iScience

Article

Identifying stem cell numbers and functional heterogeneities during postembryonic organ growth

Diana-Patricia Danciu, Julian Stolper, Lázaro Centanin, Anna Marciniak-Czochra

anna.marciniak@iwr.
uni-heidelberg.de (A.M.-C.)
dpdanciu@math.
uni-heidelberg.de (D.-P.D.)

Highlights

Markov process model of stem cell dynamics during postembryonic organ growth

Numerical and analytical approaches to fit the model to experimental clonal data

The model identifies the number of stem cells that participate in organ growth

Mathematical modeling uncovers a novel heterogeneous behavior of growth stem cells

Article

Identifying stem cell numbers and functional heterogeneities during postembryonic organ growth

Diana-Patricia Danciu,[1,2,*] Julian Stolper,[3,4] Lázaro Centanin,[3] and Anna Marciniak-Czochra[1,2,5,*]

SUMMARY

Uncovering the number of stem cells necessary for organ growth has been challenging in vertebrate systems. Here, we developed a mathematical model characterizing stem cells in the fish gill, an organ displaying non-exhaustive growth. We employ a Markov model, stochastically simulated via an adapted Gillespie algorithm, and further improved through probability theory. The stochastic algorithm produces a simulated dataset for comparison with experimental clonal data by inspecting quantifiable properties. The analytical approach skips the step of artificial data generation and goes directly to the quantification, being more abstract and efficient. We report that a reduced number of stem cells actively contribute to growing and maintaining the gills. The model also highlights a functional heterogeneity among the stem cells involved, where activation and quiescence phases determine their relative growth contribution. Overall, our work presents a method for inferring the number and properties of stem cells required in a lifelong growing system.

INTRODUCTION

Stem cells are essential during organ growth in all higher vertebrates. Once the organisms reach maturation and acquire a definitive body size, their adult stem cells (aSCs) are responsible for maintaining homeostasis, i.e., they continue to proliferate with the goal of replacing the cells lost regularly. Unlike mammals and other higher vertebrates, fish increase their body size throughout their entire life. Accordingly, fish organs must adapt to this permanent growth, by either increasing in size as happens with gills (Stolper et al., 2019) and retina (Centanin et al., 2014; Tsingos et al., 2019), or in numbers as is the case with neuromasts - mechano-sensory organs sensing the water flow (Ghysen and Dambly-Chaudière, 2007; Dambly-Chaudière et al., 2003; Wada et al., 2013; Seleit et al., 2017). Fish aSCs carry the task of organ remodeling, and they are not only able to maintain homeostasis, but also to drive growth. In this work, we focus on the respiratory organ of fish - the gill or the branchia, recently introduced as a suitable model for studying aSCs and organ development (Stolper et al., 2019). The model organism under study is the Japanese rice fish (*Oryzias latipes*), colloquially known as medaka, which is convenient because of its rapid development (Figure 1B) and isogenic genome (Wittbrodt et al., 2002). Two different stem cell populations have been reported (Stolper et al., 2019) to drive growth and maintain homeostasis, with the growth stem cells being restricted to the growing edge of the organs.

Each gill is composed of four pairs of branchial arches, which in turn comprise a sequence of filaments, each of which is built from multiple stacked lamellae (Leguen, 2018) (see Figure 1A). Branchial arches grow along two orthogonal axes: longitudinally by elongating their filaments and transversally by adding new ones (Figure 1C). Both these actions are performed by stem cells located at the growing edges of the respective domains: growth stem cells at the periphery of branchial arches (*br-archSCs*) generate new filaments, which contain growth stem cells at the tip (*filamSCs*), driving their elongation by adding new lamellae. Homeostatic stem cells are present along the mid-axis of filaments, at the base of each lamella, and are responsible for replenishing the cell population supporting the tissue. Development of branchial arches presents a hierarchical setting: A *br-archSC* divides so that its daughter cell creates the new filament, where it becomes a *filamSC* that drives its elongation from the tip by leaving a trail of homeostatic stem cells along the way, which have the role of maintaining the lamellae (Figure 1D).

We have recently shown (Stolper et al., 2019) that the growth *br-archSCs* include four different fate restricted stem cells which are recruited together into a newly forming filament, giving rise to four different

[1]Institute of Applied Mathematics, Heidelberg University, 69120 Heidelberg, Baden-Württemberg, Germany

[2]Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, 69120 Heidelberg, Baden-Württemberg, Germany

[3]Centre for Organismal Studies (COS), Heidelberg University, 69120 Heidelberg, Baden-Württemberg, Germany

[4]Murdoch Children's Research Institute, University of Melbourne, 3052 Parkville, VIC, Australia

[5]Lead contact

*Correspondence:
anna.marciniak@iwr.uni-heidelberg.de (A.M.-C.),
dpdanciu@math.uni-heidelberg.de (D.-P.D.)
https://doi.org/10.1016/j.isci.2022.103819
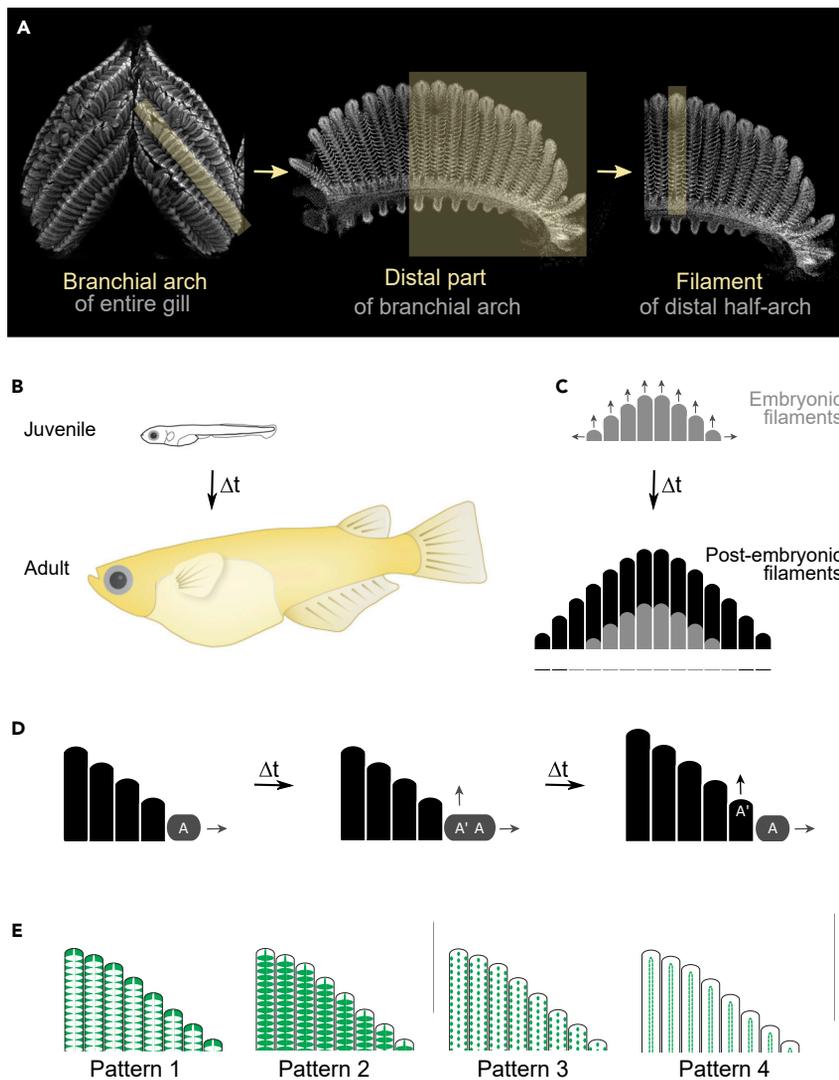
Figure 1. Biological background of medaka gills

(A) Structure of a gill: entire gill (left) contains four pairs of branchial arches (middle), each of which is composed of a sequence of filaments (right). Our study focuses on half-arches (distal and proximal).

