (E) Simulated NSC divisions and virtual measurements with a redivision probability  $p_{\text{rediv}} = 0$  at 2 labelings  $\Delta t = 48$  h apart. (F) Without redivisions, the simulated S-phase NSCs are within the boundaries of random patterns. (G) The fitted radii are highly variable with maximum likely values from 50 to 150  $\mu\text{m}$ . Labeling intervals that are also available from experimental data (see Fig 2C) are shown in black, all others in gray. (H) In contrast to the simulations with redivisions, we now find no indication for aggregated patterns for all labeling intervals ( $p$ -value = 0.88 for a constant model with nonzero shift from a strength of 1). Labeling intervals that are also available from experimental data (see Fig 2D) are shown in black, all others in gray. CI, confidence interval; NSC, neural stem cell.

<https://doi.org/10.1371/journal.pbio.3000708.g004>

extended methods from spatial statistics for quantification and parameter inference. Moreover, we found a prominent cell cycle reentry in daughter NSCs.

The quantification of patterns of NSCs in S-phase is challenged by their heterogeneity and the relatively weak signal: Only 70% of the 36 hemispheres show aggregated spatial patterns, the remaining 30% qualify as random, so we wondered how this heterogeneity comes about. Since we estimate NSCs to redivide with only 38% probability, we believe that de novo divisions of previously quiescent cells dilute the aggregation patterns arising from reoccurring divisions. Our agent-based simulations are able to reproduce aggregated patterns and their frequency, however, with a smaller interaction radius than observed. Contributions to this deviation might come from the abstracted morphology of the hemisphere and individual cell shapes (see S4G and S4H Fig), the discrepancy between a 2D model and 3D effects in the brain, or the previously reported underestimation of radii when using a Euclidean distance measure [29].

A cell division model allowed us to estimate cell cycle and S-phase length distributions from double-labeling data. It is interesting to consider that, similarly to variations in the G1 and G2 phases [46,47], S-phase length can be variable too. Lengthening of S-phase occurs during, e.g., development [48,49]. The work by Arai and colleagues [50] and Turrero García and colleagues [51] report distinct lengths of S-phases according to distinct types of neural progenitors in the developing neocortex in mouse and ferret, respectively. In the adult mouse subependymal zone [52], the *gfap*<sup>+</sup> B-type cells displayed very rapid S-phases of about 4 to 5 h, which is below the values we measured here of about 18 h by identifying double-labeled nuclei (see Fig 3F). Since we still detected occasional double-labeled nuclei with 24-h intervals, but not later, we can state with confidence that these nuclei were already in S-phase as the first label was present (until up to 4 h after the injection) and finished their S-phase as the second labeled was administered. Hence, the group of NSCs defined by *gfap*:GFP<sup>+</sup> in the zebrafish telencephalon undergoes collectively and comparatively long S-phases.

The emergence of complex patterns from simple rules has been analyzed extensively, e.g., for artificial systems [53], biology-inspired models [54], and biological phenomena [44]. Here, we first use an interaction model to determine the strength and the radius of the observed pattern of S-phase NSCs. This extends traditional Gibbs or Cox models [38] to the analysis of a particular multi-type spatial point pattern where NSCs in S-phase appear at discrete locations defined by the presence of all NSCs. However, fitting an interaction model to spatiotemporal data does not necessarily imply that cell–cell interactions are present. We thus model quiescent and NSCs in S-phase as agents in a continuous space with stochastic cell cycle kinetics to quantitatively compare 2 hypotheses: Are extrinsic effects like cell–cell interactions or signaling waves required to generate the observed aggregated spatiotemporal division patterns, or do redivisions of NSCs suffice? We find that agent-based simulations with redividing NSCs suffice to explain the spatiotemporal patterns observed in the zebrafish brain.

According to this model, coordination of cell cycle entries in the population can be explained by the internal synchronization due to the cells' history. Such a model is supported by interesting studies following cell families in cultured cell lines, detecting correlations of cell cycle parameters between siblings and cousin cells as a result of inherited factors but independent on their location [55,56]. This cell intrinsic–driven behavior is in contrast to cellular systems endowed with a clear environmental regulation, in particular in stem cell compartments constituted by a distinctive 3D architecture like in the hair follicle in the skin. There, the precise location of stem cells is associated with their cycling behavior and fate [57], indicating that the position within an environment impacts on the stem cells activity. In the neurogenic zone studied here, no distinctive organization of the tissue would so far hint toward morphological features specifically associated with an NSC's activity. Our analysis of cellular volumes in this study does not indicate that a distinctive density dependent niche would be in place. In this

context, it is also interesting to note that 2 intermingled types of progenitors with radial morphology in the killifish telencephalon reveal distinct proliferative behaviors [58] even if sharing the same environment, arguing for a small contribution of environmental effects on stem cell activities.

Nevertheless, we cannot reject the existence of cell extrinsic mechanisms that might contribute to the spatiotemporal patterns, such as local diffusive signaling activity in delimited groups of cells, a functional activity of extended cell–cell contacts (as has been observed in NSCs by Obermann and colleagues [20]) or the activity of the Notch signaling pathway [7]. As a result of either mechanisms, levels of molecular heterogeneity have been observed: For instance, variable levels of the Zinc finger protein *Fezf2* regulate Notch activity levels and quiescence of NSCs [25]. Likewise, the expression of miR-9 involved in keeping quiescence upstream of Notch signaling is found only in a subset of quiescent cells [14] several days after a division. And in a recent study, a subpopulation of NSCs that expresses low levels of *Elavl3* has been characterized as mostly nondividing cells in transit toward neuronal differentiation [59].

Repeated divisions of NSCs were not expected, given the low percentage of divisions in the whole NSC population. NSCs were able to enter successive rounds of cell division within a few days, the earliest starting 24 h after the previous S-phase. This contrasts with the early model of adult NSCs established in the mouse dentate gyrus and subependymal zone, where radial astrocytes are quiescent and can replenish the transient amplifying progenitor population after all fast dividing cells have been eradicated by an ARA-C treatment [60]. The latest characterizations by single cell sequencing differentiate between quiescent versus active radial astrocytes; however, these most probably represent alternating states, as also suggested by a continuum of transcriptional states [61,62]. The definition of an active radial astrocytes here cannot distinguish between a repetitive or sporadic division behavior. Recently, however, evidence for several rounds of divisions within a few days in NSCs has been reported in the mouse subependymal zone from clonal analysis data using the Troy-driven recombination [63]. Further, live imaging of *Ascl*-driven recombination in the dentate gyrus *in vivo* demonstrated the existence of several divisions of radial astrocytes in a short time window [64]. In zebrafish, 2 studies based on *in vivo* imaging of the whole brain have followed NSCs, 1 observing single-labeled NSCs and their behavior of division and differentiation [28], the other considering the whole dividing NSC population with the help of 2 transgenic lines, *her4:RFP* and *mcm5:GFP*, highlighting the NSCs and the cell cycle, respectively [34], over the course of 2 weeks. Rapidly redividing daughter NSCs, as observed by our double-labeling approach in living fish combined with confocal imaging that allows for a clear distinction between nuclei close to each other, have not been reported yet. It is likely that they have been missed due to the difficulty to resolve distinct sister cells with intravital imaging in previous studies.

