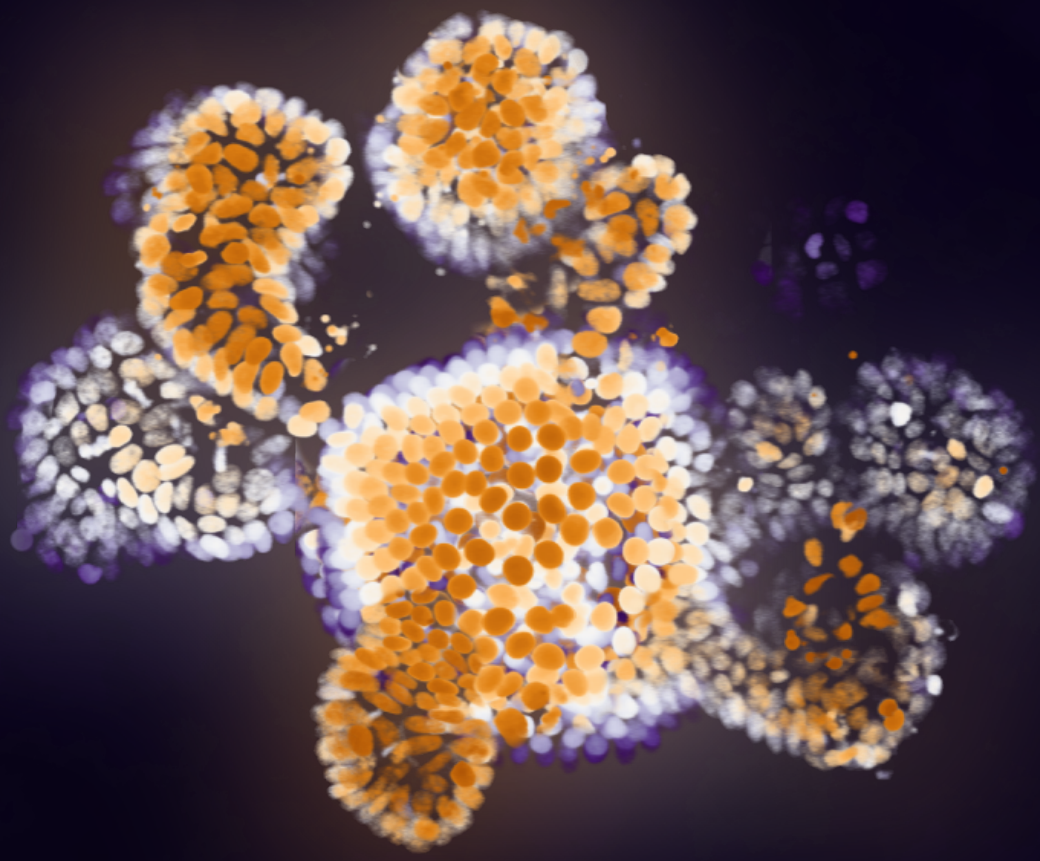
Homeostasis in intestinal organoids at the single cell level

Rutger Kok



Homeostasis in intestinal organoids at the single cell level

Homeostasis in intestinal organoids at the single cell level

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op woensdag 10 april 2024 om 15:00 uur

door

Rutger Nico Ulbe KOK

doctorandus in de chemische wetenschappen,
Universiteit Utrecht, Nederland,

geboren te Naarden, Nederland.

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. S.J. Tans	Technische Universiteit Delft, promotor
Prof. dr. P.R. ten Wolde	Vrije Universiteit Amsterdam, copromotor
Onafhankelijke leden:	
Prof. dr. M.E. Tanenbaum	Technische Universiteit Delft
Prof. dr. L.J.W. van der Laan	Erasmus Universitair Medisch Centrum Rotterdam
Prof. dr. T.S. Shimizu	Vrije Universiteit Amsterdam
Dr. K.F. Sonnen	Hubrecht Institute
Prof. Dr. C. Joo	Technische Universiteit Delft, reservelid
Overig lid:	
Dr. J.S van Zon	AMOLF

Dr. J.S. van Zon heeft in belangrijke mate aan de totstandkoming van het proefschrift bijgedragen.

Trefwoorden:	organoïden, dunne darm, homeostase, celvolgen
Afgedrukt door:	Proefschrift-AIO
Voor- & achterkant:	Fluorescente celkernen van een organoïd. Het organoïd is een digitale constructie, gebaseerd op meerdere darmorganoïden.

Een digitale versie van dit proefschrift is beschikbaar op
`https://repository.tudelft.nl/` en
`https://ir.amolf.nl/`.

Contents

Summary

How does the small intestine maintain itself? What mechanism does it use to achieve homeostasis, and how optimal is this mechanism? We approached these questions mainly using cell tracking in organoids, for which we developed new cell trackers. We found that the intestinal crypt uses a surprisingly simple and effective strategy. Finally, we showed that cell segmentation and tracking is possible in organoids without fluorescent markers.

As cell tracking is such an important method in this thesis, we strived to automate this process. In **Chapter 2** we developed a cell tracker named OrganoidTracker, which combines machine learning approaches with a min-cost flow solver. Like any cell tracker, the output of the cell tracker contains mistakes. The graphical user interface of OrganoidTracker finds inconsistencies in the cell trajectories, and allows the user to efficiently correct those.

In **Chapter 3** we set out to explore the design space in which stem cell niches can function. A stem cell niche needs to keep both the number of proliferative and the number of non-proliferative cells constant over time. We wondered what proliferation strategy keeps fluctuations in the number of proliferative cells minimal. Previously, it was assumed that a system with asymmetric divisions, which are divisions where only one of the two daughters cells will proliferate, leads to minimal fluctuations. However, if we put part of the cells in a stem cell niche with limited space, then an improved strategy becomes available. Within the stem cell niche, cells must maximally proliferate, and once a cell is pushed out of the stem cell niche, it must stop proliferating as soon as possible. This leads to two very distinct groups of cells, with maximally different proliferation behavior, and no asymmetric divisions.

While such a system is optimal for maintaining a stem cell niche (given our assumptions), it is still unknown what system the intestinal crypt uses for controlling proliferation. In **Chapter 4**, we used time-lapse microscopy to track all cell in the crypt, and combined this data with clone size distributions from mice. Together, the data convincingly shows that the crypt uses a two-compartment system for controlling proliferation, with a minimal number of asymmetric divisions.

In the next chapter (**Chapter 5**), we investigated the underlying molecular mechanism for the stem cell maintenance. While it is known that the Wnt signaling pathway controls stemness, it is not known how cells interpret this signal in the noisy and dynamic environment. Often, a spatial model is assumed, where cells differentiate if they move too far away from a Wnt-producing cell, but this has never been verified.

We imaged organoids with a marker for Axin2, one of the most common Wnt target genes. As it turns out, cells use a timer model instead, which makes cells differentiate if Paneth cell contact has been lost for about 10 hours.

In our thesis, we have always tracked cells using fluorescent markers, as for organoids this seemed to be the only possible option. However, recent work showed for 2D cell lines that it is possible to reconstruct cell nuclei and membranes [1, 2]. In **Chapter 6** we show that this technique surprisingly works for the complex 3D tissue of intestinal organoids, and that the performance of this approach reaches a level that allows cell segmentation and cell tracking for over 60 hours.

Taken together, this thesis has studied the way in which the intestinal crypt achieves homeostasis. We used theoretical approaches, observational studies and a molecular investigation to conclude that the crypt operates under a two-compartment system, with both compartments demarcated using a Wnt signaling timer. We have provided tools to track cells in organoids, both for organoids with a fluorescent nuclear marker, but also for label-free organoids.

Samenvatting

Hoe onderhoudt de dunne darm zichzelf? Welk mechanisme gebruikt het om in homeostase te blijven, en hoe optimaal is dit mechanisme? We hebben vooral naar deze vragen gekeken door cellen te volgen in organoïden, waarvoor we nieuwe celvolgprogramma's ontworpen hebben. We hebben ontdekt dat de dunne darm een verrassend simpele en effectieve strategie gebruikt om zichzelf te onderhouden. En tot slot hebben we gevonden dat het mogelijk is om cellen te volgen in de organoïden zonder fluorescente markers.

Omdat het volgen van cellen zoveel gedaan wordt in dit proefschrift, hebben we dit proces zoveel mogelijk geautomatiseerd. In **hoofdstuk 2** hebben we een celvolgprogramma ontworpen genaamd OrganoidTracker, die machinaal leren combineert met een min-kosten-stroomoplosser. Zoals alle celvolgprogramma's maakt ook deze celvolger fouten. Wat bijzonder is aan OrganoidTracker is dat het een grafische gebruikersinterface heeft die zoekt naar inconsitenties in de celtrajecten, zodat de gebruiker ze vlot kan corrigeren.

In **hoofdstuk 3** hebben we gekeken naar de mogelijke ontwerpen voor stamcelniches. Een stamcelniche moet zowel het aantal delende cellen als het aantal niet-delende cellen constant houden. We vroegen ons af wat voor delingsstrategie de fluctuaties in het aantal delende cellen minimaliseert. Voorheen werd aangenomen dat je hiervoor asymmetrische delingen nodig had, wat delingen zijn waarbij slechts één van de twee dochters blijft doorgaan met delen. Echter, als we een deel van de cellen in een stamcelniche plaatsen met beperkte ruimte, dan wordt er een betere strategie mogelijk. Binnen de stamcelniche moeten cellen maximaal blijven delen, maar zodra een cel buiten de stamcelniche terecht komt moet deze zo snel mogelijk stoppen met delen. Dit systeem leidt tot twee heel verschillende groepen cellen, met maximaal verschil in hun delingsgedrag, en met nauwelijks asymmetrische celdelingen.

Alhoewel zo'n systeem in theorie optimaal is voor het onderhouden van een stamcelniche (onder onze aannames althans), is het nog steeds onbekend wat voor systeem de crypten in de dunne darm daadwerkelijk gebruiken. In **hoofdstuk 4** hebben we *time lapses*van de organoïden onder een microscoop gemaakt, gecombineerd met kloongroottedistributies van muizen. Samen laat deze data overtuigend zien dat de crypt een tweecompartimentssysteem gebuikt om proliferatie te regelen, wat tot gevolg heeft dat

In het volgende hoofdstuk (**hoofdstuk 5**) hebben we het onderliggende moleculaire mechanisme voor dit stamcelonderhoud bestudeerd. Men wist al dat de Wnt-signaalweg bepaalt of een cel een stamcel is. Echter, hoe de cellen dit signaal interpreteren was nog

niet bekend. Vaak wordt een ruimtelijk model aangenomen, waarbij stamcellen gaan differentiëren zodra ze te ver weg bewegen van een Wnt-producerende cel, maar dit is nooit getest. Wij hebben video's van organoïden met een marker voor *Axin2* opgenomen, wat één van de meest algemene gen-doelwitten is van Wnt. Het bleek dat cellen geen ruimtelijk model volgen, maar een zandlopermodel: stamcellen differentiëren zodra het contact met de Panethcellen is verloren gedurende ongeveer 10 uur.

In alle hoofdstukken tot dit punt hebben we de cellen gevolgd aan de hand van fluorescent markers, omdat dit voor organoïden de enige optie leek. Echter, voor 2D-cellijnen bleek recent dat het mogelijk is om de celkernen en -membranen te reconstrueren zonder fluorescentie [1, 2]. In **hoofdstuk 6** laten we zien dat deze techniek verrassend genoeg ook werkt voor organoïden, en dat de methode zelfs goed genoeg presteert om cellen 60 uur te kunnen volgen.

Samengenomen hebben we in dit proefschrift onderzocht hoe de dunnen darm homeostase bereikt. We hebben theoretische methodes gebruikt, observaties en een moleculair onderzoek om te concluderen dat de crypt werkt met een tweecompartimentssysteem, waarbij een Wnt-zandloper de afbakening van de compartimenten verzorgt. We hebben software voor celvolgen ontwikkeld, niet alleen voor organoïden met fluorescent markers, maar ook voor organoïden zonder markers.

1 | Introducing cell tracking in organoids

Using parts of
Betjes, M.A., Zheng, X., **Kok, R.N.U.**, Van Zon, J.S., & Tans, S.J. (2021).
Cell tracking for organoids: lessons from developmental biology. *Frontiers in Cell and Developmental Biology*, 1347.

Abstract

Development and maintenance of organs and organoids is often studied using static images. In combination with genetic markers, this allows for inference of certain dynamic information. However, this method does not yield complete lineage trees, nor provide temporal dynamics of cells. Therefore cell tracking, a technique that allows one to track individual cells through space and time, holds great promise to elucidate the organizational principles of organs and organoids. Applied extensively in the past decade to embryonic development and 2D cell cultures, cell tracking can reveal the cellular lineage trees, proliferation rates, and their spatial distributions, while fluorescent markers indicate differentiation events and other cellular processes. Here, we review a number of recent studies that exemplify the power of this approach, and illustrate its potential to organoid research. We will discuss promising future routes, and the key technical challenges that need to be overcome to apply cell tracking techniques to organoid biology.

1.1 | Lineage tracing

While the development and maintenance of organs is one of the most fundamental problems in biology, our understanding of it is far from complete. A hallmark of this process is the differentiation of cells in time, in terms of proliferative potential and cell type, with individual cells giving rise to complex lineages that organize in space to shape tissues and organs. Thus far, these differentiation dynamics have often been studied using the lineage tracing method [3–5]. Here, cells are labeled with a heritable marker such as fluorescent genes or a genetic barcode, for instance using Cre-Lox recombination [6, 7] or lentiviral transduction [8–10]. This label can be detected in progeny after a certain period by fluorescence microscopy or single-cell sequencing, and hence allows inference of genealogical relations between cells.

However, lineage tracing does not yield complete lineage trees nor provide information on the temporal dynamics of cells, such as their movements, growth rates, transient signaling, and timing of differentiation events (**Fig. 1**), which limits progress on many important questions. For instance, it remains largely unclear when and where cell fates are actually set, whether differentiation is either a consequence or a cause of spatial organization, how size and shape homeostasis is achieved, or how lineage dynamics are remodeled upon injury or disease. We also know little about the possible interplay with cellular metabolism, and the plethora of molecular signals from adjacent cells, for instance from the immune system. Elucidating the spatio-temporal dynamics is central to resolve these crucial issues and to elucidate the organizational principles of organ development [11].

1.2 | Cell tracking

A different technique, here referred to as cell tracking, has the potential to reveal these developmental dynamics. As opposed to lineage tracing based on static snapshots, cells are here followed in real time over multiple generations, which can thus provide temporal dynamics, complete lineage trees, as well as spatial organization and cellular movements. Furthermore, cell tracking can readily be combined with the large spectrum of microscopy techniques that have been developed to study cell biology. For instance, the expression of fluorescent proteins directly identify differentiation events, cell-cycle progression, cytoskeletal structures, the dynamics of key molecular signals like Wnt or Notch, while FRET sensors may detect more rapidly changing signals such as calcium and metabolites [12–14]. Cell tracking has been applied extensively to study the early development of embryos, at increasing levels of sophistication [15–17], but poses challenges to the study of organs, given the challenges of time-lapse microscopy deep within tissues at later stages of development, even as intra-vital imaging is possible at lower resolution and throughput [18].