(B) Development of medaka fish.

(C) The gills grow along two axes, by elongating filaments and by adding new ones. New filaments are generated by stem cells residing in niches at the basal extremities of branchial arches, whereas their elongation is driven by stem cells at the tip of each filament.

(D) Simplified scheme for the postembryonic growth of branchial arches: when a stem cell A in the niche divides (left), its progeny A' generates a new filament (middle) and drives its elongation from the tip (right). This process continues throughout the fish's life.

(E) Four different cell lineages have been reported (Stolper et al., 2019) which, when labeled, give rise to four different filamental patterns. These patterns are recorded in the dataset arrays as entries 1-4 (see Results and Figure 2A).

patterns when labeled (Figure 1E). This means that the different stem cell types coordinate their activity and division to generate a new filament. Such a concept opens many questions and avenues to explore: How do the stem cells coordinate their behavior to work as an ensemble? How do the fate restricted cells get recruited to the new filament and regulate their division so that they maintain the ratio of different cell types within the filament? A first step in approaching such questions is to get an estimation for the number of *br-archSCs* of each fate, which are at the base of the hierarchy, being essential for setting in motion the organ development mechanism.

Accordingly, this work develops mathematical tools for counting the stem cells choosing each fate, responsible for generating filaments. Defining the number of stem cells involved in a lifelong process has proven difficult in most systems, mainly because of the lack of specific markers. Stem cell counting has been possible in intestinal crypts by the existence of a specific marker (Snippert et al., 2010) and it has been inferred in the hematopoiesis system by using genetic tricks that allow a combinatorial label (Pei et al., 2017; Busch et al., 2015; Sun et al., 2014) or by modeling the hematopoietic system as a modified Moran process with Wright-Fisher drift and by using an approximate Bayesian computation framework for parameter estimation (Lee-Six et al., 2018). In medaka fish gills, stem cell specific markers are not available. Therefore, in the absence of specific markers, lineage tracing experiments are performed with a ubiquitous permanent labeling, and data are only available at one time point because the fish needs to be euthanized. The essential property of fish gills, which facilitates our analysis despite data restrictions, is their modular structure. Therefore, as the permanent label is inherited by progeny upon division, the experimental images resulting from lineage tracing experiments can provide a history of *br-archSCs* divisions.

RESULTS

Stochastic model suggests that gill stem cells are not homogeneous in their division behavior

In our previous paper (Stolper et al., 2019), we have studied the nature of the filament-generating cells, by investigating two different hypotheses to determine their self-renewal capacity. Are these stem cells (SCs), capable of creating multiple filaments, or are these cells a group of progenitors, each of which is capable of producing exactly one filament? We tested each of the two hypotheses, based on clonal data acquired using lineage tracing tools (Gaudi toolkit) described by Centanin et al. (2014) and recording the labeling status of each filament along a branchial arch. The experimental procedure permanently labels a small number of cells at embryonic stages with a nuclear-tagged Green Fluorescent Protein (GFP). Briefly, the Gaudi^{RSG} medaka line expresses a Red Fluorescent Protein (RFP) ubiquitously, and this RFP prevents the expression of a nuclear GFP (nGFP). Upon induction with Tamoxifen or after a heat shock, the RFP is removed from the genome allowing nGFP expression. Because this constitutes a modification in the genome of the cell, the same modification will be found in all its progeny, which will be recognised by the nGFP expression.

In the gill system, if a *br-archSC* is labeled with nGFP, then the filaments originating from it will be green as well. In addition, along the same lines, filaments coming from an unlabeled cell will not be green. Accordingly, each branchial arch is described by an array of binary values corresponding to labeled (1) and unlabeled (0) filaments (Figure 2A and supplemental Table S1). Because the focus is on postembryonic filaments, for each branchial arch, we selected the eight most peripheral filaments from each half of the arch (from now on referred to as "mini-arches") and rearranged them so that in the resulting table the first column consists of the oldest filament of the mini-arch, i.e., the eighth filament counted from the *br-archSCs* niche (Figure 2A and Tables S2 and S3). The choice of selecting only eight filaments is based on the fact that only branchial arches with more than 25 filaments are considered, and taking into account that normally there exist five to eight embryonic filaments in the middle of the arch, which were generated before induction of genetic recombination. As presented in our previous paper (Stolper et al., 2019), there are at least four different fate-restricted *br-archSCs* in the peripheral niche (Figure 1E). These give rise to different patterns of nGFP that are easily distinguishable in the filaments and will be hereafter referred to as Patterns 1-4. Therefore, our dataset consists of arrays containing values in the set {0, 1, 2, 3, 4} corresponding to unlabeled filaments or the four different possible patterns, respectively, or combinations of patterns recorded by values of the form $\overline{ab}$, $\overline{abc}$, or $\overline{abcd}$ with $a, b, c, d \in \{1, 2, 3, 4\}$. Combinations of more than two patterns are not found in our experimental data - observing such cases is very difficult with the use of a single-color label, because certain patterns (e.g., Patterns 3 or 4) can be easily hidden under more prominent patterns (Patterns 1 or 2). Therefore, the only combinations of patterns observed in our data are $\overline{ab} \in \{12, 13, 23, 34\}$. A quantification of the occurrence of simple and mixed patterns can be seen in supplemental Figure S1. Each pattern is analyzed separately, so the original data are remodeled into four datasets, one for each pattern, with binary values (Figure 2A and Tables S2 and S3).

For the comparison between models and data, the notion of switches was introduced, defined as the transition from a labeled to an unlabeled filament, i.e., binary transitions (1-to-0 or 0-to-1). Accordingly, for each mini-arch we counted the switches and the labeled filaments, and recorded them as a pair of the form (s, f) with $s \in \{0, 1, ..., 7\}$ switches and $f \in \{0, 1, ..., 8\}$ labeled filaments (Figures 2A and 2B). For a mini-arch of eight filaments, out of the total number of 8×9 associations (s, f) only 33 are possible (for example, for $s = 0$ the only options are $f \in \{0, 8\}$, whereas for $s = 7$ one must have $f = 4$) - supplemental table S4. The frequency of observing each such pair (s, f) in the data can be computed and is shown in Figure 2C.
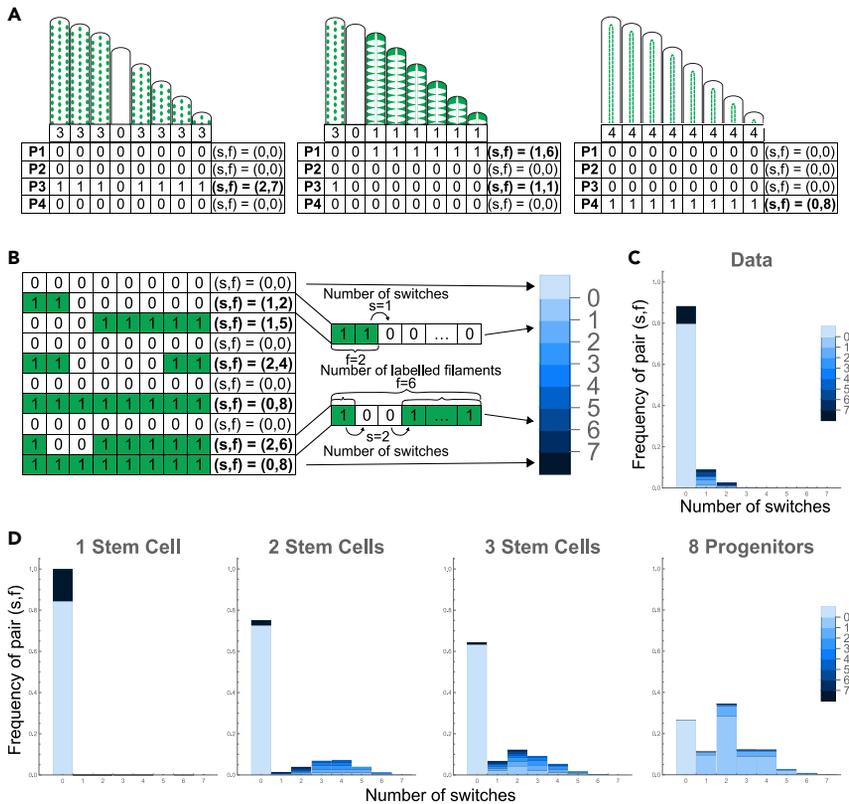
Figure 2. Approach and results of the homogeneous model