How these reoccurring divisions of adult stem cells come about will be important to assess in the future. Single cell sequencing data, such as performed by Cosacak and colleagues [65], might help associating these events with specific molecular pathways. Studies on human epithelial cell lines have shown that mother cells can relay distinct levels of *CCND1* mRNA and p53 protein to their daughters, which after completing mitosis will then rapidly decide for the next round of division, depending on a resulting bimodal level of activity of cyclin-dependent kinase (CDK2) [66,67]. It could well be that in NSCs too, specific cell cycle regulators are transmitted to daughter cells, thereby permitting for immediate new rounds of division. Other mechanisms of daughter cells' decisions for quiescence or new cycle have been reported, for instance, in the adult mouse dentate gyrus, where degradation of *Ascl1a* in daughter cells mediated by the ubiquitin ligase *Huwe1* takes place and promotes a reentry in quiescence [15].

Hence, in stem cell populations, individual phases of quiescence and exit thereof might well be predictable according to inherited factors, allowing us to understand how the kinetics of

tissue maintenance are regulated. Beyond this study, which aimed at understanding a regulation within days-scale time window, long-term tracing studies will help understanding the consequences of those patterns on the organization of the resulting neuronal circuits.

## Methods

### Zebrafish maintenance and transgenic lines

Zebrafish of the transgenic line *gfap:GFP* (mi2001) [30] were bred and maintained in the fish facility of the Helmholtz Zentrum München. Experiments were conducted under the animal protocol 55.2-1-2532-83-14, in accordance with animal welfare rules of the government of Oberbayern.

### Labeling of consecutive S-phases in vivo and immunostainings

BrdU or EdU were dissolved at a concentration of 1 mg/ml in saline solution (0.07% NaCl) containing methylene blue and injected intraperitoneally (5  $\mu$ l/ 0.1 g body mass) into the fish at precise time points. Fish were over-anesthetized in MS-222 or placed in ice water for euthanasia, decapitated, and the brains dissected and fixed in 4% PFA overnight. Whole mount brains were processed for Click chemistry with Azyde-Alexa-Fluor-647 to detect EdU, following the manufacturer's protocol (C-10269, Thermo Fisher Scientific, Darmstadt, Germany); see also [68]. For the subsequent BrdU immunoreaction, brains were treated with HCl 2M at 37°C for 30 min, washed in sodium tetraborate buffer 0.1 M, pH8 and in PBS, and incubated with Mouse-anti BrdU (Phoenix-Flow Systems, San Diego, USA, PRB1-U) 1:800 overnight. Mouse or Rabbit anti-PCNA (clone PC-10, Santa Cruz, Santa Cruz, USA, sc-56; Abcam ab15497) was diluted 1:800 or 1:100, respectively. Following secondary antibodies were incubated for 1 h at room temperature at 1:1,000 concentration: Goat-anti-mouse Alexa-Fluor-555, Goat-anti-mouse Alexa-Fluor-405, Goat-anti-Rabbit-Pacific Blue, and Goat-anti-Rabbit Alexa-Fluor-555 (Thermo Fisher Scientific).

### Imaging of whole mount brains

Brains were mounted in vectashield (Vector Laboratories, Burlingame, USA) or in PBS with 40% glycerol between 2 coverslips separated by 8-layered parafilm spacers. Imaging was performed using a Leica SP5 confocal microscope (Wetzlar, Germany) with a 20 $\times$  glycerol immersion objective at a resolution of 2048  $\times$  2048 pixels, and for close up views with a 63 $\times$  glycerol immersion objective with a resolution of 1024  $\times$  1024 pixels.

### Identification of *gfap:GFP*+ cells and PCNA/EdU/BrdU-labeled cells

NSCs that are labeled by GFP in the transgenic *gfap:GFP* line were identified using the Single Cell Identification Pipeline (SCIP; [29]). In this pipeline, single cells are automatically identified from an image 3D stack, exploiting the fact that all NSCs are located on top of the hemisphere on a 2D surface. SCIP returns x, y, and z coordinates for all NSCs. Nuclei labeled by PCNA immunocytochemistry, BrdU immunocytochemistry, or by EdU-click chemistry were identified semiautomatically: Nuclei were first identified with SCIP and then visually verified 1 by 1 using Fiji to avoid false positives and ensure correct assignment to GFP-positive or negative cells, using the 2 channels in consecutive z-planes of the confocal stacks. Images are displayed as maximum intensity projections of z stacks, single planes, or 3D visualization using the Plugins Clear Volume [69] or 3D viewer [70].

In experiments with  $\Delta t > 18$  h, the majority of Time 1-labeled cells were found as doublets, i.e., a pair of small cells close to each other (see Fig 3I and 3M and S3D–S3J Fig), since the time

span allowed the mother cell to reach its mitotic (M) phase. For  $\Delta t = 72$  h, even triplets occurred (see Fig 3Q). Doublets and triplets were handled as single-division events in the subsequent analysis. Some of the doublets incorporated the second dU label and were categorized only to Time 1 in the spatiotemporal analysis, in order to assess distances specifically between the distinct sets of cells undergoing S-phases at consecutive time points. The proportion of redividing NSCs was calculated by dividing the number of doublets (or triplets) containing at least 1 GFP+BrdU+EdU+ cell by the total number of GFP+BrdU+ doublets (or triplets).

### Three-dimensional segmentation of *gfap*:GFP+ cells and PCNA/EdU/BrdU-labeled cells

Four-channel confocal images of 4 whole mount hemispheres taken at the 63 $\times$  objective (1024  $\times$  1024 pixels) were used. Pixel classification of each single channel was performed in Ilastik [71] to discriminate cells from cell borders and background. The Ilastik autocontext pixel classification was used for BrdU and PCNA channels containing background staining of blood vessels. The Fiji MorphoLibJ package [72] was then used on the pixel-classified GFP channel to perform either a 3D distance-transformed watershed segmentation or a marker transformed watershed segmentation using a white top hat filtered image as marker input, depending on the image quality and the precision of the resulting labeled image. Single objects in the segmented images were evaluated and quantified using the Fiji 3D Manager [73], and 3D reconstructions were obtained with the 3D Viewer [70]. Small objects with a volume below 30  $\mu\text{m}^3$ , corresponding to segmented radial processes detached from segmented cell bodies, were filtered out from the analysis, giving a total number of 834, 961, 293, and 283 NSCs, respectively. Co-localization of segmented NSCs with the division markers was performed by quantification in the 3D Manager on binarized images and verified manually. The analysis of volumes of dividing, redividing, and quiescent NSCs, as well as their direct neighbors in the 4 hemispheres was performed in R using the spatstat package [38].

### Spatial statistics

Spatial statistics were applied on NSCs labeled in S-phase, i.e., *gfap*:GFP+ cells with EdU or BrdU staining. The set of NSCs without EdU or BrdU staining served as the substrate to, e.g., evaluate patterns for randomly dividing NSCs. Spatial analysis is performed with MATLAB (MATLAB version 9.7.0.1296695 (R2019b) Update 4).