Yet, in recent years organoids have emerged as a model system for studying development and disease at the cellular level, including patient derived systems, which are ideally suited for this approach. Organoids are self-organizing cellular assemblies, which are grown in vitro and recapitulate organ structure and functionality to a striking degree [19, 20]. Due to their in vitro nature, the growth and maintenance of organoid tissue can be observed directly by time-lapse microscopy [21]. However, cell tracking approaches have only scarcely been applied to organoids thus far. This is likely because of the technical challenges of large-scale cell tracking in organoids, which in practice requires (semi-) automatic cell tracking.

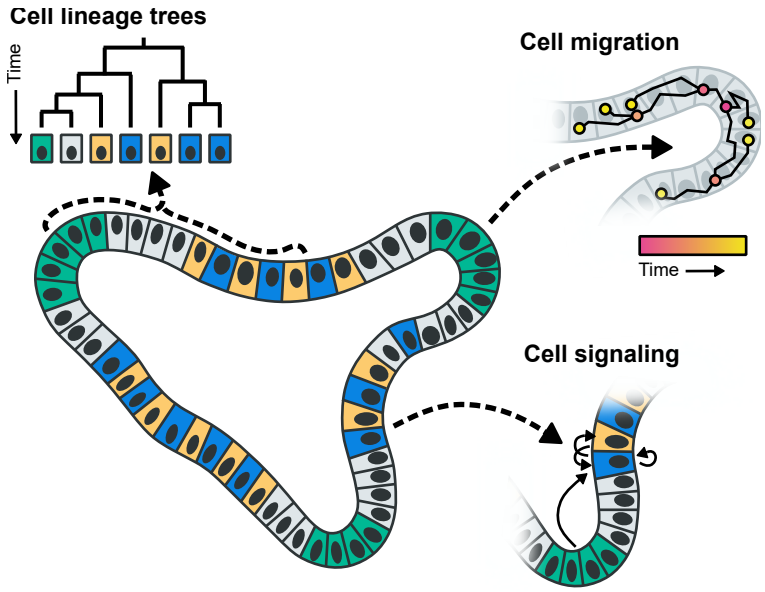


Figure 1 • Organoid properties that can be studied by cell tracking
Cell tracking is a technique in which (nearly) all cells are followed in time over multiple generations, using 3D time-lapse microscopy and automated image analysis, along with a host of fluorescence detection methods. Tracked cell positions allow one to reconstruct lineage trees, follow migration, growth and division dynamics, while fluorescence reporters may be used to study differentiation events, signaling pathways, and metabolic states, which are key to understand cellular decision-making during development.

1.3 | Automatic cell tracking

Automated tracking of individual cells in time has become a powerful approach to study cellular dynamics in cell lines and embryos [22]. Pioneering examples include tracking of embryos of C. elegans, fruit flies, zebrafish and mice [23–25]. In a recent paper, 10^4 cells were tracked in growing mouse embryos over 48 hours, during gastrulation and early organogenesis [26]. The authors imaged cell nuclei using adaptive multi-view light sheet imaging. Cells were tracked using a Gaussian mixture model, with the center of each nucleus determined by fitting their fluorescent signal to a 3D Gaussian function. By assigning cells present at the end of the experiment to different tissues, e.g. the heart field or the neural tube, based on anatomical features, and by following the tracks of these cells backward in time, it was possible to reconstruct how these cells flowed out of the primitive streak and assembled into tissues. This revealed that the both the timing and position of cells as they left the primitive streak was key to determining their cell fate. Moreover, by tracking cell divisions in time, the authors showed that the orientation of cell divisions changes several times during neural tube closure, with important impact on tissue morphology.

Despite the advance in analyzing mouse embryogenesis, the study also underscores the formidable technical challenges that remain when studying development, including in organoids. While the tracking error rate was low enough to reconstruct the general flow of cells, it was too high to automatically reconstruct lineages in most parts of the embryo. This is because a single erroneous switch in cell identity can corrupt large parts of the lineage tree. Instead, the authors used a mosaic Cre/loxP reporter to sparsely labels small subsets of cells. This strongly reduced cell identity mistakes, by increasing the spacing between tracked cells, but severely reduced the lineage information that was captured. While such analysis of cell flow coupled to fate is useful for many applications, acquiring more exhaustive lineage information is of particular importance for organoids, for instance to identify the rare differentiation events and correlations between them, or to reveal spatial interactions on short length scales, including those between neighboring cells that originate from cell-cell signaling.

Following each cell without error requires both fine-tuned image analysis algorithms to accurately identify all nuclei and their positions, as well as a careful balance between limiting phototoxicity and increasing temporal and spatial imaging resolution. Similar to the embryo systems discussed above, organoids have extended 3D structures that lead to out-of-focus light and scattering, resulting in decreased resolution. In addition, cell nuclei tend to be more closely packed compared to early embryos and, particularly in epithelia, nuclei move rapidly along the apicobasal axis during division [27]. These properties require comparatively fast imaging (one 3D image every 5-15 min) at high spatial resolution (better than 1 μm/pixel), and generally complicates identification of all nuclei in each frame, for instance using gaussian fitting, and linking them through time without error. Light-sheet imaging may be used [21, 28], which can limit resulting phototoxicity. This technique has also been used to study the flow of embryonic renal cells in kidney organoids during kidney rudiment re-aggregation [29]. However, more broadly available confocal imaging is often sufficient for organoid time-lapse imaging studies.

Apart from imaging, the dense 3D tissues found in organoids also pose challenges for nuclei identification using established image analysis approaches, such as Gaussian mixture models. An important recent advance in this regard is the use of neural networks and machine learning. This approach, which is based on a training procedure that uses manually analyzed datasets to learn to identify nuclei, was shown to improve performance in non-organoid systems with closely packed cells [16, 30]. Another issue is that cell tracking software can be difficult to use for non-experts [31]. Cell trackers often need to be reprogrammed, reconfigured, or retrained in the case of a neural network approach, upon changes in the system studied, imaging parameters, or fluorescent reporters, although algorithms that work for a wider range of microscopy images are developed [32]. Finally, an important practical problem is that software packages are not always well suited to correct for the tracking errors they invariably generate. This feature is less important when studying properties such as cell flow, but is important for lineage analysis in organoids, where differentiation events may be strongly influenced by stochasticity or neighbor interactions.

As a consequence, manual cell tracking approaches are still used, even for systems with hundreds of cells [33]. noids is to incorporate algorithms that autoA promising new direction for organoid systems is to combine automated tracking based on neural networks with manual error correction steps. Such an approach was used to reconstruct cell lineages by tracking 50 cells during embryonic brain regionalization in brain organoids [34]. Key to scaling up such a hybrid approach from a limited number of lineages to entire orgamatically identify possible errors and allow for efficient manual correction of these errors. Recently, we developed such a hybrid approach to perform lineage tracking for whole intestinal organoids, which we will present in the next chapter.

In the future, we expect that automatic cell tracking approaches will continue to improve, driven in part by advances in machine learning methods [22]. Currently cell tracking studies focus primarily on cellular movement and divisions. With automatic cell tracking becoming more accessible, a range of new applications will open up in organoid research, including the study of cellular differentiation, tissue renewal, shape and symmetry changes, and may involve simultaneous measurements of key regulatory and metabolic signals [35, 36].

2 | OrganoidTracker: efficient cell tracking using machine learning and manual error correction

Published as:
Kok, R.N.U., Hebert, L., Huelsz-Prince, G., Goos, Y.J., Zheng, X., Bozek, K., Stephens, G.J., Tans, S.J., & Van Zon, J.S. (2020). *PLOS One*, 15(10), e0240802.

Abstract

Time-lapse microscopy is routinely used to follow cells within organoids, allowing direct study of division and differentiation patterns. There is an increasing interest in cell tracking in organoids, which makes it possible to study their growth and homeostasis at the single-cell level. As tracking these cells by hand is prohibitively time consuming, automation using a computer program is required. Unfortunately, organoids have a high cell density and fast cell movement, which makes automated cell tracking difficult. In this work, a semi-automated cell tracker has been developed. To detect the nuclei, we use a machine learning approach based on a convolutional neural network. To form cell trajectories, we link detections at different time points together using a min-cost flow solver. The tracker raises warnings for situations with likely errors. Rapid changes in nucleus volume and position are reported for manual review, as well as cases where nuclei divide, appear and disappear. When the warning system is adjusted such that virtually error-free lineage trees can be obtained, still less than 2% of all detected nuclei positions are marked for manual analysis. This provides an enormous speed boost over manual cell tracking, while still providing tracking data of the same quality as manual tracking.

2.1 | Introduction

Since the 1960's, researchers have used algorithms to automatically detect cells from microscopy images [37]. Traditionally, automated cell tracking algorithms have been rule-based, i.e. they relied on sets of discrete rules to detect cells. Such approaches have been successfully employed to automatically detect and track cells in highly com-

plex, multicellular systems, such as early fruit fly, zebrafish and mouse embryos [23]. However, organoids often represent adult (epithelial) tissues and therefore pose specific challenges for automated cell tracking using rule-based approaches, compared to early embryogenesis. First, during cell divisions in adult epithelia, cell nuclei rapidly move from the basal to the apical side of the cell, and back again after the division [27]. This rapid nuclear movement is challenging as cell trackers typically rely on detection of fluorescently labeled cell nuclei for tracking. Second, in epithelia nuclei are closely-packed, with microscopy images often showing overlap in the fluorescence signal between adjacent nuclei due to limits in optical resolution. Finally, in adult epithelia cells rapidly turn over, which makes cell death a common occurrence [20]. The resulting cell debris has an visual appearance similar to nuclei of live cells, and hence must be ruled out as false positives. In recent years, systematic comparison of the performance of cell trackers has shown that convolutional neural networks outperform rule-based approaches in accuracy of cell detection [16]. Convolutional neural networks have already been used successfully for microscopy cell images in 2D [38–42] as well as in 3D [43–45]. Therefore, we decided to use convolutional neural networks as a basis for developing an approach to track most, if not all individual cells in time-lapse movies of growing organoids.

A key challenge for using convolutional neural networks is obtaining enough annotated microscopy data for training. Training data is typically generated by hand initially, and then often augmented by rotating, rescaling and varying the brightness and contrast of the images [43], or even by using neural networks that generate synthetic microscopy images to supplement the training data [44]. To detect fluorescently-labelled nuclei in time-lapse movies, two different approaches have emerged. In the first approach, neural networks are trained to segment individual nuclei [46]. This has the advantage that it provides information not only on cell position, but also nuclear shape and volume, that can be used to identify the same nucleus in subsequent frames. However, this comes with the disadvantage that creating manually segmented training data is very challenging. In the second approach, neural networks are trained to only predict the center locations of each nucleus [39, 45, 47]. While this provides little information on nuclear shape, generating training data is significantly easier, as only a single point needs to be annotated for each nucleus. Here, we follow the latter approach, allowing us to make use of an extensive collection of manually annotated imaging data generated by us. We combine it with the technique of providing the image input coordinates to the neural network [48], allowing for optimized cell detection for various depths within the image stack.

To link cell positions generated by the convolutional neural network into cell trajectories, and to correctly keep track of cell identities during cell divisions, we use a min-cost flow solver [49] that we further optimized for detection of cell divisions in organoids. Tracking all cells and their divisions in entire organoids raises the possibility of directly measuring cell lineages, e.g. the division patterns by which adult stem cells generate differentiated cells. To achieve this, cell detection and linking needs to take place with high accuracy, as a single switch in cell identities can have a profound impact on the resulting cell lineage. For that reason, we followed to approach in Refs. [50–52] and implemented tools for manual curation of the automatic tracking data to eliminate such mistakes, with the software flagging tracking data that is likely

erroneous for subsequent examination by the user.

We tested the ability of this approach to perform automated tracking of all cells in intestinal organoids, which are organoids that model the small intestinal epithelium [20]. The intestinal epithelium rapidly renews itself; in mammalians every five days the stem cells replace almost all of the differentiated cells. This results in a flow of cells from the stem cell compartment (crypt) to the compartment where fully differentiated cells die (villus). Intestinal organoids contain all cell types from the intestinal epithelium [53], making them a highly valuable model system for intestinal biology. Using our approach we could obtain tracking data of the same quality as the manually annotated training data, but for entire organoids and at vastly higher speed. Using the tracked data, we could for the first time establish complex cell lineages of up to five generations and visualize spatial patterns of cell division and cell death over the entire organoid. Such data will allow tissue-level study of cell divisions, movement and death at high time resolution, and will likely give new insight in how and when cells make their cell fate decisions.

2.2 | Results

Overview of cell tracking approach

We are interested in tracking cells in time lapse videos of growing intestinal organoids. In the videos, we observe the nuclei of the cells which was made fluorescent using H2B-mCherry. The videos are up to 65 hours in length with the time resolution set to 12 minutes, resulting in videos of 84 to 325 time points. Each pixel is 0.32 x 0.32 µm in size, and the separation between two z-layers is 2 µm. As can be seen in **Fig. 1**, the images have irregular fluorescence intensity, which is not only caused by intrinsic variations of nucleus fluorescence, but also by the location of the cells: some parts of the organoids are oversaturated, while parts deeper inside the organoids are undersaturated. Especially in the oversaturated parts of the images, the nuclei can appear to visually overlap. In addition, the organoids exhibit high nuclear density, with large variation in the shape, orientation and intensity of nuclei. Moreover, the requirement to limit phototoxicity results in low-contrast images. Finally, images suffer from low resolution in the z-direction. All of these factors make it challenging to automatically recognize nuclei.

The imaged area of the organoid typically contains over 200 cells, which makes manual tracking unfeasible. Instead, we had previously obtained (to be submitted) manual tracking data for a part of the imaged area, namely the crypt region. For that region, the center position of about 70 nuclei was recorded manually. The center position of each nucleus was gauged by eye and may therefore be a few pixels off. Additionally, links were created between the same nuclei in different time points, so that the nuclei could be tracked over time. In total, manual tracking was done for 2826 time points, coming from 17 different organoids.