(A) Toy mini-arches used to describe how the data is prepared for the analysis. For each mini-arch, the distribution of patterns is recorded as 1-4 (first row of each table), whereas mixed patterns are recorded as $\overline{ab}$, $\overline{abc}$, or $\overline{abcd}$ with $a, b, c, d \in \{1, 2, 3, 4\}$ (Figure S1 and Table S1). For each pattern, the configuration is rewritten such that each mini-arch is described by an array of binary entries (bottom four rows of the tables, showing the distribution for each pattern P1-P4). Finally, for each such array, the number of switches s (1-to-0 and 0-to-1 transitions) and that of labeled filaments f is determined, and recorded as a pair.

(B) Example of 10 mini-arches extracted from the experimental data, adjusted for studying Pattern 1. For each mini-arch the pair (s, f) of switches and labeled filaments are recorded and shown in the last column. Focusing on two mini-arches, the switches are indicated and the labeled filaments are counted, the number of which corresponds to a blue tone on the color bar shown, which will be used in the following plots.

(C) Plot of the 8-filament long mini-arches in the entire dataset, for Pattern 1. The plot shows the frequency of observing a certain pair (s, f) of switches and labeled filaments (y axis), according to the number of switches (s) observed on the x axis, and the number of labeled filaments (f) represented by the color code. One can notice that most mini-arches are composed of filaments carrying the same label/pattern (0 switches), and out of these, most are fully unlabeled. This ratio between entirely labeled and entirely unlabeled mini-arches is regulated by the experimental labeling efficiency, which is approximated from the data. For pattern 1, the labeling efficiency is mathematically estimated to be 15.53% (i.e., cells have a probability 0.1553 of being labeled) - see STAR Methods for the description of the approximation method and CIs.

(D) Comparison between various simulated scenarios. The two extreme scenarios described previously (Stolper et al., 2019), the stem cell model (left) and the progenitor model (right), are plotted as the data in C, for comparison. The stem cell model, even though extreme, provides a much better fit to the data from C, than the progenitor model does. The central plots show results of simulations for scenarios in which 2 (middle-left) and 3 (middle-right) stem cells reside in the niche, for Pattern 1. Here a homogeneous division behavior is assumed, i.e., stem cells are randomly chosen for division and filament generation, with all having the same probability of being selected. In such a homogeneous setting, the more stem cells are introduced into the system, the worse the fit to the data becomes, and the progression can be followed from left to right.

We have recently shown (Stolper et al., 2019) that a stem cell model considering all filaments in the mini-arch to be generated by one single stem cell provides a better fit to the data than a progenitor model in which each filament is created by a different progenitor cell (Figure 2D, far left and far right, reconsiders the two scenarios, for 8-filament long mini-arches, for Pattern 1). Because we often observe switches in the branchial arches, consequently more than one stem cell are contributing filaments to the branchial arch. In addition, interesting is the

often occurrence of long stretches of consecutive filaments carrying the same label, indicating that if more than one stem cell contributes filaments, they are not homogeneous in their division behavior. This insight is supported by the comparison of the experimental data obtained from clonal analysis to the two modeled cases where two and three functionally homogeneous stem cells contribute filaments to the branchial arch (middle plots in Figure 2D). The simulations are performed by first selecting the number of labeled stem cells, $L \sim$ Binomial(n,$probLab$), where n is the total number of stem cells in the niche (in the two cases considered in Figure 2D middle, $n = 2$ or $n = 3$) and $probLab$ is the experimental labeling efficiency approximated from the data (e.g., $probLab = 0.1553$ for Pattern 1) via a combinatorial approach (see STAR Methods). Next, a weighted random choice selects whether a labeled or an unlabeled stem cell will divide, at each time step, with weights given by the number of labeled and unlabeled br-$archSCs$, for each of the eight filaments in the arch. The labeling efficiency is different for each pattern and represents an average over the whole dataset. The approximation method only considers the oldest filament in each mini-arch, as it provides an indication of the first cell which divided postembryonically. The labeling efficiency can differ significantly across the stem cell types, because of the experimental method of labeling and the current state of the cells of interest at the time of induction. The heat-shock stimulus is applied by adding water at a specific temperature, which means that not all cells in the body will receive the same change in temperature (ΔT), depending on their position within the niche. In addition, in the case of Tamoxifen-induced combination, the drug is delivered by the circulatory system and cells located closer to blood vessels would be more exposed to the inducing agent. For example, Pattern 1 is observed by the labeling of pavement or respiratory cells, which occupy more than 90% of the filament surface and are found in the epithelium of filaments and lamellae (Laurent, 1984; Laurent and Dunel, 1980; Leguen, 2018; Olson, 2002). At the other extreme, Pattern 4 is observed as elongated nuclei located inside the filament tissue between the epithelium and the skeletal member, along the efferent edge, which are thought to be neuroepithelial cells (Laurent, 1984; Laurent and Dunel, 1980). Such variability in population size and location among the stem cell types may lead to large differences in the labeling efficiency for the four patterns. Following this initial study, we concluded that the filament-generating cells are indeed stem cells (br-$archSCs$), but simply having a small number of functionally homogeneous br-$archSCs$ contributing filaments to the branchial arches is not sufficient for explaining the data.

A Markov approach shows that fish gills originate from a small number of functionally heterogeneous stem cells

As opposed to a homogeneous scenario in which br-$archSCs$ (both labeled and unlabeled) divide randomly giving rise to an uncorrelated succession of labels, the often occurrence of stretches of consecutive identically-labeled filaments observed in the data and the previously tested cases (of two and three homogeneous SCs - Figure 2D, middle) suggest successive generations (divisions) of the same br-$archSC$. This heterogeneity idea for the gill system is derived and quantified from the data via the mathematical study, as a purely experimental approach could not assess such a behavior. The heterogeneous scenario means that not all br-$archSCs$ in the niche have equal probabilities of being selected for division at a specific time step, and corresponds to the concept of stem cell activation and quiescence phases, in which when a stem cell divides, it becomes activated and divides multiple times before another SC takes over. The heterogeneity is incorporated in the model through the "probability of division" parameter p, representing the probability that the stem cell which has just divided will be the next one to divide again. If this probability $p = 1$, then all filaments in the mini-arch will carry the label of the first stem cell which divided, whereas a probability of division $p = 0.5$ corresponds to the most homogeneous scenario with intermingled labeled and unlabeled filaments with comparable incidence. This correlation between the cell which has just divided and the one about to divide, for $p > 0.5$, will be referred to as a "heterogeneous division behavior" or a "functional heterogeneity". As previously mentioned, this behavior is suggested by the occurrence of long stretches of consecutive filaments with the same label, but quantifying this heterogeneity parameter in practice, experimentally, is not possible, thus leading to the need for mathematical modeling.