### Three-dimensional distance matrix

We determine the distance between any 2 cells by calculating the shortest path on the hemisphere manifold to account for the bending of the hemisphere surface. To this end, we used the fitted hemisphere surface and calculated the stepwise shortest paths between 2 cell locations on it (see [29] for a detailed description).

### Ripley's K statistics

Ripley's K [31] is a measure for the deviation of a point pattern from spatial homogeneity. It has been previously applied in geographic information science [74], in the context of spatial economic analysis [75], archeological studies [76], single-molecule localization [32], or single cells behavior [33]. In all those applications, it is assumed that the point pattern occurs from a Poisson point process on a homogenous space. Here, however, division events can only occur where NSCs are located on the 2D hemisphere manifold. As the underlying NSC distribution could be inhomogeneous, we adapted the measure to account for this possible inhomogeneity

by sampling a point pattern only from discrete NSC locations and thus call it discrete Ripley's K.

For S S-phase NSCs we calculate the spatial  $K_S$  value for increasing radii  $r$  along

$$K_S(r) = \left(\frac{S}{A}\right)^{-1} \sum_{i=1}^S w(i, r) \sum_{j=1, j \neq i}^S \frac{I(\text{dist}(i, j) \leq r)}{S} \quad (1)$$

This function counts NSCs in S-phase within a radius  $r$  around the S-phase NSC  $i$ . The indicator function  $I$  is 1 if cell  $i$  and  $j$  are closer than  $r$  and 0 otherwise. The term is normalized by the total amount of NSCs in S-phase  $S$  and by the S-phase NSC density  $S/A$ . The hemisphere area  $A$  is calculated as the sum of all triangle areas between all NSCs obtained via Delaunay triangulation (see S1A Fig) [77]. The edge correction term  $w(i, r)$  is 1 if the disc with radius  $r$  around NSC  $i$  does not cut the hemisphere edge and else calculated as the fraction of disc inside the hemisphere.

To obtain the spatiotemporal  $K_{ST}$  between 2 sets of NSCs in S-phase, labeled at Time 1 and Time 2 we modified Eq 1:

$$K_{ST}(r) = \left(\frac{S_1}{A}\right)^{-1} \sum_{i=1}^{S_1} w(i, r) \sum_{j=1}^{S_2} \frac{I(\text{dist}(i, j) \leq r)}{S_2} \quad (2)$$

Here, we count  $S_2$  (NSCs in S-phase at Time 2) cells around  $S_1$  (NSCs in S-phase at Time 1) cells within  $r$ . In contrast to Eq 1 the whole term is normalized by the density of  $S_1$  cells while the indicator function is divided by the number of  $S_2$  cells.

To compare observed K values to random spatial distributions, we sample the amount of observed  $S$  and  $S_2$  cells, respectively, ( $S_1$  cell locations are fixed) from all NSCs 20 times and calculate the random sampling discrete K value. To evaluate whether the observed K value differs from random sampling, we check if the observed K value is below the 5% quantile (which would indicate spatial dispersion) or above the 95% quantile (which would indicate spatial aggregation) of the 20 sampled discrete K values.

To make Ripley's K plots comparable between hemispheres, we standardized the results per hemisphere similar to a z-score, such that K values are 1 when they are on the 95% quantile and  $-1$  when they are on the 5% quantile.

To classify S-phase patterns as aggregated, random, or dispersed, we calculated the mean z-score between 30 and 150  $\mu\text{m}$ . If the mean z-score is above 1 we classify the pattern as aggregated, between  $-1$  and 1 as random and below  $-1$  as dispersed.

## Model-based analysis

### Interaction model

To determine the spatial extent and the nature of the temporal interaction of divisions at different time points, we extend a spatial interaction model [29] with 2 parameters: the interaction strength ( $g$ ), where  $g = 1$  means no dependencies, below 1 dispersion between the 2 populations, and above 1 aggregation. The second parameter the model fits is the interaction radius ( $r$ ) in  $\mu\text{m}$ .

$$\log(L(g, r)) = \sum_{i=1}^{S_1} \log\left(g \sum_{j=1}^{S_2} I(\text{dist}(i, j) \leq r)\right) / \sum_{i=1}^{S_1} \log\left(g \sum_{k=1}^N I(\text{dist}(i, k) \leq r)\right) \quad (3)$$

For each hemisphere, we fit the parameters ( $g, r$ ) of the model to locations of cells in S-phase of 2 time points to detect spatial dependencies of NSCs in S-phase at the later time point ( $S_2$ ) regarding NSCs in S-phase at the earlier time point ( $S_1$ ). The parameters are fitted with a log-

likelihood approach (Eq 3), where  $S_x$  is the number of cells in S-phase at time point  $x$ , while  $N$  is the whole NSC population. The indicator function  $I$  is 1 if the distance between 2 cells ( $i, j$  or  $i, k$ ) is smaller or equal  $r$ . In the numerator term, we iterate over all  $S_1$  cells and count the number of  $S_2$  cells having a smaller or equal distance than  $r$ , while in the denominator, we normalize the equation by counting the not affected (non-S-phase) cells of  $N$  within  $r$  around  $S_1$  cells.

Eq 3 calculates the log likelihood for 1 hemisphere. To calculate the log likelihood across several hemispheres, every single likelihood per hemisphere is summed up and optimized in parallel to form a combined likelihood. We use the PESTO toolbox [78] to maximize the likelihood including uncertainty estimation via posterior sampling [79] as can be seen exemplarily in Fig 2B for hemispheres with  $\Delta t = 32$  h and for all hemispheres in S2C Fig.

### Cell division model

The observed fractions of redividing NSCs and DLS (Fig 3F and 3T) suggest an upper limit for the S-phase length of 32 h (since we observe no DLSs for  $\Delta t \geq 32$  h, Fig 3F) and a lower limit for all other cell cycle phases of 9 h (since we observe no redividing cells at  $\Delta t = 9$  h, Fig 3T). A simple model with a fixed cell cycle length of, e.g., 32 h + 9 h, is however not able to explain a plateauing of redividing NSCs for  $\Delta t \geq 24$  h. Instead of constant cell cycle and S-phase length, we thus assume them to be distributed as delayed exponential distributions:

$$f(x; \beta, d) = \begin{cases} 0 & x < d, \\ \frac{1}{\beta} e^{-\frac{1}{\beta}(x-d)} & x \geq d \end{cases}$$

where  $\beta$  is the scale parameter of the exponential distribution and  $d$  is the delay. Note that [40] use a different notation for the 2 parameters. Additionally, we assume a redivision probability  $p_{rediv}$  (S4A Fig), which is lower bounded by the observed redivision frequency of 15%, since we are not able to observe all redivisions (S4B Fig) in our snapshot experiments with an EdU labeling of 1 h.