In our cell tracking approach, we use a convolutional neural network for cell detection, a min-cost flow solver [49] for linking cells over time and a small set of rules to detect

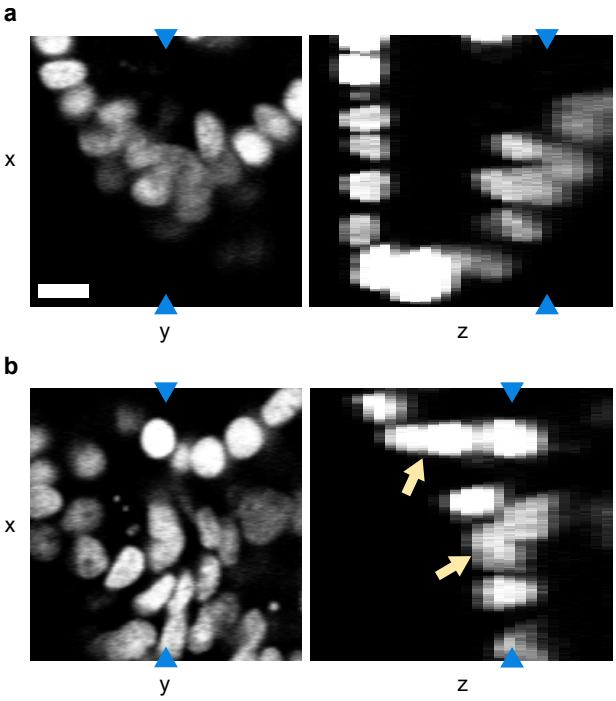


Figure 1 • Example of a tracked organoid

Microscopy image slices in the xy and xz planes using H2B-mCherry to visualize the cell nuclei. Blue arrows indicate that both image slices in the same panel represent the same pixels. Note that the resolution in the z-direction is lower. The images show that there is both variation in intensity and visual overlap in the nuclei (indicated by the orange arrows), especially in the z-direction. The scale bar is 10 µm.

improbable cell tracking data that the user needs to verify and correct. This complete approach is displayed schematically in **Fig. 2**. Step 1 is formed by the convolutional neural network: using raw images, it outputs images with peaks at the center location of the nuclei. This network needs to be trained using ground-truth data, which is done using the manual tracking data. Then, in Step 2 the cells are linked using a min-cost flow solver that can handle cell divisions [49], which optimizes for the smallest cell movement in between time points, for the smallest changes in nucleus volume in between time points and for the longest cell tracks. Finally, in Step 3 the program reports all cell tracks that it sees as suspicious based on a small set of rules, which the user needs to inspect. For us, this workflow resulted in cell tracks of the same quality as manual tracking though obtained vastly faster.

Step 1: automated cell detection

The nucleus center positions from our manually annotated tracking dataset were used to train a fully convolutional neural network, where the input and output are 3D images of the same shape. We trained the network with 3D microscopy images as the input images. Here, each image represents a single time point. The network architecture follows a standard U-Net [46] type architecture with all the convolutional layers replaced by the CoordConv layer [48] adapted for 3D images (**Fig. 3a**). The CoordConv layer concatenates the absolute location of pixels to each convolution input. Compared to the simple convolution layer which is translation equivariant, this layer is useful for cases where the absolute position of the pixels is important to the task.

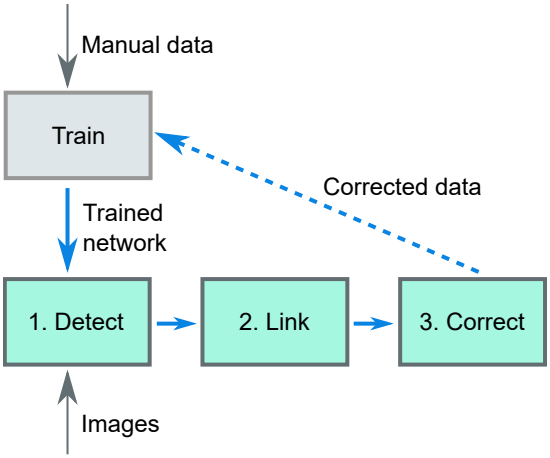


Figure 2 • Overview of the tracking software
Using ground-truth data of nucleus locations in microscopy images (obtained from manual detection) a convolutional neural network is trained. This trained network is then used to detect (step 1) cells in new images. The detections are then linked over time (step 2), after which the user manually corrects the output with the help of an error checker (step 3). In principle, the manually corrected data can be used as additional ground-truth data to improve the training of the convolutional neural network, leading to an efficient, iterative improvement in performance of the convolutional neural network (dashed line).

For the output images of the network, we want images where the brightness of each pixel corresponds to the chance of that pixel being a nucleus center. As multiple pixels close together can have a high likeliness of being a nucleus center, this results in mostly dark output images with bright spots at the locations of the nucleus centers. To make the neural network produce such output images given microscopy images at the input images, it needed to be trained on pairs of microscopy images and artificially generated output images. The generated output images are black images that contain bright, Gaussian-shaped spots with standard deviations of 2 px at the locations of the nuclei (**Fig. 3b**). The locations of the nucleus centers were determined from the manual tracking data. The Gaussian shape of the spots acknowledges the fact that there is an uncertainty in the exact location of the nucleus center. Because the resolution in the z axis of the images is lower than in the x and y axes, the Gaussians for the labeled volumes are in 2D and don't cross several xy-planes.

We trained the network on 9 organoids, containing 1388 time points in total. For training, columns of 64x64 px from random locations within the images were used. This made the training algorithm use less memory compared to using the full 512x512 px, and also randomized the x and y coordinates of the nuclei, but unfortunately it does prevent the network from using information in the images that is located further away from the nucleus. To make the training more robust, random perturbations were applied to the training images [46]. The images were randomly flipped in the x, y and/or z

direction, rotated a random number of degrees (0-360), scaled to 80%-120% of their original size, made brighter or darker with at most 10 % of the maximum intensity in the image and finally the contrast was adjusted according to $I_{new} = (I - \langle I \rangle) \cdot c + \langle I \rangle$ where I_{new} is the resulting intensity of a pixel, I the original intensity of that pixel, $\langle I \rangle$ the mean intensity of all pixels in the image and c the contrast factor, which varied from 0.5 to 1.5. We use a weighted mean squared error as the loss function between the network output and the labeled volume. Because the labeled volumes were mostly composed of zeroes, we gave more importance to the Gaussian spots by applying weights that correspond to the percentage of non zero values in the labeled volume.

Once the network was trained, it generated output images that show where the nucleus centers are located (**Fig. 3c**). Each pixel in the 3D image represents the probability of that pixel being the nucleus center, resulting in a probability distribution with small peaks at the location of the nucleus centers. We interpolated linearly the empty space between the slices so that the resulting volume had the same resolution in the z axis as in x and y. This allows us to apply a 3D peak detection algorithm (`peak_local_max` in scikit-image 1.1.0 [54]) to detect these local maxima in the interpolated 3D volumes. The resulting 3D coordinates are considered to be the locations of the nucleus centers in the full 3D volume. We then map back these coordinates to the closest image slice.

To evaluate the performance of the network, we needed to know how many of the detections are true positives or false positives, and how many false negatives there are. To do this, we compared the automatic tracking data to manual tracking data of 8 organoids (1438 time points) that were not used for training the neural network. Because these images are from separate organoids, we can use this tracking data to evaluate the model generalization. One challenge in the performance evaluation was that it is difficult to measure the number of false positives from the neural network, as only 30% to 40% of all cells visible in the images were tracked. Therefore, at any location where the neural network reports the presence of a nucleus while the manual annotations do not, we cannot a priori be sure whether there is a false positive or whether that part of the image was simply not manually annotated. To overcome, we used the following approach.

Any nucleus center detected by the neural network was assigned to the closest nucleus center from the manually tracking data, under the condition that the distance was no longer than 5 µm. Every nucleus center cannot have more than one assignment. Each successful assignment was a true positive. Then, any manually tracked nucleus center that was left with no assignments became a false negative. Finally, any nucleus center from the neural network that was left with no assignments was regarded as a false positive if it was within 5 µm from a manually tracked nucleus center, otherwise it was rejected. This ensured that misdetections within the manually tracked area were still detected.

We measured three values to quantify the performance of the network: the precision, recall and the F_1 score. The precision is the fraction of the nucleus detections done by the neural network that were correct. Mathematically, it is defined as

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.1}$$

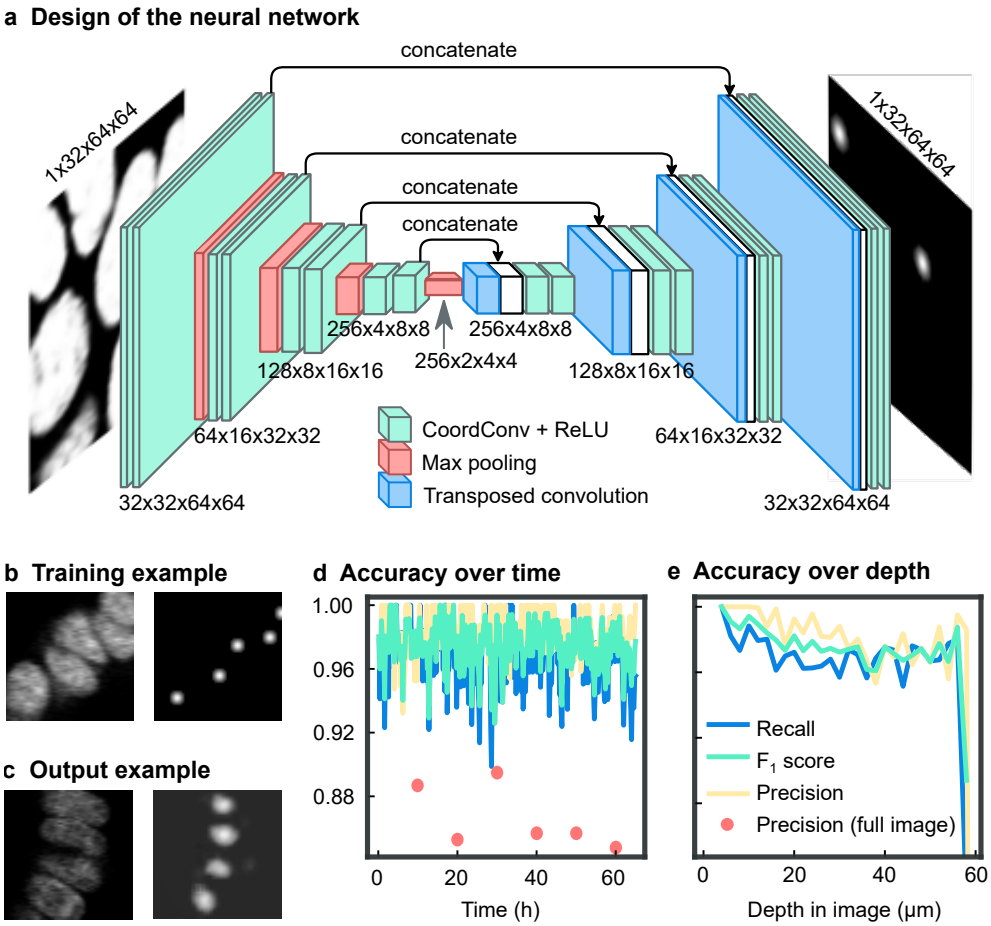


Figure 3 • Automated detection of cell nuclei by a convolutional neural network
(a) Schematic overview of the network. The network is a standard convolutional network with absolute pixel coordinates added as input values. This makes it possible to adjust its detection network both for cells deep inside the organoid and cells near the objective. **(b)** Example of training data. On the left a slice of a 3D input image, on the right Gaussian predictions created from manual tracking data. The brightness of each pixel represents the likeliness of that pixel being the nucleus center. **(c)** Example of a network prediction, which shows the probability of each pixel being a nucleus center, according to the neural network. **(d)** Accuracy of the neural network over time compared to manual tracking data obtained in a subsection of a single organoid. For six time points (annotated with red circles) tracking data was compared to manually annotated data for an entire organoid. In this case, we observed a lower precision. For this organoid, the recall is 0.97 while the precision is 0.98. **(e)** Same data as in Panel d, now plotted over z. The accuracy of the network drops at the deepest imaged parts, likely because part of each nucleus falls outside the image stack here.

with TP the number of true positives and FP the number of false positives. The recall is the fraction of nucleus detections in the ground truth that was detected by the neural network, and it is defined as

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.2}$$

with FN the number of false negatives. For a neural network it is important that both the precision and the recall are high. In one extreme example, if a neural network would mark almost every pixel as a nucleus center, it would have a high recall, but a precision of almost zero. In the other extreme, if the neural network would only find back the most obvious nuclei, it would have a high precision but a low recall. To express the quality of a neural network as a single number, the harmonic average of the precision and recall was calculated, which is known as the F_1 score:

$$\text{F}_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{2.3}$$

The precision turned out to be 0.98 with a standard deviation of 0.007, which means that 98% of the detections done by the neural network were consistent with the manual data. The recall turned out to be on average 0.96 with a standard deviation of 0.02, which means that 96% of all manually tracked cells were detected by the neural network. Based on these values for the precision and the recall, the F_1 score of the network is 0.97.

Fig. 3d displays the accuracy over time, which shows that the accuracy did not degrade over time, when the organoids have grown larger. As we can see in **Fig. 3e**, the accuracy is the highest close to the objective and then drops before levelling off from 20 µm. In the deepest part of the organoid, the accuracy drops further. This is because the full nucleus is no longer visible; part of the nucleus now falls outside of the image stack.

The manually tracked region may not be representative for the whole organoid as there are also untracked regions which mostly consist of fluorescent material coming from dead cells. To check the performance of the network for the organoid as a whole, for six time points all nuclei were manually annotated in the entire image stack. This resulted in 1318 nuclei being annotated. The recall for these images was still 0.96, but the precisions, shown in **Fig. 3d** as red dots, decreased to 0.87. By observation of the locations of false positives, we observed that this was mainly because dead cell material was recognized by the neural network as if it were a live cell nucleus. Together, this indicates that the recall is approximately consistent for the entire image, but that the precision is lower in regions with more dead cell material. To obtain correct cell tracking data, any spurious annotations of dead cell material will need to be filtered out later in the tracking process. This could be done either during the automatic linking step or during the manual error correction step.

In addition to calculating the F_1 score of the neural network for our organoids, we also calculated the detection accuracy (DET) score. The DET score is another measure for the accuracy of identifying cells [55]. The software available to calculate DET scores requires cell segmentation masks, not just cell positions as is the case for our

cell tracker [16]. Therefore, we drew a pseudo-mask of the nuclei, which was a sphere with a of 5 µm around each detected position. In closely-packed regions of the organoid, this sphere is larger than the nucleus. Therefore, multiple spheres can overlap. If this happens, each pixel in the overlapping region of multiple spheres was assigned to the closest detected position. To make sure that detected positions outside the manually tracked area were not regarded as false positives, we removed all positions outside the manually tracked region. The approach was the same as for the F_1 score: any detected nucleus center more than 5,µm away from a nucleus center in the manual tracking data was deleted. The DET score, calculated in this manner for the 8 organoids, was 0.93 ± 0.03, with a score of 1.00 corresponding to error-free detection. These scores are in the high range of scores obtained by other trackers for developmental data sets of similar complexity [16]. Overall, these results indicate that our approach can accurately detect most cell nuclei in our organoids. The next step is to link the nucleus detections at different time points together, such that nuclei can be followed over time.

Step 2: linking of cell positions

To follow cells over time, the locations of the same cell nucleus at different time points need to be linked together. A link goes from a detected nucleus center at time point t to the same nucleus center imaged at time point $t+1$. Normally, every nucleus has one link to the next time point and one link to the previous time point. However, in case of a division a nucleus will split into two nuclei and therefore the nucleus will also have two links to the next time point. A straightforward way to create these links is to always assume that the nearest detected nucleus in the previous time point represents the same nucleus; this is called nearest neighbor linking. By going back in time, in theory we get detection of cell divisions for free: if two nuclei at time point $t+1$ both have the same, single nucleus at time point t as their closest nucleus, a division is generated.