Accordingly, because of the assumption of single-step memory of the system that takes into account which cell divided last, the heterogeneity scenario can be modeled by a Markov process on the generated filaments (Figures 3A and 3B). If a labeled filament was generated in the previous time step, we expect a higher probability that the next filament produced will also be labeled. The transitions within the two-state Markov Chain, between a labeled and an unlabeled filament, are described by the conditional probabilities shown in matrix P (1), from Box 1. These conditional probabilities depend on the number of labeled (L) and unlabeled (U) stem cells in the niche, on the total number $n = L + U$ of stem cells in the niche, and on the probability of division p (see STAR Methods for derivation). The model assumptions are summarized in Box 1.
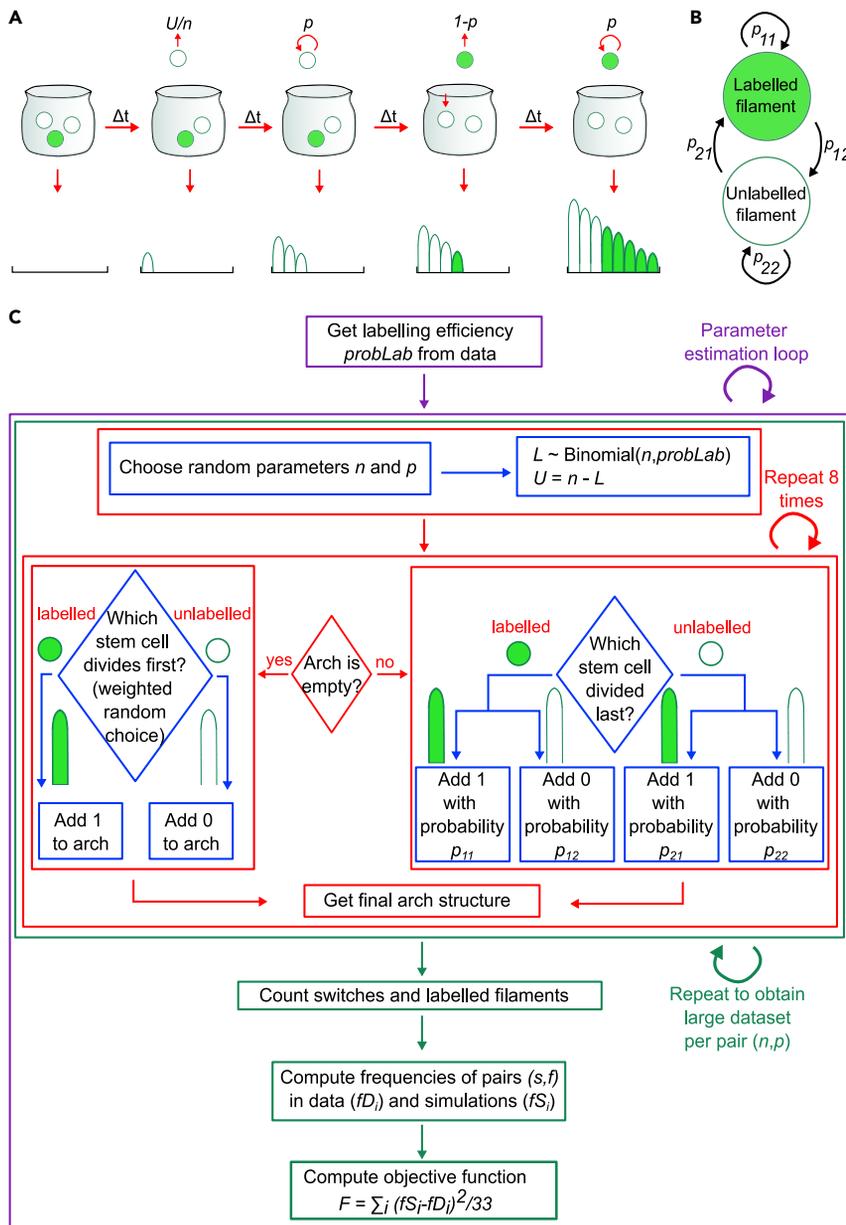
Figure 3. Approach for the heterogeneous model

(A) Schematic of the model, according to the assumptions in Box 1. In the initial time step, a stem cell (green and white circles) is randomly chosen for division. The probability of this cell to divide again is p, and if $p>0.5$ more consecutive divisions of the same cell take place, filling the mini-arches with filaments (green and white oval shapes). At a later point, a new cell may be selected, with a probability $1 - p$, and the probability of this cell to divide again is changed to p. This process continues until eight filaments are generated.

(B) Two-state Markov Chain diagram, describing the transitions between filaments of labels 1 and 0 within the mini-arch, via the probabilities $p_{11}, \ldots, p_{22}$. These probabilities are computed in (1).

(C) Flowchart of the stochastic algorithm presented in the main text and explained in Box 2. The green and white circles represent labeled and unlabeled cells, whereas the green and white oblong shapes are labeled and unlabeled filaments, respectively.

To simulate the previously deduced heterogeneity hypothesis, a stochastic algorithm was developed, which generates a dataset of *in silico* mini-arches, to compare to those from the experimental data. The two are compared by inspecting the frequencies of observing each of the pairs (s, f) of switches and labeled filaments. In accordance with the assumptions presented in Box 1 and based on the conditional

---

Box 1. Model assumptions

The model presented below is based on the following assumptions:

◇ One labeled stem cell (1) in the peripheral niche produces a labeled filament, whereas one unlabeled stem cell (0) gives rise to an unlabeled filament.

◇ The number of labeled stem cells in the niche, L, depends on the labeling efficiency $probLab$ and on the total number of stem cells in the niche n: $L \sim \text{Binomial}(n, probLab)$, and the number of unlabeled stem cells is $U = n - L$.

◇ The cell which has just divided will divide again in the next time step with probability p, where the time steps correspond to division events and have random durations.

The probabilities for the transitions from one filament to another, with varying labeling status, for $n \geq 2$ and $p \in [0.5, 1)$, are found in the matrix P, with entries p_{ij} corresponding to having a filament of label j following a filament of label i in the mini-arch, with $i, j \in \{1, 2\}$, 1-labeled and 2-unlabeled.

$$P = \begin{pmatrix} p + (1-p)\dfrac{L-1}{n-1} & (1-p)\dfrac{U}{n-1} \\ (1-p)\dfrac{L}{n-1} & p + (1-p)\dfrac{U-1}{n-1} \end{pmatrix}$$ (Equation 1)

---

probabilities in (Equation 1), the algorithm shown in the flowchart from Figure 3C and outlined in Box 2 was implemented.

The algorithm fills an initially empty array with eight binary entries, corresponding to the filaments in a mini-arch, by taking into account the previously added value. A large dataset of simulated mini-arches is produced. The algorithm is repeated inside a parameter estimation loop, so it is run multiple times with different starting guesses for an array of values of parameters $n \in \{1, \ldots, 10\}$, $p \in [0.5, 1)$ for which the objective function (2) is computed to be minimized. The objective function is a mean of square errors (MSE) between the frequencies in the experimental and simulated data, of observing a pair $(s, f) \in \Pi$, with $\Pi = \{(0, 0), (0, 8), \ldots, (7, 4)\}$ the set of 33 possible pairs (see STAR Methods and Table S4 for details). Finally, the best parameters n and p are obtained, which provide the most accurate fit between the simulation results and the data, performed by minimizing the objective function. For this purpose, the *Mathematica* routine NMinimize was used, employing the Nelder-Mead and Differential Evolution methods.

Furthermore, an analytical approach can be developed, to skip the step of *in silico* data generation, such that instead of computing the approximate frequencies of observing a pair (s, f) in the respective simulated data, exact probabilities of each such event can be directly calculated. By using the entries of the probability transition matrix (1) together with the initial distribution $\lambda = (\lambda_1, \lambda_2) = \left(\frac{L}{n}, \frac{U}{n}\right)$, the analytical probabilities of observing a certain pair (s, f) of switches and labeled filaments in the model, for each parameter pair (n, p), are computed as a sum of probabilities of possible Markov Chain trajectories producing the required number of switches and labeled filaments. For example, the probabilities of producing an entirely

Table 1. Summary of parameter notations

Parameter	Domain	Description
L	$\mathbb{N}$	Number of *labeled* stem cells in the niche
U	$\mathbb{N}$	Number of *unlabeled* stem cells in the niche
$n = L + U$	$\mathbb{N}$	Total number of stem cells in the niche
$probLab$	$(0, 1)$	Labeling efficiency
p	$[0.5, 1)$	Probability of division
s	$\{0, 1, \ldots, 7\}$	Number of switches within a 8-filament long mini-arch
f	$\{0, 1, \ldots, 8\}$	Number of labeled filaments within a mini-arch

---

Box 2. A stochastic algorithm for heterogeneous stem cell division

1. The algorithm starts by approximating the labeling efficiency based on the data, as explained by Stolper et al. (2019) and in STAR Methods.