To infer these 5 parameters (scale parameter and delay of cell cycle,  $d_{cc}$  and  $\beta_{cc}$ , and S-phase,  $d_{sp}$  and  $\beta_{sp}$ , and redivision probability  $p_{rediv}$ ) from (i) the observed redivision frequency; and (ii) the fraction of DLS NSCs (see sketches in S4C and S4D Fig) at every labeling interval  $\Delta t = 9$  h, 18 h, 24 h, 32 h, 48 h, and 72 h, we apply ABC [41]. Given the 5 parameters ( $d_{cc}$ ,  $\beta_{cc}$ ,  $d_{sp}$ ,  $\beta_{sp}$ , and  $p_{rediv}$ ), we simulate the dynamics of 10,000 dividing cells over 400 to 500 h. Using a fixed endpoint (Time 2), we can determine Time 1 by subtracting the according labeling interval. With 2 virtual measurements we calculate the observed redivision and DLS frequency via counting all cells being in S-phase at Time 1 and determine the status of these Time 1 S-phase cells at Time 2 (see S4B Fig): If an S-phase of a cell is labeled by both virtual measurements, it is classified as DLS, and if an offspring cell is labeled in S-phase, the cell is classified as redividing cell. The simulated redivision frequency is then the number of redividing cells divided by the number of S-phase cells at Time 1, and the DLS fraction is the number of DLSs divided by the number of S-phase cells at Time 1, accordingly.

In order to fit the observed data, we define a distance function between the observed and simulated redivision and DLS fractions per labeling interval. To this end, we calculate the sum of differences between means and SDs, respectively:

$$distance_{ABC} = \sum_i (|\text{mean}(obs_i) - \text{mean}(sim_i)| + |\text{std}(obs_i) - \text{std}(sim_i)|)$$

where  $i \in \{9 \text{ h}, 18 \text{ h}, 24 \text{ h}, 32 \text{ h}, 48 \text{ h}, \text{ and } 72 \text{ h}\}$ . We optimize this distance function employing ABC with 50 epochs, evaluating 500 particles per epoch (ABC parameters).

### Agent-based spatiotemporal simulation

To assess the contribution of redivision events to the emerging spatial patterns of NSCs in S-phase, we simulated an NSC population using a cellular Potts model as implemented in Morpheus [45]. Morpheus is a modeling and simulation environment where cells act and interact as agents in space and time. We approximate a base division rate  $p_{div}$  from the average observed S-phase cell fraction ( $1.9 \pm 0.7\%$ ,  $n = 36$  hemispheres, mean  $\pm$  SD) and S-phase length estimation (approximately 18 h):  $p_{div} = 0.019 / 18 \text{ h} = 1 \times 10^{-3}$  divisions per hour. This base division rate suffices for control simulations, but for simulations with redivisions, we decreased  $p_{div}$  slightly. To obtain a similar amount of observed divisions as in the control simulations, we fixed  $p_{div} = 9 \times 10^{-4}$  divisions per hour to account for redivisions, which divide after 1 cell cycle independently of the base division rate with probability  $p_{rediv} = 0.38$  (inferred from the cell division model above).

At the beginning of the simulation, every cell has to wait at least 1 cell cycle duration until it is available for a spontaneous division via  $p_{div}$ . This leads to a small bias in the beginning (roughly 1/30th of the whole simulation time), but after all cells are older than their cell cycle time, there should be no impact anymore. We estimate the rate of NSC differentiation from the proportion of doublets with 1 *gfap*:GFP+ and 1 *gfap*:GFP- cell. This proportion is roughly 10% in our data. *Gfap*:GFP- cells are simulated for 1 more cell cycle duration and are then excluded, mimicking differentiated cells transitioning away from the stem cell pool inside the brain [27]. Cell cycle length is inferred from the cell division model (see above) with delay  $d_{cc}$  and  $\beta_{cc}$  from the delayed exponential distribution. Other predefined parameters are minimum and maximum cell size, determined by measuring real NSC size and simulated via sigmoidal cell growth. Morpheus implements basic cell-cell interactions and kinetic assumptions [45]. Our simulation starts with 500 cells and runs until the colony size reaches the observed  $2,356 \pm 460$  (mean  $\pm$  SD) cells (see Fig 4A–4E). We analyze cells in S-phase at the simulated measurement time points and apply the same spatial statistics as for the real data.

### Figures

Plots were prepared in Matlab, R, and Excel. Figures were adjusted in Adobe Indesign and Inkscape.

### Supporting information

**S1 Fig.** (A) Delaunay triangulation on the identified NSCs to calculate discrete Ripley's K. (B) A synthetically generated random pattern of NSCs in S-phase. (C) Discrete Ripley's K correctly identifies the pattern in (B) to be within the 90% CI of randomly sampled patterns. (D) A synthetically generated dispersed pattern with an interaction radius of 100  $\mu\text{m}$ . (E) Discrete Ripley's K correctly identifies the pattern in (D) below the random interval. (F) Standardized K shows S-phase NSC aggregation in 24 (thick lines) out of 36 hemispheres. Hemispheres that show a mean standardized K value higher than 1 between 30 and 150  $\mu\text{m}$  are classified as aggregated (thick lines). Each trace represents the spatial distribution of S-phase NSCs in 1 hemisphere. (G) All hemispheres of the dataset. Gray dots represent NSCs magenta dots represent NSCs in S-phase. Aggregation patterns are classified according to the standardized K analysis in (F). (H) Cells in S-phase are equally labeled by EdU or BrdU when those are administered concomitantly. BrdU, 5-bromo-2'-deoxyuridine; CI, confidence interval; EdU,

5-Ethynyl-2'-deoxyuridine; NSC, neural stem cell.  
(TIF)

**S2 Fig.** (A) Experimental setup for measuring spatiotemporal patterns. With different labeling intervals  $\Delta t$  between the 2 S-phase labelings, we systematically profile spatiotemporal effects. (B) Standardized K shows spatiotemporal aggregation of S-phase NSCs in 22 out of 36 hemispheres for different  $\Delta t$ , while the remaining 14 patterns are classified as random. (C) Parameter inference on all labeling intervals  $\Delta t$ . Most likely interaction strength and radius for all hemispheres per labeling interval determined via posterior sampling density. Whiskers cover the 95% CIs for strength and radius. CI, confidence interval; NSC, neural stem cell.  
(TIF)