Unfortunately, nearest-neighbor linking does not provide us with accurate lineage trees. We can see in **Fig. 4a** that nearest neighbor linking creates unrealistically short cell cycles. Moreover, although rare, there is nothing that prevents a mother cell from having three or more daughters. Clearly, the assumption of the nearest nucleus always being the same nucleus does not hold. For this reason it is necessary to let the linking algorithm also consider links to nuclei further away than the nearest neighbor. In our method we consider all links that are at most two times further away in distance than the nearest neighbor. For most detected cell positions, this results in multiple linking options for creating links (**Fig. 4c**). Each of those options is given a score (an energy), of which the sum of all created links is minimized. This minimization is carried out using a min-cost flow solver adapted for tracking of dividing cells, developed by Haubold *et al.* [49]. Here, we used the following rules:

1. *An ordinary link is scored as score* $= \Delta X + \sqrt[3]{\Delta V}$*, with* ΔX *the change in position in µm, and* ΔV *the change in volume of the nucleus in µm*^{3} This ensures that links with the lowest change in position and volume are favored.

2. *Every track end is given a penalty of 1.* This favors creating longer tracks over multiple shorter tracks, which allows for the occasional long-distance link.

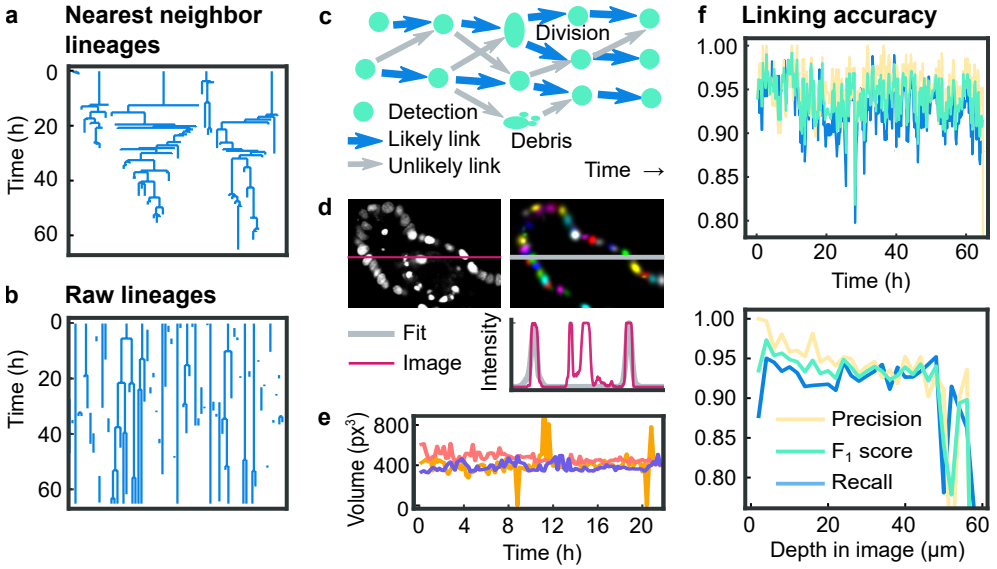


Figure 4 • Overview and results of the linking algorithm
(a) Lineage trees obtained by linking using the nearest-neighbor method. **(b)** Raw lineage trees obtained by our linking algorithm. **(c)** Example of a network of links. There are many possibilities to link the cell detections in different time points together. Of all possible links (displayed as arrows), the most likely ones (displayed as blue arrows) are selected using a scoring system. **(d)** Original microscopy image, Gaussian fit and intensity profile of both. The intensity profile runs along the marked row in the image. We have chosen a section that highlights the fact that our approach can select nuclei and ignore cell debris, even though both have a similar fluorescence intensity profile. **(e)** Nucleus volume of three nuclei extracted from a Gaussian fit. The volume ($V = \mathrm{Cov}(x,x) \cdot \mathrm{Cov}(y,y) \cdot \mathrm{Cov}(z,z)$) is calculated for three non-dividing cells. The outliers are caused by errors in the Gaussian fit. **(f)** Accuracy of the linking algorithm presented here for a single organoid, over time and over the image depth. Note that the precision is higher than for nearest-neighbor linking.

3. *Every track start is given a penalty of 1.* Again, this favors the creation of longer tracks.

4. *Every cell position that is ignored by the linking algorithm results in a penalty of 1.* This ensures that as much as possible of the data from the neural network is linked. If this rule would not exist, a globally optimal score could be reached by not creating any links at all.

5. *Cell divisions have a scoring system based on the change in volume and intensity*. This scoring system is described in . A score below 1 means that a cell division here is unlikely, in which case no division is allowed here.

6. *If the cell division score is at least 1, the score from rule 1 is halved.* In the time around a cell divisions, the cell nuclei are expected to quickly move and change their volume, so it is appropriate to lower the penalty for position and volume changes.

The scores resulting from rules 1 to 5 are multiplied by a factor specified by the user of the software. Here, these factors are 20, 100, 150, 150 and 30, respectively. By changing these values, the scoring system can be tuned. For example, if the neural network accidentally detects dead fluorescent material as living cells, resulting in many short paths, this can be fixed by increasing the penalty for appearances and disappearances of cells. In this way, the high penalties at the beginning and end of the track together are larger than the penalty of ignoring those detected nucleus positions. As a result, the track is not created at all. To give another example, if the cells in a system tend to move slowly, the movement penalty can be increased, so that the likeliness of creating long-distance links is decreased. We used this in our system, to avoid creating a link to a neighbor nucleus when the correct nucleus was not detected by the neural network. For us this works well, because a cell track that abruptly ends will easily be found if the user rechecks all track ends, while a switch in cell identity resulting from a link to another cell is harder to detect.

For evaluating rule 1 and 5 it is necessary to know the volume of the nuclei. This volume is estimated using a 3D multivariate Gaussian fit. While nuclei do not always have a Gaussian-like shape, they can still be approximated by a Gaussian function in order to get an estimate of their volume [56]. Because we only want to fit Gaussian functions to living nuclei and not to fluorescent material coming from dead cells, we cannot fit the image as a whole. Instead, we need to separate dead cell material from the living nuclei, and only fit Gaussian functions to the latter. This is done by first performing a segmentation to divide the image into clusters, each containing zero, one or multiple cell positions detected by the neural network. The segmentation starts by performing an adaptive threshold on the original image I_{image} to separate the foreground and background, resulting in a mask named I_{mask}. This mask is then improved by first removing all pixels from I_{mask} that in the original image I_{image} have a negative iso-intensity Gaussian curvature [56], which means that all pixels of lower intensity in between two places with higher intensity are removed. This is expected to remove the pixels in between two nuclei. Next, all holes in the mask are filled, resulting in the mask named $I_{mask,filled}$. On this mask, three rounds of erosion are performed with a 3x3x1 cross-shaped kernel on the mask. This step makes the area of foreground

smaller, which makes the foreground fall apart into separate clusters. For that reason, this eroded mask was called $I_{mask,eroded}$. From the convolutional neural network, we know how many cells there are in each cluster and where their nucleus centers are. A Gaussian fit is then performed on $I_{mask,filled}$, which was the mask before erosion was applied. One Gaussian function is fit for each nucleus position and each Gaussian function is centered around the position given by the neural network, using a default intensity and a default covariance matrix. The following function is used to fit a single nucleus:

$$G\left(\vec{x}\right) = a\, e^{\left(\vec{x}-\vec{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\vec{x}-\vec{\mu}\right)} \tag{2.4}$$

Here, a is the maximum intensity of the function, μ is a vector representing the mean and $\boldsymbol{\Sigma}$ is the covariance matrix. Together, these three variables describe the intensity, the position and the shape of the cell nucleus, respectively. $\left(\vec{x}-\vec{\mu}\right)^T$ denotes that we are taking the transverse of $\left(\vec{x}-\vec{\mu}\right)$, which results in a row vector. These functions are fitted to a blurred version of the original image, $I_{image,blurred}(\vec{x})$, so that the nuclei have a more Gaussian-like shape. For all pixels $\vec{x}$ in the cluster, the quadratic difference between that image and the outcome of the Gaussian functions describing the nuclei is minimized:

$$\text{difference} = \sum_{\vec{x}} \left(I_{image,blurred}\left(\vec{x}\right) - \sum_i G_i\left(\vec{x}\right) \right)^2 \tag{2.5}$$

In **Fig. 4d**, we can see an example of the Gaussian fit. The panel shows that not only the Gaussian fit correctly follows the intensity profile of the cells, but also that the fit ignores the debris of dead cells, which was previously filtered out by the neural network. To show the consistency of the Gaussian fit over time, we have plotted a volume ($\text{Cov}(x,x) \cdot \text{Cov}(y,y) \cdot \text{Cov}(z,z)$) over time for three arbitrary cells in **Fig. 4e**. As can be seen in that figure, the Gaussian fit predicts a volume over time without significant noise. However, sometimes the fit predicts a volume more than twice the median volume, and sometimes only predicts a very low volume. This can happen if the fit of two nuclei results in one large Gaussian shape that covers both nuclei, and one very small Gaussian shape.

In **Fig. 4b** the raw lineage trees obtained from the linking algorithm are displayed. By comparing both these lineage trees and lineage trees obtained from nearest-neighbor linking (**Fig. 4a**) to lineage trees from manual tracking data (**Fig. S1**) it is clear that the method proposed here is more accurate than nearest-neighbor linking. To quantitatively measure the accuracy of our method, we need to calculate the precision (**Eq. 2.1**) and recall (**Eq. 2.2**). For this it is necessary to define true positives, false positives and false negatives in terms of links. We considered all detected nuclei at most 5 µm from the location of the manually annotated nucleus. If any of these nearby nuclei centers at time point t had a link to any of the nearby nuclei centers at time point $t+1$, this was considered to be the same link and hence a true positive. If no link existed between the nearby positions, this was considered to be a false negative. Naturally, every link in the manual tracking data can only be assigned to one corresponding link

the automatic tracking data. Every remaining link in the automatic tracking data with no corresponding link in the manual tracking data was therefore in principle a false positive. However, as not the whole organoid was tracked, links were rejected if, at both time points, they did not have any nucleus centers in the manual tracking data within 5 µm.

Using this approach, we calculated the precision, recall and F_1 score for the same 8 organoids (1438 time points) as used for evaluating the detection performance. The recall of all organoids is 0.91 with a standard deviation of 0.04, the precision is 0.94 with a standard deviation of 0.02 and the F_1 score is therefore 0.93. Note here that the recall score is limited by the recall of the neural network: if the neural network fails to detect a nucleus, the two links of that nucleus will automatically also not be created. As can be seen in **Fig. 4f**, the results are relatively constant over time, but not over depth: the accuracy is worse near the edge of the imaged area. In addition, we calculated the tracking accuracy (TRA) score, a different measure of how well cells are tracked in time [16, 55], using the approach outlined for calculation of the DET score. We obtained a TRA score of 0.92 ± 0.04 averaged over 8 different organoids, with a score of 1.00 corresponding to error-free tracks. Like the DET score, these TRA score are in the high range of scores obtained by other trackers for developmental data sets of similar complexity [16]. Overall, our analysis shows that our automated cell detection and tracking approach yields cell tracking data of high quality, although not yet sufficient for error-free tracking. Hence, a subsequent step of manual data correction is required to achieve this.

Step 3: manual error corrections

While the lineages trees obtained from the previous step already show promising results, still the lineage trees are not correct: divisions are missed, cell are switched in identity, cell tracks end prematurely and cell divisions are created where there are none. Based on these results, manual error correction is needed to obtain reliable tracking data. Time-wise, it would not be practical to revisit each nucleus at every time point. For this reason, the program flags all nuclei that violate one the following rules, so that the user can manually make corrections:

1. *Cells must have a link to the previous and to the next time point.* This detects cells that are appearing out of nowhere and cells disappearing into nothing. These cases usually indicate that a cell was not detected at a time point, or that a cell was detected where there was none. In case the cell actually dies, or if the cell goes out of the microscope view, the user can ignore this rule violation.

2. *At least 10 hours must have passed before a cell divides again.* A violation of this rule indicates that a cell identity switch has happened. The exact number of hours can be changed.

3. *Cells must not have moved more than 10 µm in between two time points.* Such a quick movement can point to a switch in cell identity. The exact number of micrometers can be changed.

4. *The nucleus volume of a dividing cell must be larger than the volume of the daughter cells combined.* This rule helps identifying misdetected divisions.

5. *The volume of a nucleus may not shrink more than 3 times.* This rule helps finding switches in cell identity and missed cell divisions.

6. *Two cells may not merge together, and cells may not have more than two daughters.* Even though linking does not allow these situations, this rule can be violated if a person makes a mistake while manually correcting the tracking data.

Any violations of the above rules are automatically flagged by an error checker that runs as the last step of the linking process. The program guides the user through the list of errors, prompting the user to either suppress them, or to manually correct the tracking data. While correcting the tracking data, the error checker immediately checks every edit done by the user. As a result, if the user accidentally creates a cell division with three daughter cells, the user will immediately see a cross appearing on the mother cell. The strictness of the rules can be varied, with stricter rules leading to more false warnings, but a lower risk for missing tracking errors. When we adjusted the parameters such that we got tracking data of the same quality as manual tracking data, 1 to 2 % of all cell detections needed to be manually verified.

To evaluate the efficiency and performance of the manual data correction step based on the above rules, we compared the output of our cell tracker before and after manual data correction against manually annotated data in our training data set for two organoids. Before data correction, the recall was 0.93 ± 0.02 and the precision 0.95 ± 0.01. After data correction, the first organoid produced identical lineage trees compared to the manual tracking data, while the other organoid contained one tracking mistake: the wrong cell was declared dead. To evaluate how many of the warnings were false positives, we quantified what percentage of these errors led to a change in the tracking data. If the annotated nucleus position was moved or deleted, or if there was a change in the locations of the nuclei it was linked to, the warning was considered to be a true positive. For both organoids, we found that 80% of the warnings were true positives.

Comparison to existing methods

Currently, other methods exist that combine automated tracking with manual correction, albeit without using neural networks, with one example being TrackMate [51]. Using such a rule-based approach to cell detection, rather than using neural networks, has the advantage that no large set of training data is required. As the creation of such datasets is often a time-consuming and laborious process, one could wonder if our effort to create such a training dataset was justified here. How well would TrackMate perform when applied on our images?

To test this, we used the so-called Downsampled Laplacian of Gaussian detector of TrackMate, which is the only detector suitable for spots larger than 20 pixels, according to its documentation. The detector has three parameters, the spot size, the downsampling factor and the detection threshold. By manual optimization, we arrived at a spot size of 10 µm, a downsampling factor of 2 and a threshold of 1. Like for our neural

network, we measured the performance by only considering the manually tracked region of the organoid. The recall was on average 0.66, with a standard deviation of 0.04 across the organoids. This means that in our hands, TrackMate missed a third of all cells in the image. For comparison, the neural network only missed 4% of all cells. The precision was higher: 0.969 with a standard deviation of 0.006, which is almost as high as the precision of the neural network (0.98). This indicates that TrackMate is conservative at recognizing nuclei, but when it does recognize a nucleus, it is almost always correct. However, in total this results in an F_1 score of only 0.79, which is considerably lower than the score of the neural network, which was 0.97.