2. With this value at hand, the main part consisting of the stochastic simulations begins, which is visualized through the big green rectangle in the flowchart.

   (a) For each mini-arch to be simulated, the program chooses random parameters n and p based on which the mini-arch will be filled with filaments, i.e., the 8-cell-long array will be filled with binary values.

   (b) Out of the total number n of stem cells of a particular fate, the number of labeled ones, $L \sim \text{Binomial}(n, probLab)$ and the remaining unlabeled cells $U = n - L$. Then the *in silico* mini-arch generation begins.

      i. One starts with an empty array representing the mini-arch before any filament has been generated.

      ii. The first stem cell to divide and generate a filament is selected by a weighted random choice of whether to add a 1 or a 0 to the empty array, with weights given by probabilities of choosing a labeled cell $\frac{L}{n}$, or an unlabeled cell $\frac{U}{n}$, respectively.

      iii. After the first entry in the array has been added, the following ones depend on the previously inserted value. The previous dividing stem cell will divide again with probability p, so the labeling status of the new filament depends on that of the previous filament via the conditional probabilities (1).

      iv. The procedure stops when the array is filled with eight entries representing one virtual mini-arch.

   (c). Items (i)-(iv) are repeated multiple times to generate a large table of simulated mini-arches for each pair (n, p) of parameters.

3. Once the simulated dataset is obtained, the pair (s, f) is computed for each virtual mini-arch, as it was previously done for the experimental data.

4. Subsequently, the frequency of observing the pair (s, f) is calculated.

5. These frequencies together with those from experimental data are used to construct a mean-squared-error objective function

$$F = \frac{1}{33} \sum_{i=1}^{33} (fD_i - fS_i)^2, \qquad \text{(Equation 2)}$$

where fD_i and fS_i are the frequencies in the data and the simulation results of observing pair $i \in \Pi$, with $\Pi = \{(0,0), (0,8), \ldots, (7,4)\}$ the set of the 33 possible pairs (s, f) (Table S4). The objective function is used within the parameter estimation, to find the best parameters n and p.

---

unlabeled mini-arch (with 0 switches and 0 labeled filaments) and an entirely alternating mini-arch (with seven switches and four labeled filaments) read:

$$\mathbb{P}(s = 0, f = 0) = \mathbb{P}(00000000) = \mathbb{P}(X_1 = 0, X_2 = 0, \ldots, X_8 = 0) = \lambda_2 \cdot p_{22}^7 \text{ and}$$
$$\mathbb{P}(s = 7, f = 4) = \mathbb{P}(10101010) + \mathbb{P}(01010101) = \lambda_1 \cdot p_{12}^4 \cdot p_{21}^3 + \lambda_2 \cdot p_{12}^3 \cdot p_{21}^4$$

Similarly, this method can be employed for each of the 33 possible pairs (s, f) to obtain a list of probabilities with entries corresponding to each pair, depending on the number of stem cells in the niche n, the probability of division p and the labeling efficiency *probLab*. As in the previous numerical method, an MSE-objective function is minimized to obtain the best parameter set. In the analytical approach, the frequencies fS_i in (Equation 2) (Box 2) are replaced by the exact probabilities of observing a pair $(s, f) \in \Pi$, which can be computed as summarized above.

In order to obtain an initial general overview on the parameters' influence on the data, the MSE-objective function (scaled by 10^6) is plotted against the parameter $p \in [0.5, 1]$ for various values $n \in \{1, \ldots, 10\}$ (Figure 4A). These plots suggest that few stem cells are sufficient to generate filaments in the branchial arches, as long as their probability of division is high, indicating a highly heterogeneous division behavior that corresponds to activation and quiescence phases. The plots also show that provided the probability of division

Parametric line plots

A

Pattern 1 · Pattern 2 · Pattern 3 · Pattern 4

MSE ×10⁶

Probability of division

Legend: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Model fit to data

B

Pattern 1: n=2, p=0.919 · Pattern 2: n=4, p=0.779 · Pattern 3: n=2, p=0.955 · Pattern 4: n=10, p=0.912

Frequency of pair (s,f)

Number of switches

Residual plots

C

Pattern 1: n=2, p=0.919 · Pattern 2: n=4, p=0.779 · Pattern 3: n=2, p=0.955 · Pattern 4: n=10, p=0.912

Residuals (fS_i-fD_i)

Model (predicted fS_i)

Figure 4. Results of the heterogeneous model

(A) Parametric line plots showing the MSE-objective function (scaled by 10^6) with respect to the probability of division p for multiple values $n \in \{1, ..., 10\}$ (individual curves, see legend). The smaller the objective function, the better the fit to the data. Hence the curve with the smallest minimum indicates the best parameter values. Plots correspond to patterns 1 to 4, from left to right.

(B) Comparisons between frequencies and probabilities of observing a pair (s, f) in experimental data (D) and model results (M), respectively, for each of the four patterns (1 to 4 from left to right). The x-axis represents the number of switches and the color code corresponds to the number of labeled filaments (from light blue - zero to dark blue - eight). Parameters were estimated by minimizing the objective function computed according to the analytical probabilities, and can be read in the titles of each plot.

(C) Residual plots for comparing the frequencies of pairs (s, f) observed in the data with those predicted by the model, for each pattern (1-4), for assessing the goodness of fit. Each point represents the residual value ($fS_i - fD_i$) plotted against the model predictions (fS_i), for each pair $(s,f) \in \Pi$, with correspondingly displayed labels.

p is high, a greater number of stem cells *n* also fit the data equally well, but we are interested in the minimum number of *br-archSCs* needed for the organ growth, as the rest of the cells would simply not be selected for division within the experiment time span. In other words, more stem cells may reside in the niche, but only few of them are actively involved in filament generation, whereas the rest are dormant, to conserve energy and protect themselves from possible mutations or metabolic damage. In the case of pattern 4, there is a higher variability in the values for *n*, in the sense that multiple *n* values give the same MSE-minimum for similar *p* (Table S5 shows values of the objective function for various combinations of *n* ∈ {2, …, 20} and *p* ∈ [0.9, 0.95]). We attribute this effect to practical non-identifiability arising from the experimental bias and labeling efficiency approximation - see Discussion. In pattern 2, the probability of division *p* is smaller than in the other patterns, suggesting that the activation phases of SCs of fate 2 are shorter. Subsequently, the parameters *n* and *p* are estimated by minimizing the respective objective functions, as discussed before (Box 2).