**S3 Fig.** (A) Example of reoccurring divisions taking place in *gfap:GFP*<sup>-</sup> progenitors (yellow arrow). (B) Experimental setup for observing NSCs in S-phase at 3 time points within 2 days. (C) Reoccurring divisions have been quantified in 6 hemispheres out of 4 brains (% of BrdU<sup>+</sup>-clones labeled with a second cell cycle label) and are observed both in NSCs (*gfap:GFP*<sup>+</sup>, green dots) and in *gfap:GFP*<sup>-</sup> progenitors (gray dots) after 1 day (BrdU+EdU<sup>+</sup>) or after 2 days (BrdU+PCNA<sup>+</sup>). Higher percentages of redivisions are observed compared to the experiments with a short pulse of EdU labeling at time 2 (injection 1 h before killing, see Fig 3), probably due to more S-phases entries labeled with a longer duration of EdU availability in this experimental setup. (D-J) Maximum intensity projection and close-up view of a 4-channel image depicting recurrent divisions marked with 3 cell cycle labelings at 3 time points: 2 days earlier by BrdU (E), 1 day earlier by EdU (F), and at the time of fixation by PCNA (G). (H) *gfap:GFP*<sup>+</sup> daughter cell pairs (green arrows) and *gfap:GFP*<sup>-</sup> daughter cell pairs (white arrows) with respectively 1 daughter entering a second round of division. (I,J) Overlays. Scale bar: 10  $\mu$ m. (K) Quantification of reoccurring divisions in NSCs performed on 63 $\times$  magnification images from 4 hemispheres. The average proportions of dividing NSCs labeled at distinct time points reveal that about 40% of all PCNA<sup>+</sup> NSCs have been generated by a recent division (compare gray quadrants PCNA+EdU<sup>+</sup> and BrdU+EdU<sup>+</sup> to the yellow PCNA-only quadrant) (L) 3D segmentation and 3D visualization of all NSCs performed on the same image as shown in (D–I), representing a total of 834 NSCs. (M–P) Close-up view of the boxed area delineated in D and L depicting BrdU-labeled sister cells, 1 of which is labeled with PCNA (yellow arrows). (M) Maximum intensity projection. (N) Single z-plane. All NSCs are delimited by white lines obtained by a 3D watershed segmentation applied after Ilastik pixel classification of the original GFP channel. 3D segmentation of the BrdU and PCNA channels are represented in cyan and yellow, respectively. (O) Labeled NSCs visualized in 3D (top view). (P) 3D reconstruction of the BrdU-labeled sister cells, 1 of which is PCNA-labeled. (Q) Volumes of all segmented NSCs from the hemisphere shown in L. No significant difference was observed between the volumes of NSCs found in reoccurring division (PCNA+EdU<sup>+</sup> and PCNA+BrdU<sup>+</sup>), division (PCNA only) or quiescent state (negative for all cell cycle labels) according to Kruskal–Wallis test ( $p$ -value = 0.10). No significant difference can be detected either between the mean volume of the direct neighbors of those groups, arguing against the formation of distinctive niche areas within the NSC population ( $p$ -value = 0.37). BrdU, 5-bromo-2'-deoxyuridine; EdU, 5-Ethynyl-2'-deoxyuridine; GFP, green fluorescent protein; NSC, neural stem cell.  
(TIF)

**S4 Fig.** (A) We devise a cell division model to describe dividing NSCs with 5 parameters: the minimal cell cycle length  $d_{cc}$  and minimal S-phase length  $d_{sp}$  and their variability  $\beta_{cc}$  and  $\beta_{sp}$  respectively, parametrizing a lag-exponential distribution and the redivision probability  $p_{rediv}$ . Note that we do not have any information about the length of G2 and thus do neither model

nor visualize this phase in the graphs. (B) We use observed redivision proportions (see Fig 3T) and the percentage of DLS cells (see Fig 3F) to feed a division model (S-phases labeled in red) and take simulated measurements at T1 and T2 to infer the redivision probability  $p_{re-div}$ . We observe a difference between the model ( $p_{re-div} = 0.38$ ) and observed redivisions frequency (0.15): From 10 simulated cells dividing at T1, 4 cells (cell 1, 3, 5, and 10, shown in bold) redivide. Thus, 40% cell redivide, which is close to  $p_{re-div}$ . However, our measurement at T2 only picks up 2 cells (cells 3 and 5) corresponding to only 20%, in line with the observed redivision frequency. (C, D) The cell division model fits the proportion of redividing cells (C) and DLS cells (D) for different labeling intervals. (E) We find an extremely variable delayed exponential distributed cell cycle with a minimal length ( $d_{cc}$ ) of 22.2 h and a mean ( $d_{cc} + \beta_{cc}$ ) of 107.5 h. (F) The S-phase length is narrow with 16.6-h minimal length ( $d_{sp}$ ) and  $\beta_{sp} = 1.5$ h. (G) Snapshot of an agent-based simulation using the Morpheus software corresponding to Fig 4A (Time point 2) with NSCs (green), NSCs in S-phase (red), and dividing intermediate progenitor cells (black). (H) Boxed area in (G) shown in higher magnification to visualize simulated cell shapes. (I, J) Proportion of observed redividing and DLS cells for different labeling intervals in agent-based simulations vs. experimental data. (K) Cell cycle length measured in the agent-based simulations compared to the probability density function of the estimated delayed exponential function in (E). (L) Fraction of redividing cells measured in 20 agent-based simulations. DLS, double-labeled S-phase; NSC, neural stem cell. (TIF)

**S1 Table. For each brain hemisphere (Experiment) with a particular labeling interval, we detail the number of NSCs, number of S-phases at labeling time 1 (T1) and labeling time 2 (T2), number of NSCs in S-phase at T1 and T2, number of redivisions, number of NSC redivisions, number of DLS, and number of DLS in NSCs.** DLS, double-labeled S-phase; NSC, neural stem cell. (XLSX)

**S1 Data. Data tables to support S3C and S3K Fig.** (XLSX)

## Acknowledgments

We thank Carolin Loos and Lisa Bast for their advice regarding parameter inference, Yannik Schälte for his support with ABC, Benjamin Ballnus for his insights in MCMC sampling, Elmar Spiegel and Hannah Busen for statistical consulting, and the rest of the Marr lab for helpful comments. Also, we thank Walter de Back for his help to set up Morpheus and Jovica Ninkovic for his valuable remarks. Furthermore, we thank Moritz Thomas for their notes on the manuscript and Elly Pachel for her help on code/analysis reproducibility. The experimental work was carried out in the lab of Hernán Lopez-Schier.