Even though one third of the nuclei were not detected, we still made an attempt at linking using the LAP tracker of TrackMate, using its default settings. For the 8 testing organoids together, the recall was 0.55 with a standard deviation of 0.04 and the precision was 0.79 with a standard deviation of 0.04. This results in an F_1 score of 0.65. This result means that almost half of the correct links are already missing. Extensive manual correction would be required to establish cell tracks at similar quality as we produced with our approach. Overall, these results underscore to importance of efficient and accurate cell detection, as we achieved here using neural networks.

Analysis of tracking data

In **Fig. 5** we show the results of this whole-organoid tracking approach. First, this approach makes it possible to connect lineage dynamics to spatial position of cells in an organoid. In **Fig. 5a** 2D slices of microscopy images of an organoid are displayed, with its 3D reconstruction shown in **Fig. 5b**. In the reconstruction, the cells have been colored by lineage: selected cells were colored based on which cell they originated from in the first time point. Videos of both panels are included as Video S1 and Video S2. In **Fig. 5c** lineage trees are shown of the same organoid and using the same coloring. We can see that cells of the same lineage tree tend to stay connected in space, although cells of multiple lineage trees do intermingle. In addition, we can see that the largest lineage trees reside at the lowest position in the crypt region, which is the growing protruded region in **Fig. 5b**.

Next, we used our tracking data to study dynamics of cell divisions in space and time. We observed that almost all cell divisions are happening in the crypt (Video S3), which is consistent with the literature [57]. We could see that at the bottom of the crypt, the number of divisions is the largest. However, there were a few cells at the bottom of the crypt that, in sharp contrast to their neighbors, did not divide, or divided only once. These are likely to be Paneth cells, which are differentiated cells that reside in the stem cell compartment [58]. In **Fig. 5d** we saw that divisions were initially synchronized, while later in time this synchronization was less apparent.

Software

Source code and software packages are available at `https://github.com/jvzonlab/OrganoidTracker`. Documentation including tutorials is available at the same ad-

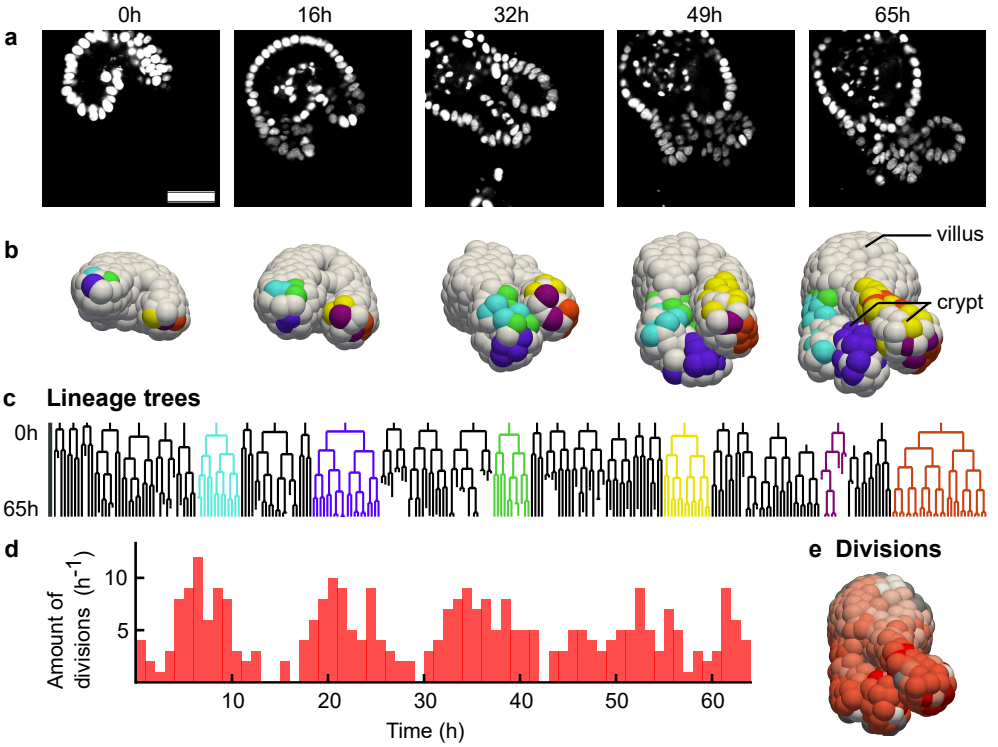


Figure 5 • Analysis of a fully tracked organoid

(a) Single microscopy image slices of an intestinal organoid from a time lapse movie of 65 hours, in steps of 16.25 hours. H2B-mCherry was used to visualize the nuclei. Scale bar is 40 µm. **(b)** Digital reconstruction of the same organoid as in Panel a. Cells of the same color originate from the same cell at the beginning of the experiment. **(c)** Selected lineage trees. The colors match the colors in the reconstruction from Panel b. **(d)** Number of cellular divisions over time in a single organoid. **(e)** Number of times each cell divided during the time-lapse integrated over time, from 0 (white) to 5 (red). For gray cells, we could not count the amount of divisions because the cell was outside the image at an earlier point in time.

dress. The trained model used for this article has been made available as File S1; this model is suitable for 3D confocal images of cell nuclei, with a resolution of approximately 0.32 µm/px, z-separation of 2 µm and low background noise. Another model is available for trained for the lower resolution of 0.40 µm/px, still with the same z-separation and noise tolerance.

The released software is capable of manual tracking, training the neural network, using the neural network to detect nuclei in microscopy images, linking nucleus detections from different time points and assisting with the correction of automated tracking data. To correct the warnings, one would normally use an interface to jump from warning to warning. Instead, one can also jump from warning to warning within a single lineage tree. Additional tools are available to highlight all cells within a (sub)lineage, or to jump from a point in a lineage tree to the cell tracking data.

The software was written in Python 3.6, making use of the established software packages Mahotas 1.4.4 [59], matplotlib 2.2.2 [60], numpy 1.15.1 [61], OpenCV 3.4.1 [62], PyQT 5.9.2 [63] and Tensorflow 1.9.0 [64]. Image loading support is provided by tiffile 2020.2.16 [65] and nd2reader 3.1.0 [66]. The software saves all tracking data as JSON files.

3D visualization is possible by exporting the nucleus positions to Paraview [67], which can render those points as spheres. The software is also able to load and save files in the data format of the Cell Tracking Challenge [16], which makes it possible to exchange data with most other cell trackers. The software supports loading scripts written in Python, which can be used for data analysis. These scripts can be reloaded while the program is running, making it unnecessary to restart the program if you made a change to your analysis script.

2.3 | Discussion

We have developed a program for cell tracking based on nuclear markers inside organoids that uses a convolutional neural network for detection and a min-cost flow solver for linking cell positions. For our dataset 91% of all links in the ground truth were recognized. This is mainly because the neural network was only able to recognize 96% of the cells, and a single missing detection already means that the linking algorithm cannot reconstruct the cell track. However, the software facilitates manual error-correction so from this data cell lineages at the same quality as for manual tracking can still be obtained.

The convolutional neural network and the min-cost flow solver used for linking are both very generic: they make no assumptions on the microscopy pictures, other than that they are grayscale 3D images. In contrast, the Gaussian fit, which is used to provide an estimate of the nucleus volume and to recognize cell divisions, is quite specific to our images: it assumes bright nuclei on a dark background with not too much noise. Currently, eliminating the Gaussian fit is not possible, as the recognition of cell divisions depends on this.

As this approach uses a supervised form of machine learning, it is necessary to have a

diverse set of training data available. Although the network will already automatically generate variations of your training data by adjusting the brightness, scale and rotation, still the network can have suboptimal performance for microscope images taken with different settings, like for example images with a different contrast or resolution. To solve this, one needs to retrain the network on training data created with those settings.

It is possible to think of further improvements to the algorithm. One could start training the neural network with a smaller training dataset, which results in a cell tracker of which its output requires a large amount of manual curation to get error-free results. After this correction step is over, it is then possible to use the manually corrected data to retrain the neural network. In this way, a feedback loop is created that improves the neural network every time more tracking data becomes available (**Fig. 2**). In this paper no such feedback loop has been used: because we already had a large amount of training data beforehand, for us it was not helpful to retrain the network.

In general, we believe the way forward is to make the neural network as accurate as possible, and then give the linking algorithm the smallest possible amount of options to create links. As long as the biologically correct option is still part of this smaller set of options the linking algorithm can choose from, this approach would reduce the chance of the linking algorithm making mistakes. One possible step in this direction would be to let the neural network also classify cells: if the network could point out where the mother and daughter cells are, the linking algorithm could be restricted to only creating divisions at those locations.

In conclusion, we are now able to observe a growing organoid at the cellular level using cell tracking in microscopy videos. For this, we have developed a cell tracker using a convolutional neural network for cell detection, a min-cost flow solver [49] for linking of cell positions and a small set of rules to assist with data curation. It is our hope that cell tracking, with its coupling of lineage tree dynamics to time and space, will lead to in a better understanding of organoid growth and maintenance.

2.4 | Supplementary information

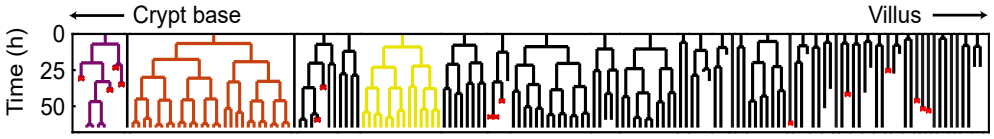


Figure S1 • Full lineages of the manual tracking data
This is the same organoid that is displayed in **Fig. 5**. Red crosses denote the positions where a cell death was observed in the microscopy images. This figure will also become part of an upcoming publication.

File S1. Trained neural network. The trained neural network used in this paper.

Download at
https://doi.org/10.1371/journal.pone.0240802.s001

Video S1. Microscopy video of a growing organoid. Pixels are colored by depth. Stills of this organoid were included in the main text (**Fig. 5a**).

Download at
https://doi.org/10.1371/journal.pone.0240802.s002

Video S2. Reconstruction of an organoid using spheres. Manually selected lineages have been colored. Stills of this video were included in the main text as **Fig. 5b**.

Download at
https://doi.org/10.1371/journal.pone.0240802.s003

Video S3. Visualization of dividing cells in an organoid. Nuclei are again drawn as spheres, but are now colored by the time to the next division. The closer the next division, the redder the sphere of a nucleus becomes.

Download at
https://doi.org/10.1371/journal.pone.0240802.s004

Appendix S1. Cell division scoring system.

The fluorescent H2B-mCherry marker labels the histones, which normally results in the entire nucleus being observed. However, during a cell division the nuclear envelope breaks down and instead we observe the compacted chromosomes. It would therefore not be correct to say that we observe the nucleus during a cell division, as there is no nucleus during certain parts of the division. However, to make the text easier to read we will still call these compacted chromosome structures ”nuclei” in the following text, instead of constantly writing ”nuclei or compacted chromosome structures”.

The scoring system takes into account both the mother cell and two daughter cells. The scoring system is designed such that lower values are more optimal; the score can therefore be considered as a penalty score or an energy. If the score is above -1, the possibility of a cell division is not even considered by the min-cost flow solver.

As stated in the main text, the linking algorithm starts by defining all possible links, which are the nearest-neighbor links along with links that span a distance at most twice as long as the distance to the nearest neighbor. Division scores are calculated for every case where a nucleus at one time point has at least two possible links to the next time point. If there are three or more possible links, then there are also three or more possible

daughters. As biologically only two daughter cells are possible, scores are calculated for all possible combinations of two daughter cells and the lowest score is chosen.

Every potential mother cell gets a score using the following equation:

$$\begin{aligned}\texttt{score} &= \texttt{mother\_intensity\_delta} + \texttt{daughter\_intensity\_difference} \\ &+ \texttt{daughter\_intensity\_delta} + \texttt{mother\_volume}\end{aligned} \tag{2.6}$$

Of these scores, `mother_volume` carries the heaviest weight. By default, it is 10 which is so high that it makes it impossible for the min-cost flow solver to create a division here. Only if the volume of the mother nucleus is almost as large as the volume of the new daughter cells combined, the score becomes 0 which allows for a division.

$$\texttt{mother\_volume} = \begin{cases} 0 & \text{if } V_{mother}/(V_{daughter1} + V_{daughter2}) > 0.95 \\ 10 & \text{otherwise} \end{cases}$$

The other rules are then used to pick the most likely division candidate, if multiple possibilities of nuclei show such a volume change. The next rule makes use of the observation that the daughter nuclei are located at a different position than the mother nucleus was. Therefore, the fluorescence intensity at the location of the mother cell before the division should have decreased after the division. For non-dividing nucleus, it is rare that nuclei move this rapidly, so non-dividing nuclei will get a lower score.

$$\texttt{mother\_intensity\_delta} = \begin{cases} -1 & \text{if } I_{current}/(I_{next} + 0.0001) > 2 \\ 0 & \text{if } 2 \geq I_{current}/(I_{next} + 0.0001) > 1.4 \\ 1 & \text{otherwise} \end{cases}$$

Here, $I_{current}$ is the average pixel intensity at the location of the mother nucleus at the last time point where a single nucleus was visible, with the intensities of the image normalized from 0 to 1. The location of the mother nucleus is derived from the Gaussian fit: we look at all pixels at the XY plane of the nucleus center where the Gaussian fit still has 20% of its maximum intensity. I_{next} is the intensity at exactly the same pixel locations in the next time point. The $+0.0001$ is used to prevent a division by zero for the rare cases where the image becomes completely black at the location where the mother location was.

For the daughter nuclei the opposite should happen: where there was no nucleus before, there now is a nucleus. The intensity should therefore go up. For a single daughter, the score is calculated as follows:

$$\texttt{daughter1\_intensity\_delta} = \begin{cases} -1 & \text{if } I_{current}/(I_{prev} + 0.0001) > 2 \\ 0 & \text{if } 2 \geq I_{current}/(I_{prev} + 0.0001) > 1 \\ 1 & \text{otherwise} \end{cases}$$

Here, $I_{current}$ is the intensity of the daughter cell nucleus, in the first time point where the two daughter nuclei were visible. Again, the Gaussian fit is used to find out which pixels are inside the daughter cell nucleus, in exactly the same way as this was done for the mother nucleus. I_{prev} is the intensity at the same pixel locations, but for the previous time point.

For daughter nucleus 2 the same equation is used. Both values are then summed:

$$\begin{aligned}\texttt{daughter\_intensity\_delta} = &-1 + \texttt{daughter1\_intensity\_delta} \\ &+ \texttt{daughter2\_intensity\_delta}\end{aligned}$$

1 is subtracted here to relax the scoring system and make divisions more likely.

Finally, we expect both daughter nuclei to have the same intensity, as they were created out of a single nucleus minutes ago. The only deviations of intensity between the two nuclei should occur if one daughter nucleus ends up further away from the microscope objective than the other. Therefore, we can lower the score for any difference in the average intensity of the two:

$$\texttt{daughters\_intensity\_difference} = \frac{|I_{daughter1} - I_{daughter2}|}{2}$$

Again, the intensity is the average intensity of the pixels within the nucleus, as defined by the Gaussian fit.