Both methods provide agreeing results, with similar parameter values, as expected. Figure 4B presents plots based on parameters estimated with the analytical approach, showing a good fit between the model and data, and suggesting that only few stem cells for each pattern are actively participating in organ growth, with high probabilities of division. Some differences can be observed, for example, in the ratio between fully labeled and fully unlabeled mini-arches (dark and light blue in the 0-switches bar), which is influenced by the labeling efficiency, as previously mentioned. This minor mismatch comes from the approximation of the labeling efficiency from the data. Even though the method is based on realistic assumptions, in some cases it does not provide a close enough value to the true one, because the dataset is not sufficiently large for this task ($N = 470$ mini-arches). If a larger dataset were available ($N = 10^4$ mini-arches), a much better approximation could be obtained. Wilson score CIs (Wilson, 1927) for the labeling efficiency are provided in STAR Methods. Nevertheless, the bars for data (D) and model (M) have equal heights in most cases. Because the main defining characteristic of the biological system conveyed in the data is the occurrence of long stretches of filaments carrying the same label (pattern), a good fit aims at first capturing this aspect (i.e., the cases of $s = 0$ switches), as opposed to giving more weight to the substantially fewer cases of nonzero switches. In addition, the frequency of observing scenarios with nonzero switches $s \neq 0$ (for the moment ignoring the colors indicating the number of labeled filaments *f*) is also well recapitulated by the model. Concerning the more balanced distribution of occurrences with different values for the number of labeled filaments *f* for a certain number of switches $s \neq 0$ predicted by the mathematical model (i.e., colors within a bar), this is to be expected based on our model assumptions, and this symmetry does not manifest in the experimental data simply because of the dimension of the dataset. Were such an unbalanced distribution of scenarios with different *f* values for the same *s* to persist in a much larger dataset, it would suggest that an extra level of heterogeneity existed among the stem cells. Figure 4C presents plots of the residuals ($fS_i - fD_i$) against the frequencies predicted by the model (fS_i), for each pattern, for assessing the goodness of fit. In addition, Figures S2–S5 show the stacked plots (as in 4B) and residual plots (as in 4C) for the optimal, the homogeneous, and two further "bad" scenarios, for comparison; the value of the objective function is also displayed for each case. The results presented in Figure 4 thus indicate that few but highly heterogeneous stem cells are enough to contribute filaments to the medaka gills during the entire postembryonic and adult life.

DISCUSSION

This study has presented two alternative methods for determining the numbers and functional heterogeneities among stem cells during postembryonic gill growth: a numerical approach via stochastic simulations, which generates a large dataset of simulated mini-arches for comparison to experimental data, and a more abstract analytical approach, which skips the step of *in silico* data generation, is exact and less time-consuming. The two methods are based on the same assumptions and provide equivalent results. These methods contributed to the discovery of novel insights into the behavior of the stem cells responsible for building, growing, and maintaining the respiratory organ of fish: (i) few branchial arch stem cells are participating in the postembryonic and adult organ growth and, more importantly, (ii) these stem cells are functionally heterogeneous in their division behavior, in the sense that they follow phases of activation and quiescence, such that once a stem cell has divided to generate a filament, it becomes active and divides multiple times thus creating more clonal filaments, before becoming quiescent and allowing another stem cell of the same type to take over the task of filament generation.

This activation/quiescence behavior is valid for all four fate-restricted stem cells, corresponding to the four possible patterns observed in filaments. In all cases, although more variability is observed in the values for

the number of SCs n, essential for the biological system is that the heterogeneity parameter p describing the probability of division is very high. Stem cells of Pattern 2 have a smaller probability of division, relative to the others, which suggests that their activation phases are shorter, and this comes together with a higher number of SCs of the second type, which participate in organ growth. In Pattern 4 more variability is observed (Figure 4A), which is caused by the low labeling efficiency in the case of the fourth stem cell type. This low labeling efficiency may also be a consequence of the fact that Pattern 4 is difficult to spot when mixed patterns are present, being hidden beneath other more prominent patterns (e.g., Pattern 2). A suitable method to overcome this issue would be to label the different types of cells (patterns) with different colors, but such method is currently not available with the required cellular-resolution for the gill system.

The results of this study improve our understanding of the coordination between the filament-generating stem cells, which get recruited as an ensemble to a newly forming filament. The aspect of how this co-ordination is accomplished remains an open question, but our studies have shown that approximately equal numbers of the fate-restricted stem cells are responsible for organ growth, which supports the mechanism of maintaining the ratio of cell types within a filament, guaranteeing its proper development. The heterogeneous division behavior of stem cells that we now describe in the fish gill has been previously reported in other systems. For example, Bogdan et al. (2014) have studied the proliferative heterogeneity in muscle cells of mice and rats and in human mesenchymal stem cells and reported that stem cells growth rates possess multifractal characteristics. Alternating activation and quiescence phases have also been recently suggested in neurogenesis (Harris et al., 2021; Kalamakis et al., 2019; Urbán et al., 2016; Ziebell et al., 2018; Basak et al., 2018), where it is speculated that proliferating stem cells return to quiescence into a pool of temporary quiescent cells, which is separate from the main dormant stem cells. This behavior can be thought of as a defense mechanism against the possible steps which can fail during division, taking into account the small number of stem cells which carry the responsibility of filament generation, and thus, organ growth. Considering that more stem cells for the optimal probability of division do not improve nor make the fit worse, we expect that in addition to the respective optimal number of stem cells found by our model to drive the organ growth, more stem cells of each particular fate reside in the niche. However, according to our model and experimental data, the majority of the "extra" stem cells should be dormant, only becoming active when one of the main stem cells fails and dies.

Limitations of the study

This work considers the study of stem cell numbers and functional heterogeneities in gill stem cells. As stem cell specific markers are not available in the gill system, lineage tracing experiments are performed with a permanent, ubiquitous label that is inherited by progeny. This may result in errors when recording the distribution of filamental patterns, as certain patterns may be hidden under others and thus overlooked. These possible errors may also result in an underestimation of the labeling efficiency for certain patterns.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal husbandry and ethics
- METHOD DETAILS
  - Generation of clones
  - Staining protocol and imaging
  - Experimental data adjustment
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Labelling efficiency estimation
  - Transition probabilities computation

AUTHOR CONTRIBUTIONS

Model development, D-P.D. and A.M-C.; Algorithm implementation, D-P.D.; Formal analysis, D-P.D. and A.M-C.; Data curation, D-P.D., J.S., and L.C; Interpretation of the modeling results, D-P.D. and A.M-C.; Writing - original draft, review, and editing, D-P.D., J.S., L.C., and A.M-C.; Data acquisition, J.S. and L.C.; Conceptualization, L.C. and A.M-C.; Supervision, L.C. and A.M-C.

REFERENCES

Basak, O., Krieger, T.G., Muraro, M.J., Wiebrands, K., Stange, D.E., Frias-Aldeguer, J., Rivron, N.C., van de Wetering, M., van Es, J.H., van Oudenaarden, A., et al. (2018). Troy+ brain stem cells cycle through quiescence and regulate their number by sensing niche occupancy. Proc. Natl. Acad. Sci. *115*, E610–E619. https://doi.org/10.1073/pnas.1715911114.

Bogdan, P., Deasy, B.M., Gharaibeh, B., Roehrs, T., and Marculescu, R. (2014). Heterogeneous structure of stem cells dynamics: statistical models and quantitative predictions. Scientific Rep. *4*, 1–11.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S.M., Reth, M., Höfer, T., and Rodewald, H.R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542. https://doi.org/10.1038/nature14242.

Centanin, L., Ander, J.J., Hoeckendorf, B., Lust, K., Kellner, T., Kraemer, I., Urbany, C., Hasel, E., Harris, W.A., Simons, B.D., et al. (2014). Exclusive multipotency and preferential asymmetric divisions in post-embryonic neural stem cells of the fish retina. Development. https://doi.org/10.1242/dev.109892.

Dambly-Chaudière, C., Sapède, D., Soubiran, F., Decorde, K., Gompel, N., and Ghysen, A. (2003). The lateral line of zebrafish: a model system for the analysis of morphogenesis and neural development in vertebrates. Biol. Cell *95*, 579–587. https://doi.org/10.1016/j.biolcel.2003.10.005.

Ghysen, A., and Dambly-Chaudière, C. (2007). The lateral line microcosmos. Genes Dev. *21*, 2118–2130. https://doi.org/10.1101/gad.1568407.