## Author Contributions

**Conceptualization:** Carsten Marr, Prisca Chapouton.

**Data curation:** Prisca Chapouton.

**Formal analysis:** Valerio Lupperger.

**Funding acquisition:** Carsten Marr.

**Methodology:** Carsten Marr, Prisca Chapouton.

**Software:** Valerio Lupperger.

**Supervision:** Carsten Marr, Prisca Chapouton.

**Validation:** Valerio Lupperger.

**Visualization:** Valerio Lupperger.

**Writing – original draft:** Valerio Lupperger, Carsten Marr, Prisca Chapouton.

**Writing – review & editing:** Valerio Lupperger, Carsten Marr, Prisca Chapouton.

## References

1. van Velthoven CTJ, Rando TA. Stem Cell Quiescence: Dynamism, Restraint, and Cellular Idling. *Cell Stem Cell*. 2019; 24:213–25. <https://doi.org/10.1016/j.stem.2019.01.001> PMID: 30735649
2. Gonzales KAU, Fuchs E. Skin and Its Regenerative Powers: An Alliance between Stem Cells and Their Niche. *Dev Cell*. 2017; 43:387–401. <https://doi.org/10.1016/j.devcel.2017.10.001> PMID: 29161590
3. Yeh C-Y, Asrican B, Moss J, Quintanilla LJ, He T, Mao X, et al. Mossy Cells Control Adult Neural Stem Cell Quiescence and Maintenance through a Dynamic Balance between Direct and Indirect Pathways. *Neuron*. 2018;493–510.e4. <https://doi.org/10.1016/j.neuron.2018.07.010> PMID: 30057205
4. Dong J, Pan Y-B, Wu X-R, He L-N, Liu X-D, Feng D-F, et al. A neuronal molecular switch through cell-cell contact that regulates quiescent neural stem cells. *Sci Adv*. 2019; 5:eaa4416. <https://doi.org/10.1126/sciadv.aav4416> PMID: 30820459
5. Ottone C, Krusche B, Whitby A, Clements M, Quadrato G, Pitulescu ME, et al. Direct cell–cell contact with the vascular niche maintains quiescent neural stem cells. *Nat Cell Biol*. 2014;1045–56. <https://doi.org/10.1038/ncb3045> PMID: 25283993
6. Lepko T, Pusch M, Müller T, Schulte D, Ehses J, Kiebler M, et al. Choroid plexus-derived miR-204 regulates the number of quiescent neural stem cells in the adult brain. *EMBO J*. 2019; 38:e100481. <https://doi.org/10.15252/emboj.2018100481> PMID: 31304985
7. Chapouton P, Skupien P, Hesl B, Coolen M, Moore JC, Madelaine R, et al. Notch activity levels control the balance between quiescence and recruitment of adult neural stem cells. *J Neurosci*. 2010; 30:7961–74. <https://doi.org/10.1523/JNEUROSCI.6170-09.2010> PMID: 20534844
8. Ehm O, Göritz C, Covic M, Schäffner I, Schwarz TJ, Karaca E, et al. RBPJkappa-dependent signaling is essential for long-term maintenance of neural stem cells in the adult hippocampus. *J Neurosci*. 2010; 30:13794–807. <https://doi.org/10.1523/JNEUROSCI.1567-10.2010> PMID: 20943920
9. Alunni A, Krecsmarik M, Bosco A, Galant S, Pan L, Moens CB, et al. Notch3 signaling gates cell cycle entry and limits neural stem cell amplification in the adult pallium. *Development*. 2013; 140:3335–47. <https://doi.org/10.1242/dev.095018> PMID: 23863484
10. Liang J, Balachandra S, Ngo S, O'Brien LE. Feedback regulation of steady-state epithelial turnover and organ size. *Nature*. 2017; 548:588–91. <https://doi.org/10.1038/nature23678> PMID: 28847000
11. Mesa KR, Kawaguchi K, Cockburn K, Gonzalez D, Boucher J, Xin T, et al. Homeostatic Epidermal Stem Cell Self-Renewal Is Driven by Local Differentiation. *Cell Stem Cell*. 2018; 23:677–686.e4. <https://doi.org/10.1016/j.stem.2018.09.005> PMID: 30269903
12. Knobloch M, Braun SMG, Zurkirchen L, von Schoultz C, Zamboni N, Araúzo-Bravo MJ, et al. Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis. *Nature*. 2013; 493:226–30. <https://doi.org/10.1038/nature11689> PMID: 23201681
13. Knobloch M, Pilz G-A, Ghesquière B, Kovacs WJ, Wegleiter T, Moore DL, et al. A Fatty Acid Oxidation-Dependent Metabolic Shift Regulates Adult Neural Stem Cell Activity. *Cell Rep*. 2017; 20:2144–55. <https://doi.org/10.1016/j.celrep.2017.08.029> PMID: 28854364
14. Katz S, Cussigh D, Urbán N, Blomfield I, Guillemot F, Bally-Cuif L, et al. A Nuclear Role for miR-9 and Argonaute Proteins in Balancing Quiescent and Activated Neural Stem Cell States. *Cell Rep*. 2016; 17:1383–98. <https://doi.org/10.1016/j.celrep.2016.09.088> PMID: 27783951
15. Urbán N, van den Berg DLC, Forget A, Andersen J, Demmers JAA, Hunt C, et al. Return to quiescence of mouse neural stem cells by degradation of a proactivation protein. *Science*. 2016; 353:292–5. <https://doi.org/10.1126/science.aaf4802> PMID: 27418510
16. Mueller T, Wullimann MF. An evolutionary interpretation of teleostean forebrain anatomy. *Brain Behav Evol*. 2009; 74:30–42. <https://doi.org/10.1159/000229011> PMID: 19729894
17. Figueira M, Bayley P, Navratilova P, Becker TS, Wilson SW, Clarke JDW. Morphogenesis underlying the development of the everted teleost telencephalon. *Neural Dev*. 2012; 7:32. <https://doi.org/10.1186/1749-8104-7-32> PMID: 22989074