3 | Minimizing cell number fluctuations in self-renewing tissues with a stem cell niche

Published as:
Kok, R.N.U., Tans, S.J., & Van Zon, J.S. (2023). *Physical Review E*, 108(6), 064403.

Abstract

Self-renewing tissues require that a constant number of proliferating cells is maintained over time. This maintenance can be ensured at the single-cell level or the population level. Maintenance at the population level leads to fluctuations in the number of proliferating cells over time. Often, it is assumed that those fluctuations can be reduced by increasing the number of asymmetric divisions, i.e. divisions where only one of the daughter cells remains proliferative. Here, we study a model of cell proliferation that incorporates a stem cell niche of fixed size, and explicitly model the cells inside and outside the niche. We find that in this model fluctuations are minimized when the difference in growth rate between the niche and the rest of the tissue is maximized and all divisions are symmetric divisions, producing either two proliferating or two non-proliferating daughters. We show that this optimal state leaves visible signatures in clone size distributions and could thus be detected experimentally.

3.1 | Introduction

Many adult tissues, such as the mammalian intestinal epithelium and the skin epidermis, undergo constant self-renewal supported by stem cells [68, 69]. To ensure homeostasis, the number of proliferating cells needs to be kept constant over the lifespan of the organism. Large fluctuations in the number of proliferating cells could lead to disease or even the death of the organism. For that reason, tight regulation of cell proliferation is required [70–72], where every cell division must result in one proliferating and one non-proliferation cell, at least on average.

The balance of proliferation and differentiation can either be maintained strictly at the single-cell level or at the population level [73]. In the first strategy, every cell

division produces an asymmetric outcome: one daughter cell remains proliferative and the other daughter ceases proliferation and terminally differentiates. In the population level strategy, the outcome of each division is stochastic and can result in zero, one or two proliferating daughters. In this case, the balance between proliferation and terminal differentiation is maintained only on average, at the population level [74]. This maintenance strategy is therefore called the population-asymmetry model.

Unlike the single-cell level strategy, the population level strategy is inherently stochastic, and therefore potentially prone to fluctuations in the number of proliferating cells. Despite these fluctuations, the population level strategy of self-renewal is found in many stem cell systems [70]. Examples include the mammalian germline, the intestine and the epidermis[6, 18, 73, 75, 76]. In some organs large variations of the number of stem cells can occur, such as for spermatogenic stem cells in murine testes [74]. As such fluctuations are potentially dangerous, in other organs the number of stem cells appears to be tightly controlled, such as in the small intestine [57]. It is unknown how different strategies of stem cell maintenance affect fluctuations in stem cell numbers.

In many stem cell systems, cell proliferation is also organized in space, with stem cell niches that provide a local environment that maintains stem cells in an undifferentiated and proliferating state [74], while cells outside of such stem cell niches eventually cease proliferation and differentiate. As a consequence, the division patterns of proliferating cells also likely vary in space, with more divisions generating proliferating cells within the stem cell niche, and more non-proliferating cells without. How such a spatial segregation of proliferation dynamics might impact fluctuations in cell proliferation remains an open question.

Here, we use a theoretical approach to study the impact of different stem cell maintenance strategies, including whether a stem cell niche is present or not, on fluctuations in the number of proliferating cells. In particular, we compare proliferation dynamics in an uniform, unbounded system, i.e. lacking a stem cell niche, and a system with two compartments, a stem cell niche where cells are geared towards proliferation, and a differentiation compartment where cells are biased towards ceasing proliferation. We analytically derive under which conditions the two-compartment model is stable, meaning that the number of proliferating cells is stationary, and calculate the corresponding steady-state number of proliferating cells. We then systematically examine how different parameters, such as the fraction of symmetric and asymmetric division, or the size of the stem cell niche, impact the magnitude of fluctuations. We find that in the uniform model fluctuations are minimized when all divisions are asymmetric, strictly generating one proliferating and one non-proliferating daughter. Surprisingly, we find that in the two compartment model, that incorporates a stem cell niche, fluctuations are instead minimized by a very different strategy: when all divisions are strictly symmetric, generating either two proliferating or two non-proliferating daughters. Finally, our simulations show that these different strategies generate distinct clone size distributions, and could thus potentially be differentiated experimentally [72].

3.2 | Related work

Theoretical models have advanced our understanding of stem cell behavior. These models have focused on two main questions. First, what is the impact of various niche setups on cell number fluctuations? And second, what is the role of (a)symmetry in cell divisions [71, 72]?

We will first discuss the impact of niche setup on cell number fluctuations. The models used to study fluctuations can roughly be divided in three classes. A first class of models uses a uniform space with two cell types; proliferating and non-proliferating. Klein *et al.* [77] used such a model to determine the impact of various division strategies on the fluctuations of cell numbers, but only for individual lineages. Sun and Komarova [78] used the same type of model to test the impact of various feedback mechanisms on cell number fluctuations. Those feedback mechanisms required that cells are aware of the current amount of stem cells in the system. However, it is open question whether such feedback mechanisms are present in stem cell systems.

A second class of models uses a single compartment with a fixed number of stem cells, and no other cell types. These models were used to study competition between lineages. Snippert *et al.* [6] and Lopez-Garcia *et al.* [76] demonstrated that the intestinal crypt uses neutral competition, where the progeny of one stem cell eventually takes over the entire niche. Ritsma *et al.* [18] and Corominas-Murtra [79] studied the dependence of lineage survival on cell position within the niche.

Finally, in a third class of models two cell types (stem and non-stem) are distributed over two compartments. In this work, we will use a model of this class. So far, these models have only been used to study the risk of developing cancer, namely by Cannataro *et al.* [80, 81] and Shahriyari and Komarova [82]. Therefore, the question of how a compartmentalized system affect fluctuations in the number of cells remains open.

The second question is about the impact of division symmetry. The studies that focused on this question so far used a model with uniform space and both proliferating and non-proliferating cells. Klein *et al.* [77] used this model to infer division symmetry from experimental data of the skin epidermis, and concluded that asymmetric divisions were dominant. Sei *et al.* [83] used a similar setup for the intestinal crypt, and likewise concluded that asymmetric divisions were dominant. Yang and Komarova [84] used the model for a different purpose: they investigated the impact of division symmetry on fluctuations in the number of proliferating cells. The authors found that their results were mixed. Under some control mechanisms symmetric divisions provide lower fluctuations, but under most control mechanisms asymmetric divisions provided lower fluctuations. Therefore, it remains unclear what impact division symmetry has on the fluctuations in cell numbers.

3.3 | Results

3.3.1 One-compartment model

We start our analysis by looking at a model without space, corresponding to Klein *et al.* [77]. In the next section, this model will be extended to include a niche compartment. We use two cell types: proliferating and non-proliferating. Whether a cell is proliferative or not is decided at the birth of the cell; this fate cannot be changed later. Every proliferating cell will divide T hours after the birth of the cell, while non-proliferating cells will never divide. Values of T are drawn from a skew-normal distribution [85] with a skewness parameter 6.1, a location of 12.2 and a scale of 5.3. This distribution approximates cell cycle times we recently measured in intestinal organoid crypts [86].

After a division, two daughter cells are created, which can be (**I**) both proliferating, (**II**) both non-proliferating or (**III**) one can be proliferating and the other non-proliferating (**Fig. 1a**). The chances for these division types to occur are p, q and $1 - p - q$, respectively. Division types **I** and **II** are symmetric, while type **III** is asymmetric. The values of p and q are determined by two parameters. The parameter $\phi = p + q$ is the chance of a division being symmetric (**Fig. 1b**), while the other parameter, the growth rate $\alpha = p - q$, is the average increase in the number of proliferating cells per division (**Fig. 1c**). One can verify that for $\phi = 0$ all divisions are required to be of type **III**, while for $\alpha = 1$ all divisions are required to be of type **I**.

For homeostasis, in this system it is required that the growth rate $\alpha = 0$. For $\alpha > 0$ the number of proliferating cells would grow exponentially, and for $\alpha < 0$ this number would decrease exponentially. The symmetry fraction ϕ can be varied freely, as well as the initial number of proliferating cells D. We therefore perform a simulation for different combinations of D and ϕ, while keeping α at 0. The results are displayed in **Fig. 1d-g**.

In **Fig. 1d** six simulations of the number of proliferating cells are shown over time for two values of the symmetry fraction ϕ. We can observe that a higher fraction of asymmetric divisions (low ϕ) provides a system with less fluctuations in the number of proliferating cells. In addition, we see that the system with a high fraction of symmetric divisions (high ϕ) is frequently depleted of proliferating cells, while the smaller system with a high fraction of asymmetric divisions (low ϕ) remains stable for at least 10 days. In **Fig. 1e-g** the depletion rate, overgrowth rate and coefficient of variation in the number of proliferating cells are plotted as a function of both the initial number of proliferating cells D and the symmetry fraction ϕ. Here, depletion is defined as the number of proliferating cells becoming zero, while overgrowth is defined as the number of proliferating cells reaching five times the initial amount. We can see that the depletion and overgrowth rates increase for smaller D and larger ϕ, indicating a less stable system. The coefficient of variation remains high irrespective of D and ϕ, with one exception: the theoretical case where precisely all cell divisions are asymmetric ($\phi = 0$) the coefficient is zero.

As a result, in this model the best approach to minimize fluctuations and avoid depletion or overgrowth of proliferating cells would be to have a large number of proliferating

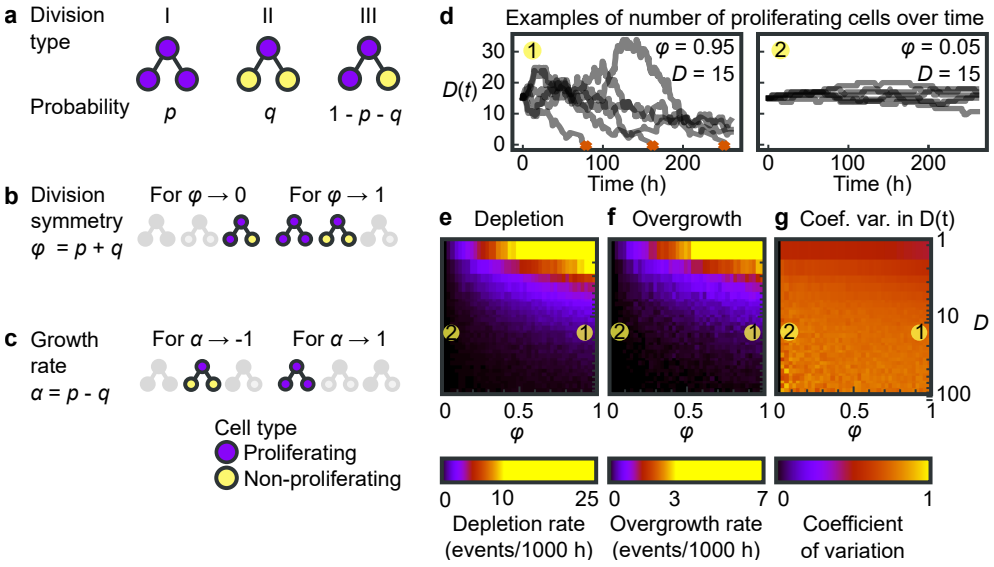


Figure 1 • The uniform model

(a) All possible division types and their probability. **(b)** Dominant division type for two different division symmetries. **(c)** Dominant division type for two different growth rates. **(d)** Two panels, each showing for six simulations the number of proliferating cells over time. The simulations use $D = 15$ and the given symmetry fraction ϕ. **(e)** Depletion rate, the rate of how often the number of proliferating cells drops to zero. **(f)** Overgrowth rate, the rate of how often the number of proliferating cells grew larger than 5 times the initial number of proliferating cells. **(g)** Coefficient of variation of the number of dividing cells.

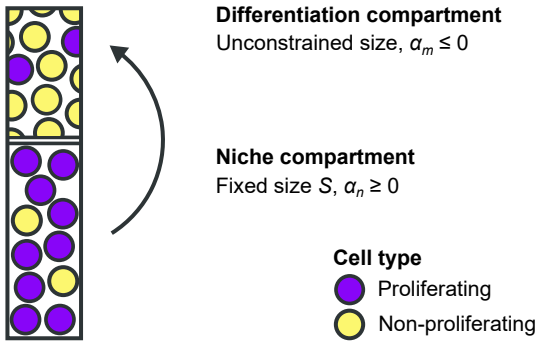
cells and to use strictly asymmetric divisions. A low amount of symmetric divisions already results in relatively large fluctuations in the number of proliferating cells.

3.3.2 Two-compartment model

We wanted to compare the performance of the uniform model, in terms of the impact of fluctuations, to a model that incorporated a stem cell niche. We therefore constructed a different model, that included two compartments, with cell proliferation differing between compartments. One compartment, which we call the niche compartment, can only contain a fixed number of cells. In contrast, the other other compartment, which we call the differentiation compartment, is unbounded in size. This two-compartment model is sketched in **Fig. 2**.

Figure 2 • Illustration of the two-compartment model

Two compartments are defined, the niche compartment and the differentiation compartment. In the niche compartment the growth rate $\alpha_n \geq 0$ while in the differentiation compartment the growth rate $\alpha_m \leq 0$. The fraction of symmetric divisions ϕ is equal in both compartments. Upon a division in the niche compartment, one random cell moves from the niche compartment to the differentation compartment.



The niche compartment is set to contain a fixed number of cells in total, denoted as S. Of this number, a variable number of cells $N(t)$ is proliferative, which makes the number of non-proliferating cells equal to $S - N(t)$. For the differentiation compartment we denote the number of proliferating cells as $M(t)$. As the differentiation compartment is unbounded in size, we do not keep track of the number of non-proliferating cells in this compartment. The total number of proliferating cells over both compartments is defined as $D(t) = N(t) + M(t)$. To keep the niche compartment fixed in size, upon every division in the niche compartment we move one random cell out of the niche compartment into the differentiation compartment.

Cell divisions still follow the rules established in the previous section (**Fig. 1a-c**). The key difference is that instead of using a single growth rate α, both compartments can now have a different growth rate. We define the niche compartment as having a growth rate of α_n and the differentiation compartment as having a growth rate of α_m. For the division symmetry fraction ϕ, we assume that it is kept equal across both compartments.

3.3.3 Analytical solution of the two-compartment model

We then examined for which values of the growth rates of the two compartments, α_n and α_m, the number of proliferating cells is stationary and asked what the resulting steady state number of dividing cells D would be, given all parameters of the system.

We addressed this problem using an analytical approach. A trivial solution is $\alpha_n, \alpha_m = 0$, which would be equivalent to the uniform model with $\alpha = 0$, as discussed above. Here, we will focus on all other possible solutions. For the derivation of the steady-state number of proliferating cells, we ignore that there is variation in the cell cycle time, and simply define the cell cycle time as T.