Harris, L., Rigo, P., Stiehl, T., Gaber, Z.B., Austin, S.H., del Mar Masdeu, M., Edwards, A., Urbán, N., Marciniak-Czochra, A., and Guillemot, F. (2021). Coordinated changes in cellular behavior ensure the lifelong maintenance of the hippocampal stem cell population. Cell Stem Cell. https://doi.org/10.1016/j.stem.2021.01.003.

Iwamatsu, T. (2004). Stages of normal development in the medaka oryzias latipes. Mech. Dev. *121*, 605–618. https://doi.org/10.1016/j.mod.2004.03.012.

Kalamakis, G., Brüne, D., Ravichandran, S., Bolz, J., Fan, W., Ziebell, F., Stiehl, T., Catalá-Martinez, F., Kupke, J., Zhao, S., et al. (2019). Quiescence modulates stem cell maintenance and regenerative capacity in the aging brain. Cell *176*, 1407–1419.

Leguen, I. (2018). Gills of the medaka (Oryzias latipes): a scanning electron microscopy study. J. Morphol. *279*, 97–108. https://doi.org/10.1002/jmor.20757.

Lee-Six, H., Øbro, N.F., Shepherd, M.S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R.J., Huntly, B.J., Martincorena, I., Anderson, E., et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. Nature *561*, 473–478.

Laurent, P. (1984). 2 gill internal morphology. Fish Physiol. *10*, 73–183.

Laurent, P., and Dunel, S. (1980). Morphology of gill epithelia in fish. Am. J. Physiol. Regul. Integr. Comp. Physiol. *238*, R147–R159.

Olson, K.R. (2002). Vascular anatomy of the fish gill. J. Exp. Zool. *293*, 214–231.

Pei, W., Feyerabend, T.B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature *548*, 456. https://doi.org/10.1038/nature23653.

Seleit, A., Krämer, I., Riebesehl, B.F., Ambrosio, E.M., Stolper, J.S., Lischik, C.Q., Dross, N., and Centanin, L. (2017). Neural stem cells induce the formation of their physical niche during organogenesis. Elife *6*, e29173. https://doi.org/10.7554/eLife.29173.

Snippert, H.J., Van Der Flier, L.G., Sato, T., Van Es, J.H., Van Den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., Van Rheenen, J., Simons, B.D., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing lgr5 stem cells. Cell *143*, 134–144. https://doi.org/10.1016/j.cell.2010.09.016.

Stolper, J., Ambrosio, E.M., Danciu, D.P., Elliott, D., Naruse, K., Marciniak-Czochra, A., and Centanin, L. (2019). Hierarchical stem cell topography splits growth and homeostatic functions in the fish gill. eLife *6*, e43747. https://doi.org/10.7554/eLife.43747.

Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322. https://doi.org/10.1038/nature13824.

Tsingos, E., Hoeckendorf, B., Suetterlin, T., Kirchmaier, S., Grabe, N., Centanin, L., and

Wittbrodt, J. (2019). Retinal stem cells modulate proliferative parameters to coordinate post-embryonic morphogenesis in the eye of fish. eLife 8, e42646. https://doi.org/10.7554/eLife.42646.

Urbán, N., van den Berg, D.L., Forget, A., Andersen, J., Demmers, J.A., Hunt, C., Ayrault, O., and Guillemot, F. (2016). Return to quiescence of mouse neural stem cells by degradation of a proactivation protein. Science 353, 292–295. https://doi.org/10.1126/science.aaf4802.

Wada, H., Dambly-Chaudière, C., Kawakami, K., and Ghysen, A. (2013). Innervation is required for sense organ development in the lateral line system of adult zebrafish. Proc. Natl. Acad. Sci. 110, 5659–5664. https://doi.org/10.1073/pnas.1214004110.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. J. Am. Stat. Assoc. 22, 209–212.

Wittbrodt, J., Shima, A., and Schartl, M. (2002). Medaka-a model organism from the far east. Nat. Rev. Genet. 3, 53. https://doi.org/10.1038/nrg704.

Ziebell, F., Dehler, S., Martin-Villalba, A., and Marciniak-Czochra, A. (2018). Revealing age-related changes of adult hippocampal neurogenesis using mathematical models. Development 145, dev153544. https://doi.org/10.1242/dev.153544.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
a-EGFP (Rabbit IgG polyclonal)	Invitrogen (Thermo Fischer)	CAB4211; RRID: AB_10709851
Alexa 488 Goat a-Rabbit	Invitrogen (Thermo Fischer)	A-11034; RRID: AB_2576217
Chemicals, peptides, and recombinant proteins		
Tamoxifen	Sigma-Aldrich	T5648
Tricaine	Sigma-Aldrich	A5040-25G
DAPI	Roth	
Deposited data		
Filamental pattern distribution (clonal data)	This paper	https://doi.org/10.5281/zenodo.5847028
Experimental models: Organisms/strains		
Wild type *Oryzias latipes*, Cab		
Transgenic *Oryzias latipes*, GaudíUbiq.iCre	Centanin et al. (2014)	
Transgenic *Oryzias latipes*, GaudíHsp70.A	Centanin et al. (2014)	
Transgenic *Oryzias latipes*, GaudíRSG	Centanin et al. (2014)	
Software and algorithms		
Fiji		https://fiji.sc/
Mathematica 12.0	Wolfram Research	
Original code	This paper	https://doi.org/10.5281/zenodo.5847028; https://github.com/dpdanciu/Gill-SCnumbers-heterogeneities

RESOURCE AVAILABILITY

Lead contact

Further information can be provided by the lead contact, Anna Marciniak-Czochra (anna.marciniak@iwr.uni-heidelberg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The clonal data used in this study is deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.

- The code generated during this study is deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table. Alternatively, it can be directly accessed on GitHub at https://github.com/dpdanciu/Gill-SCnumbers-heterogeneities

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal husbandry and ethics

All experimental animals were under the supervision of the animal welfare officer at the Heidelberg University. The fish stocks of *Oryzias latipes* (medaka) were maintained according to local animal welfare laws (Tierschutzgesetz §11, Abs. 1, Nr.1) and the European Union animal welfare guidelines. Fish were maintained and raised in a constant recirculation system at 28°C cycling between 14 hours of light

and 10 hours of darkness (Tierschutzgesetz §11, Abs. 1, Nr.1, Haltungserlaubnis AZ35-9185.64 and AZ35-9185.64/BH KIT). Fish lines being used in this study include wild-type medaka (Cab, medaka Southern population strain) and transgenic fish of the Gaudí living toolkit (Centanin et al., 2014): Gaudí $^{Ubiq.iCre}$, expressing a tamoxifen inducible Cre-recombinase, Gaudí $^{HSP70.A}$, expressing a heat-shock inducible CRE-recombinase and Gaudí ^{RSG}, containing a genetic cassette that switches from a ubiquitous expression of a red fluorescent protein (RFP) to a nuclear green signal (nGFP) upon recombination. This study has analyzed both male and female embryos staged according to Iwamatsu (2004), heat-shocked at stages 20, 24, 29, 32, 34 or 37, or induced via tamoxifen treatment at stage 36 (see Method details - Generation of clones), and then raised until adulthood, when they were imaged. Sex specific differences were not investigated in this study.

METHOD DETAILS

Generation of clones

Clonal data was generated via lineage tracing analysis. Genetic recombination in double transgenic fish (Gaudí $^{HSP70.A}$, Gaudí ^{RSG}) was induced via heat-shock. Male and female embryos were staged according to Iwamatsu (2004) and heat-shocked at stages 20, 24, 29 32, 34 or 37 using embryo rearing medium (ERM) at 42°C and transferred to 37°C for 1 to 3 hours and raised until adulthood. Sex specific differences were not investigated in this study.