18. Furlan G, Cuccioli V, Vuillemin N, Dirian L, Muntasell AJ, Coolen M, et al. Life-Long Neurogenic Activity of Individual Neural Stem Cells and Continuous Growth Establish an Outside-In Architecture in the Teleost Pallium. *Curr Biol*. 2017; 27:3288–3301.e3. <https://doi.org/10.1016/j.cub.2017.09.052> PMID: 29107546
19. Edelmann K, Glashauser L, Sprungala S, Hesi B, Fritschle M, Ninkovic J, et al. Increased radial glia quiescence, decreased reactivation upon injury and unaltered neuroblast behavior underlie decreased neurogenesis in the aging zebrafish telencephalon. *J Comp Neurol*. 2013; 521:3099–115. <https://doi.org/10.1002/cne.23347> PMID: 23787922
20. Obermann J, Wagner F, Kocijaj A, Zambusi A, Ninkovic J, Hauck SM, et al. The Surface Proteome of Adult Neural Stem Cells in Zebrafish Unveils Long-Range Cell-Cell Connections and Age-Related Changes in Responsiveness to IGF. *Stem Cell Reports*. 2019; 12:258–73. <https://doi.org/10.1016/j.stemcr.2018.12.005> PMID: 30639211
21. Pellegrini E, Mouriec K, Anglade I, Menuet A, Le Page Y, Gueguen M-M, et al. Identification of aromatase-positive radial glial cells as progenitor cells in the ventricular layer of the forebrain in zebrafish. *J Comp Neurol*. 2007; 501:150–67. <https://doi.org/10.1002/cne.21222> PMID: 17206614
22. Chapouton P, Webb KJ, Stigloher C, Alunni A, Adolf B, Hesi B, et al. Expression of hairy/enhancer of split genes in neural progenitors and neurogenesis domains of the adult zebrafish brain. *J Comp Neurol*. 2011; 519:1748–69. <https://doi.org/10.1002/cne.22599> PMID: 21452233
23. März M, Chapouton P, Diotel N, Vaillant C, Hesi B, Takamiya M, et al. Heterogeneity in progenitor cell subtypes in the ventricular zone of the zebrafish adult telencephalon. *Glia*. 2010; 58:870–88. <https://doi.org/10.1002/glia.20971> PMID: 20155821
24. Than-Trong E, Bally-Cuif L. Radial glia and neural progenitors in the adult zebrafish central nervous system. *Glia*. 2015; 63:1406–28. <https://doi.org/10.1002/glia.22856> PMID: 25976648
25. Berberoglu MA, Dong Z, Li G, Zheng J, Trejo Martinez L, del CG, et al. Heterogeneously expressed *fezf2* patterns gradient Notch activity in balancing the quiescence, proliferation, and differentiation of adult neural stem cells. *J Neurosci*. 2014; 34:13911–23. <https://doi.org/10.1523/JNEUROSCI.1976-14.2014> PMID: 25319688
26. Kroehne V, Freudenreich D, Hans S, Kaslin J, Brand M. Regeneration of the adult zebrafish brain from neurogenic radial glia-type progenitors. *Development*. 2011; 138:4831–41. <https://doi.org/10.1242/dev.072587> PMID: 22007133
27. Rothenaigner I, Krecsmarik M, Hayes JA, Bahn B, Lepier A, Fortin G, et al. Clonal analysis by distinct viral vectors identifies bona fide neural stem cells in the adult zebrafish telencephalon and characterizes their division properties and fate. *Development*. 2011; 138:1459–69. <https://doi.org/10.1242/dev.058156> PMID: 21367818
28. Barbosa JS, Sanchez-Gonzalez R, Di Giaimo R, Baumgart EV, Theis FJ, Götz M, et al. Neurodevelopment. Live imaging of adult neural stem cell behavior in the intact and injured zebrafish brain. *Science*. 2015; 348:789–93. <https://doi.org/10.1126/science.aaa2729> PMID: 25977550
29. Lupperger V, Buggenthin F, Chapouton P, Marr C. Image analysis of neural stem cell division patterns in the zebrafish brain. *Cytometry A*. 2018; 93:314–22. <https://doi.org/10.1002/cyto.a.23260> PMID: 29125897
30. Bernardos RL, Raymond PA. GFAP transgenic zebrafish. *Gene Expr Patterns*. 2006; 1007–13. <https://doi.org/10.1016/j.modgep.2006.04.006> PMID: 16765104
31. Ripley BD. The second-order analysis of stationary point processes. *J Appl Probab*. 1976; 13:255–66.
32. Khater IM, Nabi IR, Hamarneh G. A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods. *Patterns*. 2020; 100038. <https://doi.org/10.1016/j.patter.2020.100038> PMID: 33205106
33. Loerke D, le Duc Q, Blonk I, Kerstens A, Spanjaard E, Machacek M, et al. Quantitative imaging of epithelial cell scattering identifies specific inhibitors of cell motility and cell-cell dissociation. *Sci Signal*. 2012; 5:rs5. <https://doi.org/10.1126/scisignal.2002677> PMID: 22763340
34. Dray N, Bedu S, Vuillemin N, Alunni A, Coolen M, Krecsmarik M, et al. Large-scale live imaging of adult neural stem cells in their endogenous niche. *Development*. 2015; 142:3592–600. <https://doi.org/10.1242/dev.123018> PMID: 26395477
35. Corral-Rivas JJ, Wehenkel C, Castellanos-Bocaz HA, Vargas-Larreta B, Diéguez-Aranda U. A permutation test of spatial randomness: application to nearest neighbour indices in forest stands. *J For Res*. 2010; 218–25. <https://doi.org/10.1007/s10310-010-0181-1>
36. Tanimura S, Shima M. Quantitative measurements of inequality in geographic accessibility to pediatric care in Oita Prefecture, Japan: Standardization with complete spatial randomness. *BMC Health Serv Res*. 2011. <https://doi.org/10.1186/1472-6963-11-163> PMID: 21736715
37. Nicholson A. Analysis of spatial distributions of accidents. *Saf Sci*. 1998; 71–91. [https://doi.org/10.1016/S0925-7535\(98\)00056-3](https://doi.org/10.1016/S0925-7535(98)00056-3)

38. Baddeley A, Rubak E, Turner R. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press; 2015.
39. Jagiella N, Rickert D, Theis FJ, Hasenauer J. Parallelization and High-Performance Computing Enables Automated Statistical Inference of Multi-scale Models. *Cell Syst*. 2017; 4:194–206.e9. <https://doi.org/10.1016/j.cels.2016.12.002> PMID: 28089542
40. Weber TS, Jaehnert I, Schichor C, Or-Guil M, Carneiro J. Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLoS Comput Biol*. 2014; 10:e1003616. <https://doi.org/10.1371/journal.pcbi.1003616> PMID: 25058870
41. Klinger E, Rickert D, Hasenauer J. pyABC: distributed, likelihood-free inference. *Bioinformatics*. 2018; 34:3591–3. <https://doi.org/10.1093/bioinformatics/bty361> PMID: 29762723
42. Graner F, Glazier JA. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys Rev Lett*. 1992; 69:2013–6. <https://doi.org/10.1103/PhysRevLett.69.2013> PMID: 10046374
43. Alt S, Ganguly P, Salbreux G. Vertex models: from cell mechanics to tissue morphogenesis. *Philos Trans R Soc Lond B Biol Sci*. 2017:20150520. <https://doi.org/10.1098/rstb.2015.0520> PMID: 28348254
44. Manukyan L, Montandon SA, Fofonjka A, Smirnov S, Milinkovitch MC. A living mesoscopic cellular automaton made of skin scales. *Nature*. 2017; 544:173–9. <https://doi.org/10.1038/nature22031> PMID: 28406206
45. Starrauß J, de Back W, Brusch L, Deutsch A. Morpheus: a user-friendly modeling environment for multi-scale and multicellular systems biology. *Bioinformatics*. 2014; 30:1331–2. <https://doi.org/10.1093/bioinformatics/btt772> PMID: 24443380
46. Calegari F, Haubensak W, Haffner C, Huttner WB. Selective lengthening of the cell cycle in the neurogenic subpopulation of neural progenitor cells during mouse brain development. *J Neurosci*. 2005; 25:6533–8. <https://doi.org/10.1523/JNEUROSCI.0778-05.2005> PMID: 16014714
47. Takahashi T, Nowakowski RS, Caviness VS Jr. The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall. *J Neurosci*. 1995; 15:6046–57. <https://doi.org/10.1523/JNEUROSCI.15-09-06046.1995> PMID: 7666188
48. Duronio RJ. Developing S-phase control. *Genes Dev*. 2012; 26:746–50. <https://doi.org/10.1101/gad.191171.112> PMID: 22508722
49. Nordman J, Orr-Weaver TL. Regulation of DNA replication during development. *Development*. 2012; 139:455–64. <https://doi.org/10.1242/dev.061838> PMID: 22223677
50. Arai Y, Pulvers JN, Haffner C, Schilling B, Nüsslein I, Calegari F, et al. Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat Commun*. 2011; 2:154. <https://doi.org/10.1038/ncomms1155> PMID: 21224845
51. Turrero García M, Chang Y, Arai Y, Huttner WB. S-phase duration is the main target of cell cycle regulation in neural progenitors of developing ferret neocortex. *J Comp Neurol*. 2016; 524:456–70. <https://doi.org/10.1002/cne.23801> PMID: 25963823
52. Ponti G, Obernier K, Alvarez-Buylla A. Lineage progression from stem cells to new neurons in the adult brain ventricular-subventricular zone. *Cell Cycle*. 2013; 12:1649–50. <https://doi.org/10.4161/cc.24984> PMID: 23673324
53. Marr C, Hütt MT. Topology regulates pattern formation capacity of binary cellular automata on graphs. *Physica A*. 2005; 354:641–62.
54. Kauffman S, Clayton P. On emergence, agency, and organization. *Biol Philos*. 2006; 21:501–21.
55. Sandler O, Mizrahi SP, Weiss N, Agam O, Simon I, Balaban NQ. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature*. 2015; 519:468–71. <https://doi.org/10.1038/nature14318> PMID: 25762143
56. Mura M, Feillet C, Bertolusso R, Delaunay F, Kimmel M. Mathematical modelling reveals unexpected inheritance and variability patterns of cell cycle parameters in mammalian cells. *PLoS Comput Biol*. 2019; 15:e1007054. <https://doi.org/10.1371/journal.pcbi.1007054> PMID: 31158226
57. Rompolas P, Mesa KR, Greco V. Spatial organization within a niche as a determinant of stem-cell fate. *Nature*. 2013; 502:513–8. <https://doi.org/10.1038/nature12602> PMID: 24097351
58. Coolen M, Labusch M, Mannioui A, Bally-Cuif L. Mosaic Heterochrony in Neural Progenitors Sustains Accelerated Brain Growth and Neurogenesis in the Juvenile Killifish *N. furzeri*. *Curr Biol*. 2020; 30:736–745.e4. <https://doi.org/10.1016/j.cub.2019.12.046> PMID: 32004451
59. Lange C, Rost F, Machate A, Reinhardt S, Lesche M, Weber A, et al. Single cell sequencing of radial glia progeny reveals the diversity of newborn neurons in the adult zebrafish brain. *Development*. 2020; 147. <https://doi.org/10.1242/dev.185595> PMID: 31908317