Using a transport model we arrive at the following two equations for the niche compartment:

$$\frac{\partial n(t,a)}{\partial t} + \frac{\partial n(t,a)}{\partial a} = -\frac{n(t,T)}{S} n(t,a) \tag{3.1}$$

$$n(t,0) = (1+\alpha_n) n(t,T) \tag{3.2}$$

The full derivation can be found in Appendix 3.6.1. In the above equations, a is the age of a cell, with $a = 0$ for newly-born cells and $a = T$ at the moment of a cell division. $n(t,a)$ is the number of cells at a given time with a given age, and integrating over all values of a results in $N(t)$. The transport model describes the number of cells of age a to $a + da$ at time t. This number changes due to aging of cells, cell divisions and due to proliferating cells exiting the niche compartment. At every cell division, on average $1 + \alpha_n$ proliferating daughters are born. In addition, exactly one cell is ejected from the niche compartment, which has a chance of $\frac{N(t)}{S}$ of being a proliferating cell.

For the differentiation compartment, we arrive at a transport equation (**Eq. 3.8**) that resembles **Eq. 3.1**, but with a positive right hand term representing the influx of cells from the niche compartment. In steady state, we arrive at the following solution for the total number of dividing cells:

$$D = \ln(1+\alpha_n) S \frac{\alpha_m - \alpha_n}{\alpha_m} \tag{3.3}$$

This equation is a central result of our study, and it has several implications.

1. The equation has solutions for $D > 0$ if $\alpha_m < 0$ and $\alpha_n > 0$. This dictates that the system can be in a steady state for other growth rates than $\alpha_n, \alpha_m = 0$. As long as the growth rate of the niche compartment is positive and the growth rate of the differentiation compartment is negative, a steady state is reached. For other values, the systems either decays or grows without bounds.

2. The number of proliferating cells in the entire system D increases as α_n and α_m increase. The number of proliferating cells $D = 0$ independent of α_m for $\alpha_n = 0$, while it increases well beyond the niche size S for $\alpha_m \approx 0$.

3. The equilibrium number of proliferating cells in the system is independent of both the cell cycle time T and the division symmetry fraction ϕ.

4. The number of proliferating cells in the entire system D always scales linearly with the total size of the niche compartment.

3.3.4 Stochastic simulations of the two-compartment model

Impact of fluctuations From our analytical results, we have seen what values of the growth rates of the two compartments, α_n and α_m, result in a homeostatic system. However, it leaves open how sensitive this steady state is to fluctuations. For an uniform system, we saw that asymmetric divisions resulted in a system with lower fluctuations in the number of proliferating cells. We therefore asked if more division asymmetry also results in fluctuations in the two-compartment model.

We start by performing a parameter sweep, sampling all possible combinations of α_n, α_m and ϕ, while keeping the initial number of proliferating cells at D=30. The niche compartment size S is calculated using **Eq. 3.3**. As a result, S varies with both α_n and α_m. We reproduced the expected average value $\langle D(t)\rangle$=30 for all combinations of ϕ, α_n and α_m (**Fig. S1**), confirming the validity of our analytical result in **Eq. 3.3**.

In **Fig. 3a**, we show $D(t)$ for two parameter sets that differ strongly in the degree of symmetry. In the first case, almost all cells divide asymmetrically, while $\alpha_n, \alpha_m = 0$, thus corresponding to the strategy that minimizes fluctuations in the uniform model. In the second, all cells divide symmetrically, while the growth rate differs strongly between compartments. Interestingly, we find that the parameter set with high fraction of symmetric divisions results in smaller fluctuations in $D(t)$. This is in contrast with the uniform model, where asymmetric divisions always provided less fluctuations compared to symmetric divisions.

In **Fig. 3b** we quantify the coefficient of variation in $D(t)$ for different combinations of α_n, α_m and ϕ. In general, we find that fluctuations decrease when the difference in growth rate between compartment,$\alpha_n - \alpha_m$, increases. When we focus on the the subset of parameter sets with $\alpha_n = -\alpha_m$, we find that for a fixed value of the two growth rates, fluctuations are reduced by increasing the fraction of asymmetric divisions (decreasing ϕ), which appears consistent with our results for the uniform model and at odd with our observation in **Fig. 3a** that high symmetry resulted in low fluctuations.

To examine this further, we compare in **Fig. 3c** the coefficient of variation along the four lines shown in **Fig. 3b**. In the top panel, we compare two lines for parameter sets with only symmetric divisions ($\phi = 1$) that have either a larger (cyan) or smaller difference (pink) in growth rate between the two compartments. In both cases, the coefficient of variation in $D(t)$ decreases with increasing difference in growth rates, but the parameter sets with the smaller difference (pink) always showed larger fluctuations.

The bottom panel shows two other lines, where we varying ϕ while fixing $\alpha_n = -\alpha_m$. The red line uses for every value of α_n the lowest possible value of ϕ, i.e. the maximum amount of asymmetry that the growth rate permits. At $\alpha_n = 1$ all divisions

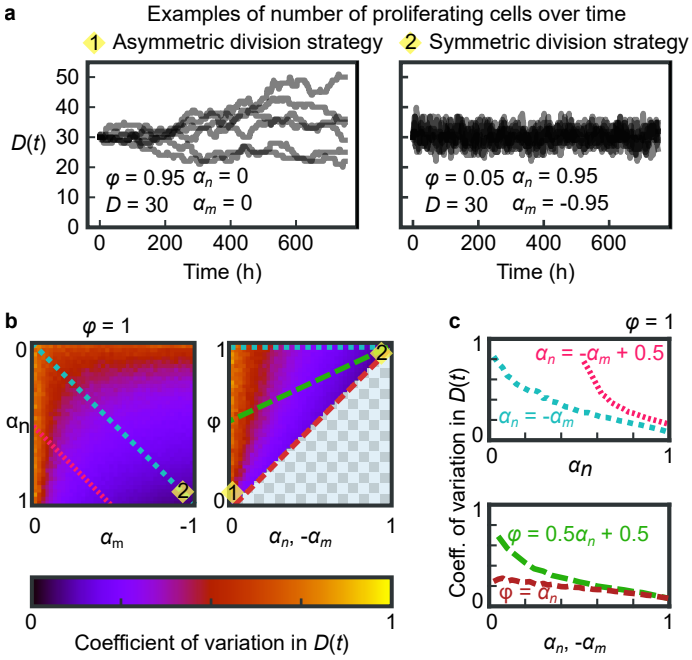


Figure 3 • Exploration of the two-compartment model
Simulations ran 10^5 hours for each parameter set. The initial number of dividing cells $D = 30$ and niche compartment size S is set according to **Eq. 3.3**. **(a)** Two panels showing $D(t)$ for six example simulations each of the optimal parameters for either an asymmetric divisions strategy (left) or a symmetric division strategy (right). **(b)** Overview of the coefficient of variation in the number of dividing cells $D(t)$. The blocked areas represent impossible combinations of parameters, such as strictly asymmetric divisions with a positive growth rate. **(c)** Coefficient of variation along selected lines. Line style matches the lines of panel b.

are symmetric. The green line has a higher fraction of symmetric divisions. The lines show that for a given growth rate α_n, the parameter set with the highest asymmetry results in the smallest coefficient of variation in $D(t)$, as seen above. However, at the same time the global minimum of the coefficient of variation is found for $\alpha_n, -\alpha_m = 1$. This is because the decrease in fluctuations is strongest when difference in the two growth rates is maximal. For these values of the growth rate only $\phi = 1$ is allowed, which is a system with only symmetric divisions. Therefore, although asymmetry generally reduces the coefficient of variation in $D(t)$, the optimal solution is a fully symmetrically dividing system.

Based on **Fig. 3b** and **Fig. 3c**, we can conclude that the in the two compartment model the best strategy to minimize fluctuations in the number of proliferating cells $D(t)$ is to use a system where all cells in the niche compartment proliferate, but each cell born outside the niche compartment immediately stops proliferating. This results in dominance of symmetric cell divisions. Other solutions, such as strictly asymmetric divisions or a combination of symmetric and asymmetric divisions result in more fluctuations in the number of proliferating cells.

Dependence of fluctuations on niche size Next, we wondered what the influence of the size of the niche compartment would be on the stability of the system. For each size, we plot the coefficient of variation along two lines. For the first line, we take the line that describes all points for which $\phi = \alpha_n, -\alpha_m$, corresponding to the red dashed line in **Fig. 3b**. This is the line with the lowest coefficient of variability in the number of dividing cells $D(t)$ for every ϕ. For the second line, we examine the dependence of the niche compartment size S for the parameters on the line with $\phi = 1$ and $\alpha_n = -\alpha_n$, corresponding to the cyan dotted line in **Fig. 3b**. This line represents the line with the largest difference in coefficient of variability of $D(t)$; it goes from the global minimum to the global maximum of the coefficient of variability in $D(t)$ for a given S.

The results are displayed in **Fig. 4a**. Consistent with the results above, for every given size, increasing the difference in growth rates α_n and α_m will always result in a system with a lower coefficient of variation in $D(t)$ and a lower depletion rate. Moreover, at larger compartment sizes S the coefficient of variation in $D(t)$ always becomes lower. Therefore, independent of the chosen parameter set, a system can always decrease its relative fluctuations by creating a larger niche. For the same reason, decreasing the niche compartment size results in more frequent depletions of the proliferating cells (**Fig. 4b**).

The observation that a smaller system results in more fluctuations is as expected. However, interestingly even for small niche sizes the system can remain stable, provided the symmetry fraction is high. Decreasing the niche compartment size from $S = 30$ to only $S = 10$ still results in a system with low fluctuations (less than 1 collapse per 10 years), provided the niche compartment maintains a high growth rate and symmetry fraction.

Determining growth rate and division symmetry by clone size distributions Experiments often measure clone size distributions using lineage tracing. For instance,

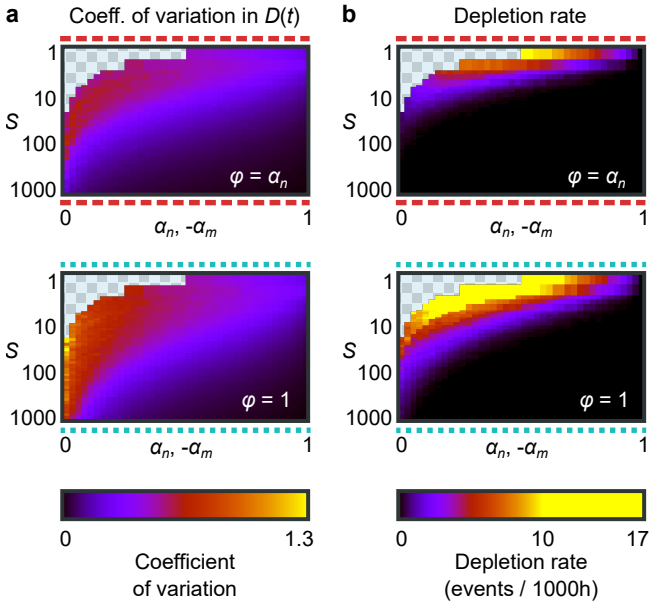


Figure 4 • Simulations for different niche compartment sizes S, each for 10^5 hours

The initial number of dividing cells D is set according to **Eq. 3.3**. The striped lines above and below the graphs correspond to the lines in **Fig. 3**. The blocked areas in the graphs are parameter ranges for which **Eq. 3.3** does not give a solution. **(a)** Coefficient of variation of $D(t)$ for various parameters. **(b)** Depletion rate for various parameters.

using a model analogous to the uniform model above, the division symmetry ϕ was estimated by fitting to experimental long-term clone size scaling distributions, measured for different time frames [77]. We therefore asked whether parameters such as growth rate and division symmetry could also be inferred from clone size distributions obtained in the context of the two-compartment system.

For short-term clone size distributions, dominance of symmetric divisions appears clearly as an enrichment of even clone sizes. In **Fig. 5a** we show simulated clone size distributions of 1000 simulation runs, taken after 48 hours. For the case where symmetric divisions are dominant ($\phi = 0.95$, displayed in **Fig. 5a**, left), we can see that even clone sizes 2, 4, 6 and 8 all occur in higher frequencies than the odd clone sizes 3, 5 and 7. This enrichment is not visible for systems where asymmetric divisions are dominant, or where neither division type is dominant (**Fig. 5a**, center and right).

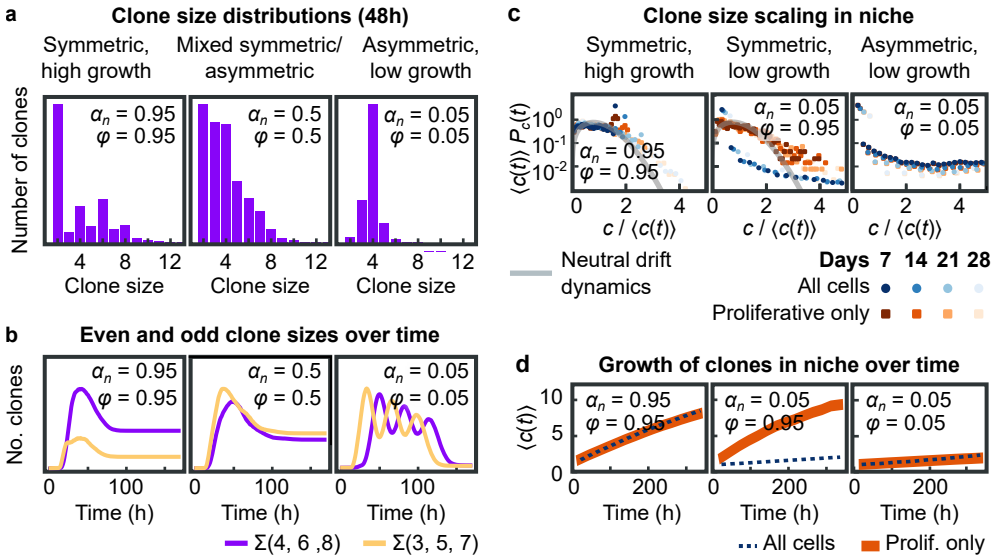


Figure 5 • Clone size distributions of cells within the niche compartment
1000 simulations, $\alpha_m = -\alpha_n$, $D = 30$ and S is set according to **Eq. 3.3**. **(a)** Short-term clone size distributions, taken after 40 hours. **(b)** Amounts of even and odd clone sizes over time. For the symmetric, high growth case there is initially a large enrichment of even clone sizes, which becomes smaller after 50 hours. For the mixed symmetric/asymmetric case, odd clone sizes are enriched. For the asymmetric case, all clones grow at almost the same rate, and therefore a single clone size is dominant at any point in time. This causes oscillatory behavior in whether even or odd clone sizes are dominant. **(c)** Clone size scaling for the symmetric, low growth case. For these parameters, there is a large contrast between the clone size distribution of all cells and the clone size distribution of only the proliferating cells. **(d)** Growth of the average clone size over time.

However, as the clone size over time in **Fig. 5b** shows, this enrichment of even clone sizes is strongest until 60 hours. In addition, during the entire simulation the enrichment is barely visible for clone sizes larger than 8 (**Fig. 5a**, **Fig. S2**). In the simulation this has two causes. First, we simulate for $\phi = 0.95$, which still results in 5% of all divisions

being asymmetric. While symmetric divisions will keep the clone size of a single lineage even, asymmetric divisions will not. Already a single asymmetric division anywhere in the lineage tree results in the clone size becoming odd.