Genetic recombination was induced via tamoxifen treatment in Gaudí $^{Ubiq.iCre}$ Gaudí ^{RSG} double transgenic fish at stage 36. Embryos were kept in ERM containing tamoxifen (T5648 Sigma, 5μM final concentration) for 3 hours, rinsed multiple times with fresh ERM to ensure removal of residual tamoxifen, and placed in a tank until they reach adulthood. Fish that resulted in a high recombination efficiency (i.e. entire branchial arch labelled) were not used in the analysis.

Staining protocol and imaging

All fish were euthanised using a 2mg/ml Tricaine solution (Sigma-Aldrich, A5040-25G), fixed in 4% PFA/PTW at 4°C overnight and the entire gills were micro-dissected to continue with the staining protocol. To permeabilise the tissue, gills were kept in acetone at −2°C for 10 minutes. After blocking with goat serum for 1 hour at room temperature, GFP staining (Rabbit a-GFP, Invitrogen, 1:750) was performed overnight at 4°C. The secondary antibody (Alexa 488 a-Rabbit, Invitrogen, 1:500) was incubated together with DAPI (final concentration: 5μg/μl) for 2 hours at room temperature. Gills were separated into single branchial arches and mounted in glycerol 50% between cover slides (Stolper et al., 2019). Whole gills were imaged using an Olympus MVX10 microscope connected to a Leica DFC500. On confocal resolution, branchial arches were imaged using Leica TCS SP8 and SP5 II microscopes. Image analysis and stitching was performed in Fiji.

Experimental data adjustment

The experimental data consisted of a table with rows of unequal lengths, each of which represented one branchial arch. Each such array contained values describing the labelling status and respective pattern or combination of patterns, for each filament in the arch. An unlabelled filament was recorded as an element $i = 0$, a filament presenting one pattern was denoted by $i \in \{1, 2, 3, 4\}$, while a filament presenting mixed patterns was presented as a multiple digit number, with digits recording the patterns present, e.g. for a combination of two patterns within one filament the options are $i = \overline{ab}$, with $a, b \in \{1, 2, 3, 4\}$, $a<b$ (Table S1, excerpt from experimental data). A total of $N = 340$ branchial arches were quantified. From these, only the branchial arches/arrays of length ≥ 25 were considered, resulting in a total of $N = 235$ rows. Subsequently, four separate data sets were created, one for each pattern, by only considering the filaments presenting that specific pattern as labelled (1), and the other cases as unlabelled (e.g. in the adjusted data for Pattern 1, the filaments with patterns 2, 3 and 4, together with the unlabelled filaments are recorded as 0) - see Table S2, data excerpt adjusted for each pattern. For all patterns, from each branchial arch, the 8 peripheral filaments from each side were selected, i.e. the first 8 and last 8 entries of each array. All such mini-arches were arranged so that they start with the oldest filament, i.e. the mini-arrays containing the first 8 entries from the original array were flipped. The final four data sets used for the model, one for each pattern, each consisted of a 8×2N = 8×470 matrix (Table S3, final adjusted excerpt).

QUANTIFICATION AND STATISTICAL ANALYSIS

Labelling efficiency estimation

The labelling efficiency was estimated for the entire experimental data set by employing a combinatorial approach. The labelling efficiency would indicate how many stem cells are labelled out of a large pool of SCs. To keep in mind is the fact that a labelled filament is produced by a labelled stem cell, so if one looks at the oldest post-embryonic filament, one finds the label of the first SC which divided in that specific branchial arch (on each side). We look at the oldest filament because, in this way, we avoid the influence of the heterogeneity parameter. Accordingly, by counting the number of labelled first filaments (i.e. oldest) out of the total number of branchial arches, we obtain an approximation of the average labelling efficiency across the entire data set. We thus obtain the following formula, where nL is the number of mini-arches which start with a labelled filament, i.e. the number of arrays that start with a non-zero (1) entry; and N is the total number of rows in the adjusted data set.

$$probLab = \frac{\text{Number of labelled first filaments}}{\text{Number of branchial arches}} = \frac{nL}{N} \qquad \text{(Equation 3)}$$

Accordingly, for each pattern the labelling efficiency *probLab* reads as in the table below, with respective 95% binomial proportion confidence intervals computed with the Wilson score (Wilson, 1927).

Pattern	Pattern 1	Pattern 2	Pattern 3	Pattern 4
probLab	0.1553	0.0638	0.0426	0.0149
Wilson CI	[0.1254, 0.1908]	[0.0451, 0.0897]	[0.0277, 0.0648]	[0.0072, 0.0304]

Transition probabilities computation

First, note that we choose $p \in [0.5, 1)$ based on our previous heterogeneity hypothesis. A probability of division of $p = 0.5$ corresponds to an entirely functionally homogeneous system, in which the probability of the previously diving cell to divide again is equal to that of another random cell of the same type to take over. A probability of division $p > 0.5$ corresponds to our hypothesis of a functionally heterogeneous system with activation and quiescence phases. In addition, $p < 1$, since $p = 1$ would be equivalent to a case where the entire mini-arch is created by one stem cell, case which we proved infeasible (Figure 2D and (Stolper et al., 2019)).

For computing the transition probabilities of the Markov process, recall the formula for conditional probabilities

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

Suppose i is the cell which has just divided and j the cell about to divide. Further, denote by cL the event of choosing a labelled cell, and by cU the event of selecting an unlabelled cell. A transition $L \rightarrow L$ i.e. a labelled filament followed by another labelled one, corresponds to a case where either the previously diving cell i was labelled and it divides again ($j = i$), or if another labelled cell is selected ($j \neq i$, with i, j labelled). All transition probabilities can be similarly considered and recorded in the transition probability matrix P with entries (4) for $n \geq 2$, recalling that $p = \mathbb{P}(j = i)$ irrespective of the labelling status. For $n = 1$, the matrix $P = I_2$ the identity matrix, but this case is not expected (Figure 2D and (Stolper et al., 2019)).

Entries of the transition probability matrix P, for $n \geq 2$:

$$p_{11} := \mathbb{P}(j \text{ labelled}|i \text{ labelled}) = \mathbb{P}(j = i) + \mathbb{P}(j \neq i)\mathbb{P}(cL) = p + (1 - p)\frac{L - 1}{n - 1}$$

$$p_{12} := \mathbb{P}(j \text{ unlabelled}|i \text{ labelled}) = \mathbb{P}(j \neq i)\mathbb{P}(cU) = (1 - p)\frac{U}{n - 1} \qquad \text{(Equation 4)}$$

$$p_{21} := \mathbb{P}(j \text{ labelled}|i \text{ unlabelled}) = \mathbb{P}(j \neq i)\mathbb{P}(cL) = (1 - p)\frac{L}{n - 1}$$

$$p_{22} := \mathbb{P}(j \text{ unlabelled}|i \text{ unlabelled}) = \mathbb{P}(j = i) + \mathbb{P}(j \neq i)\mathbb{P}(cU) = p + (1 - p)\frac{U - 1}{n - 1}$$

Objective function definition. The objective function describes a measure of the "distance" between the model and the experimental data. In this discrete system, the data are quantified based on the (s, f)

pairs, so the objective function compares the frequencies of observing such a pair, in the data and in the model. There exist 33 possible pairs describing a mini-arch since, out of the total 8×9 pair combinations, most are infeasible due to configuration dependency constraints. For example, there exists no mini-arch corresponding to a pair $(s, f) = (1, 8)$ because in an entirely labelled mini-arch ($f = 8$) no switches exits ($s = 0$). The set Π of all possible pairs was obtained by implementing a short piece of code investigating each mini-arch configuration (256 configurations) - see Table S4 for all possible configurations and details. The objective function is thus constructed as the mean squared error:

$$F = \frac{1}{33} \sum_{i=1}^{33} (fD_i - fS_i)^2,$$

(Equation 5)

where fD_i and fS_i are the frequencies observed in the data and those predicted by the simulations of observing pair $i \in \Pi$, with $\Pi = \{(0, 0), (0, 8), \ldots, (7, 4)\}$ the set of the 33 possible pairs (s, f).