60. Doetsch F, Caillé I, Lim DA, García-Verdugo JM, Alvarez-Buylla A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell*. 1999; 97:703–16. [https://doi.org/10.1016/s0092-8674\(00\)80783-7](https://doi.org/10.1016/s0092-8674(00)80783-7) PMID: 10380923
61. Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba A. Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell*. 2015; 17:329–40. <https://doi.org/10.1016/j.stem.2015.07.002> PMID: 26235341
62. Dulken BW, Leeman DS, Boutet SC, Hebestreit K, Brunet A. Single-Cell Transcriptomic Analysis Defines Heterogeneity and Transcriptional Dynamics in the Adult Neural Stem Cell Lineage. *Cell Rep*. 2017; 18:777–90. <https://doi.org/10.1016/j.celrep.2016.12.060> PMID: 28099854
63. Basak O, Krieger TG, Muraro MJ, Wiebrands K, Stange DE, Frias-Aldeguer J, et al. Troy+ brain stem cells cycle through quiescence and regulate their number by sensing niche occupancy. *Proc Natl Acad Sci U S A*. 2018; 115:E610–9. <https://doi.org/10.1073/pnas.1715911114> PMID: 29311336
64. Pilz G-A, Bottes S, Betizeau M, Jörg DJ, Carta S, Simons BD, et al. Live imaging of neurogenesis in the adult mouse hippocampus. *Science*. 2018; 359:658–62. <https://doi.org/10.1126/science.aao5056> PMID: 29439238
65. Cosacak MI, Bhattarai P, Reinhardt S, Petzold A, Dahl A, Zhang Y, et al. Single-Cell Transcriptomics Analyses of Neural Stem Cell Heterogeneity and Contextual Plasticity in a Zebrafish Brain Model of Amyloid Toxicity. *Cell Rep*. 2019; 27:1307–1318.e3. <https://doi.org/10.1016/j.celrep.2019.03.090> PMID: 31018142
66. Spencer SL, Cappell SD, Tsai F-C, Overton KW, Wang CL, Meyer T. The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell*. 2013; 155:369–83. <https://doi.org/10.1016/j.cell.2013.08.062> PMID: 24075009
67. Yang HW, Chung M, Kudo T, Meyer T. Competing memories of mitogen and p53 signalling control cell-cycle entry. *Nature*. 2017; 549:404–8. <https://doi.org/10.1038/nature23880> PMID: 28869970
68. Qu D, Wang G, Wang Z, Zhou L, Chi W, Cong S, et al. 5-Ethynyl-2'-deoxycytidine as a new agent for DNA labeling: detection of proliferating cells. *Anal Biochem*. 2011; 417:112–21. <https://doi.org/10.1016/j.ab.2011.05.037> PMID: 21683679
69. Royer LA, Weigert M, Günther U, Maghelli N, Jug F, Sbalzarini IF, et al. ClearVolume: open-source live 3D visualization for light-sheet microscopy. *Nat Methods*. 2015; 12:480–1. <https://doi.org/10.1038/nmeth.3372> PMID: 26020498
70. Schmid B, Schindelin J, Cardona A, Longair M, Heisenberg M. A high-level 3D visualization API for Java and ImageJ. *BMC Bioinformatics*. 2010; 11:274. <https://doi.org/10.1186/1471-2105-11-274> PMID: 20492697
71. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, et al. ilastik: interactive machine learning for (bio)image analysis. *Nat Methods*. 2019; 16:1226–32. <https://doi.org/10.1038/s41592-019-0582-9> PMID: 31570887
72. Legland D, Arganda-Carreras I, Andrey P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics*. 2016; 32:3532–4. <https://doi.org/10.1093/bioinformatics/btw413> PMID: 27412086
73. Ollion J, Cochenne J, Loll F, Escudé C, Boudier T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics*. 2013; 29:1840–1. <https://doi.org/10.1093/bioinformatics/btt276> PMID: 23681123
74. Hohl A, Zheng M, Tang W, Delmelle E, Casas I. Spatiotemporal Point Pattern Analysis Using Ripley's K Function. In *Geospatial Data Science Techniques and Applications*. 2017. p. 155–76. <https://doi.org/10.1201/b22052-7>
75. Arbia G, Espa G, Giuliani D, Dickson MM. Effects of missing data and locational errors on spatial concentration measures based on Ripley's K-function. *Spat Econ Anal*. 2017; 326–46. <https://doi.org/10.1080/17421772.2017.1297479>
76. Negre J, Muñoz F, Barceló JA. A Cost-Based Ripley's K Function to Assess Social Strategies in Settlement Patterning. *J Archaeol Method Theory*. 2018; 777–94. <https://doi.org/10.1007/s10816-017-9358-7>
77. Others Delaunay B. Sur la sphere vide. *Izv Akad Nauk SSSR. Otdelenie Matematicheskii i Estestvennyka Nauk*. 1934; 7:1–2.
78. Stapor P, Weindl D, Ballnus B, Hug S, Loos C, Fiedler A, et al. PESTO: Parameter ESTimation TOolbox. *Bioinformatics*. 2018; 34:705–7. <https://doi.org/10.1093/bioinformatics/btx676> PMID: 29069312
79. Ballnus B, Schaper S, Theis FJ, Hasenauer J. Bayesian parameter estimation for biochemical reaction networks using region-based adaptive parallel tempering. *Bioinformatics*. 2018; 34:i494–501. <https://doi.org/10.1093/bioinformatics/bty229> PMID: 29949983