Second, the variability in cell cycle times also contributes to the occurrences of odd clone sizes. Consider for example two proliferating sister cells. Initially the clone size is two, and after both sisters have divided it is four. However, in between the divisions of both sisters, the clone size is 3 and therefore odd. This effect occurs more often for large lineage trees, simply because there are more sister pairs, therefore increasing the chance of at least one of them making the clone size odd. In conclusion, the effect of division symmetry is visible only on the time scale of a few divisions.

Often, clone size growth is characterized experimentally by the clone size scaling function $P_c(t)$ [72], which is a function that gives the probability of finding a clone of size c at time t. Unlike the distributions discussed above, the scaling function $P_c(t)$ is not dependent on time, but only on the clone size divided by the average clone size for that time: $P_c(t) = f\left(c/\left\langle c(t)\right\rangle\right)$. Would there be a way to measure both the symmetry fraction and the growth rate from this scaling function? In other words: do the symmetry fraction and growth rate affect the observed long-time scaling behavior?

In experiments, clone size distributions are often collected only in the stem cell niche. Moreover, our simulations do not implement cell death, which results in the number of non-proliferating cells outside the niche compartment growing continuously. Such unbounded growth would prevent any long-term scaling behavior. For that reason, we will in our analysis ignore all cells outside the niche compartment.

We calculated the scaling function of different combinations of the symmetry fraction ϕ and growth rate α_n. First, we find that for all parameters examined the clone size distributions scale, meaning that they fall on the same curve independent of time, both when considering only proliferating cells or all cells in the niche compartment. In **Fig. 5c** we show the scaling functions, both for all cells in the niche compartment and for only the proliferating cells. Interestingly, the shape of the scaling function for the proliferating cells in the niche is almost independent on the growth factor α_n, as both for $\alpha_n = 0.05$ (**Fig. 5c**, left) and $\alpha_n = 0.95$ (**Fig. 5c**, center) the scaling function follows a concave shape. Instead, the scaling function for proliferative cells depends on the symmetry fraction ϕ: for high symmetry fractions the scaling function follows the concave pattern predicted by neutral drift dynamics [76], which is $\left\langle c(t)\right\rangle P_c(t) = (\pi x/2)\cdot e^{-\pi x^2/4}$, with $x = c/\left\langle c(t)\right\rangle$ (**Fig. 5c**, left and center), while for low symmetry fractions the scaling function follows a convex shape (**Fig. 5c**, right).

However, if we look at all cells, then the scaling function no longer depends on the symmetry fraction ϕ, but on the niche growth rate α_n. For high α_n the scaling function follows a concave shape (**Fig. 5c**, left), while for low α_n the function follows a convex shape (**Fig. 5c**, center and right). The scaling functions of other parameter sets are displayed in **Fig. S3**, and are consistent with these observations.

The cause of this scaling behavior can be seen in **Fig. 5d**. In this panel, we plot the average clone size over time for the same three parameter sets. As expected, the average clone size for all cells in the niche compartment increases faster for a high growth rate (**Fig. 5d**, left) compared to a low growth rate (**Fig. 5d**, center and right). However, if

we include only the proliferating cells, then the clone size growth depends on ϕ. For high ϕ (**Fig. 5d**, left and center), the clones grow faster than for low ϕ (**Fig. 5d**, right). This is because the number of proliferating cells can only grow due to symmetric divisions, as asymmetric divisions do not have an effect on the number of proliferating cells. Even though the total number of proliferating cells is independent on ϕ (**Eq. 3.3**), the proliferating cells are now distributed over more clones. We can see that the fast-growing clones in **Fig. 5d** correspond to the concave scaling functions in **Fig. 5c**, and the slow-growing clones correspond to the convex scaling functions.

3.4 | Discussion

In this paper, we investigated the impact of fluctuations for different stem cell maintenance strategies, in the context of a stem cell niche. For this, we used a model with two compartments and two cell types, namely proliferating cells and non-proliferating cells. The model assumes a stem cell niche that is restricted in size, and that cells decide whether to continue proliferating or not depending only on the basis of the identity of the compartment in which they are born. No other interactions between cells are assumed, and besides their proliferation state, no other internal cellular state is considered. The key parameters governing the dynamics of this model are each compartment's growth rate, that indicates how many proliferating cells are produced on average through each cell divisions, and the division symmetry, that describes how likely two daughters cells are to have the same proliferating or non-proliferating cell type.

We used a transport model approach to obtain an analytical solution describing the average dynamics of the two-compartment model. Combined with stochastic simulations, we found that this model allowed for two distinct strategies to minimized stem cell number fluctuations under homeostatic conditions. If the two compartments do not differ in growth rate, then the overall growth rate must be zero to balance proliferation and cells must maximize the fraction of asymmetric divisions. This special case of our model corresponds to the one-compartment model developed previously by Klein *et al.* [77], and also used by Sei *et al.* [83]. However, if the two compartments are allowed to differ in growth rate, then another strategy is possible that exhibits substantially lower cell number fluctuations. In this strategy, fluctuations are minimized when the different in growth rate between the two compartments is maximized and hence all divisions are symmetric. This finding is in contrast with the view that asymmetric divisions offer a more regulated stem cell maintenance process in stem cell niches [87, 88], but consistent with measurements performed recently by us that show that symmetric divisions dominate in growing intestinal organoids[86]. In this optimal limit, the dynamics in the niche compartment of our model becomes similar to the models of Snippert *et al.* [6] and Lopez-Garcia *et al.* [76]. Our results thus show that their models, which were based on experimental observations, correspond to a parameter regime, in our more general model, that minimizes fluctuations in the number of proliferating cells.

Our simulations indicate that the key parameters of our model, growth rate and division symmetry, can be determined experimentally through measuring clone size distribu-

tions. Clone size distributions are commonly measured using techniques such as genetic lineage tracing [3, 5]. For sufficiently short times (0-50 hr, corresponding to a few cell cycles) high division symmetry can be detected through enrichment of even-sized clones. Indeed, we observed such an enrichment recently in intestinal crypts in vivo using lineage tracing [86], in agreement with high division symmetry we observed in intestinal organoids by direct cell tracking. However, for most experiments clone size distributions are collected over much longer time scales, for which our simulations show that enrichment of even-sized clones is no longer observable. Here, we found that growth rate and division symmetry can still be determined by comparing clone size distributions either for all cells in the stem cell niche, or only proliferating cells, as they give rise to different scaling functions for clone size distributions as well as distinct increases in average clone size. Experimentally, this could be readily achieved by combining lineage tracing reporters with antibody markers for cell proliferation.

Many stem cells systems share the two-compartment architecture of the intestine, with a small niche containing stem cells that inject differentiating cells into the rest of the tissue [89]. It will therefore be interesting to experimentally test whether such tissues also exhibit dynamics consistent with the optimal limit of our model, with strong difference in growth rate between the stem cell niche and the rest of the tissue, and most divisions symmetric, producing either two proliferating or two non-proliferating daughters. Our analysis provides a starting point to test this experimentally.

3.5 | Acknowledgments

We would like to thank Max A. Betjes for useful discussions, suggestions and comments.

3.6 | Supplementary Information

3.6.1 Derivation of Eq. (3.3)

When a cell is born, it is set to be either proliferating or non-proliferating. Proliferating cells will divide T hours after they are born, while non-proliferating cells will not divide. Here, T is a normal distribution with We define $n(t,a)da$ as the number of proliferating cells in the niche compartment at time t with age bracket $(a, a+da)$. Here, age is defined as the time since the last division. The number of cells flowing in and out of this age bracket due to aging is:

$$J_+ = \frac{dt}{da} n(t, a-da)da$$
$$J_- = -\frac{dt}{da} n(t,a)da$$

The rates for cells entering and exiting the niche compartment are defined as $k_+(a)$ and $k_-(a)$, respectively. Note that in our simulation cells do not reenter the niche compartment so $k_+(a) = 0$.

Using this information, we look at how the number of cells of a particular age bracket evolves over time. This number is equal to the number of existing cells of that age bracket, plus the number of incoming cells due to aging, minus the number of cells exiting the age bracket due to aging, plus the number of cell changes due to cells entering and exiting the compartment.

$$n(t+dt,a)da = n(t,a)da + \frac{dt}{da}n(t,a-da)da \\ - \frac{dt}{da}n(t,a)da + (k_+(a)+k_-(a))dadt$$

Rearranging, we find:

$$\frac{n(t+dt,a)-n(t,a)}{dt} + \frac{n(t,a)-n(t,a-da)}{da} \\ = k_+(a)+k_-(a)$$

Using the definition of the derivative we readily find:

$$\frac{\partial n(t,a)}{\partial t} + \frac{\partial n(t,a)}{\partial a} = k_+(a) - k_-(a) \tag{3.4}$$

Growth boundary condition Next, we introduce the growth boundary condition. If there are no cells moving in between compartments then $k_+(a) = k_-(a) = 0$ and $dN = J_{in} - J_{out}$. Here, $J_{out} = \frac{dt}{da}n(t,T)da$ is the rate of cells exiting the cell cycle, and J_{in} the rate of cells entering the cell cycle. As each division produces on average $\alpha + 1$ proliferating cells, $J_{in} = (\alpha+1)\cdot J_{out}$. Together:

$$\frac{dN}{dt} = \frac{J_{in}-J_{out}}{dt} = \alpha n(t,T) \tag{3.5}$$

From the definition $N(t) = \int_0^T da\, n(t,a)$ it follows that:

$$\frac{dN}{dt} = \int_0^T da\, \frac{\partial n(t,a)}{\partial t}$$

From **Eq. 3.4** in the case where $k_+(a) = k_-(a) = 0$, we can see that $\frac{\partial n(t,a)}{\partial t} = -\frac{\partial n(t,a)}{\partial a}$. Inserting this in the above integral and evaluating it result in:

$$\frac{dN}{dt} = n(t,0) - n(t,T) \tag{3.6}$$

Combining **Eq. 3.6** with **Eq. 3.5** results in:

$$n(t,0) = (1+\alpha) \cdot n(t,T) \tag{3.7}$$

Full model Having satisfied the growth condition, we will now calculate the values of $k_-(a)$ and $k_+(a)$, which represents the movement of cells between compartments. In our model, cells cannot re-enter the niche compartment, so $k_+(a) = 0$. For expressing $k_-(a)$, we realize that the fixed size of the niche compartment means that for every cell division, one random cell must be removed from the niche compartment and added to the differentiation compartment. Therefore, $k_-(a)$ is equal to the number of cells dividing in a given time interval, multiplied by the chance that a cell being ejected is a dividing cell, multiplied by the probability that the ejected cell has the age $(a, a+da)$:

$$k_-(a)da\ dt = \left(\frac{dt}{da} n(t,T)da\right) \cdot \left(\frac{N(t)}{S}\right) \cdot \left(\frac{n(t,a)da}{N(t)}\right)$$

Here, we remind the reader that $N(t)$ is defined as the total number of proliferating cells in the niche at time t, and that $n(t,a)$ is the number of proliferating cells in the niche compartment of age a at time t.

From this equation it directly follows that:

$$k_-(a) = \frac{n(t,T)}{S} n(t,a)$$

Inserting these results in **Eq. 3.4**, we obtain **Eq. 3.1**:

$$\frac{\partial n(t,a)}{\partial t} + \frac{\partial n(t,a)}{\partial a} = -\frac{n(t,T)}{S} n(t,a)$$

The number of newly-born proliferating cells in the niche was already defined by **Eq. 3.7**.

For the differentiation compartment a similar analysis can be made, noting that here $k_-(a)$ is now negative, and $k_+(a)$ is now equal to $k_-(a)$ of the niche compartment. The number of proliferating cells in the differentiating compartment of age a at time t is defined as $m(t,a)$.

$$\frac{\partial m(t,a)}{\partial t} + \frac{\partial m(t,a)}{\partial a} = \frac{n(t,T)}{S} n(t,a) \tag{3.8}$$

The equivalent of **Eq. 3.7** for the differentiation compartment simply becomes:

$$m(t,0) = (1+\alpha_m) m(t,T) \tag{3.9}$$

Steady state for the niche compartment To solve the number of proliferating cells, we assume that the age distribution is exponential:

$$n(t,a) = f(t) \cdot e^{ra} \tag{3.10}$$

Here, $f(t)$ is a function independent of a and r is a coefficient. At $a = 0$ we obtain $n(t, a = 0) = f(t)$. From **Eq. 3.7** we obtain $n(t, a = 0) = (1 + \alpha_n) \cdot n(t, T) = (1 + \alpha_n) \cdot f(t)e^{rT}$, where we inserted **Eq. 3.10**. This results in:

$$r = -\frac{ln(1 + \alpha_n)}{T} \tag{3.11}$$

Next, we substitute **Eq. 3.10** into **Eq. 3.1**:

$$\frac{\partial f(t)}{\partial t} e^{ra} + r\, f(t)e^{ra} = -\frac{f(t)e^{rT}}{S} f(t)e^{ra}$$

By dividing by e^{ra} and rearranging, we find:

$$\frac{\partial f(t)}{\partial t} = f(t) \left(\frac{\ln(1 + \alpha_n)}{T} - \frac{f(t)}{(1 + \alpha_n)S} \right) \tag{3.12}$$

In steady state, $f' = f(t)$ so that $\frac{\partial f(t)}{\partial t} = 0$. Therefore, from **Eq. 3.12** two solutions arise. The first is the trivial solution $f' = 0$, the second

$$f' = (1 + \alpha_n) \frac{\ln(1 + \alpha_n)}{T} S$$

Inserting this solution into **Eq. 3.10** (with **Eq. 3.11**) results in:

$$n'(a) = (1 + \alpha_n) \frac{\ln(1 + \alpha_n)}{T} S\, e^{-\frac{\ln(1+\alpha_n)}{T} a} \tag{3.13}$$

Integrating over all values of a results in the number of dividing cells in the niche compartment:

$$\begin{aligned} N' &= \int_0^T da\, n'(a) \\ &= (1 + \alpha_n)S \left(1 - \frac{1}{1 + \alpha_n} \right) \\ &= \alpha_n S \end{aligned} \tag{3.14}$$

Here, we used:

$$\int_0^T da\, e^{-\frac{\ln(1+\alpha_n)}{T} a} = -\frac{T}{\ln(1 + \alpha_n)} \left(e^{-\ln(1+\alpha_n)} - 1 \right)$$

Steady state for the differentiation compartment To obtain the total number of dividing cells, we also need to calculate the number of cells in the differentiation compartment. For this compartment, we need to take into account that there is an incoming supply of dividing cells from the niche compartment. In steady state, $\frac{\partial m(t,a)}{\partial t} = 0$. To ensure homeostasis in the full system, $\frac{\partial m'(a)}{\partial a} = -\frac{\partial n'(a)}{\partial a}$. From **Eq. 3.13** we therefore obtain:

$$\frac{\partial m'(a)}{\partial a} = (1+\alpha_n)\left(\frac{\ln(1+\alpha_n)}{T}\right)^2 S\, e^{-\frac{\ln(1+\alpha_n)}{T}a}$$

As the age distribution is exponential, we can state:

$$m'(a) = A\, e^{Ba} + C$$

From inserting this equation into the one above it and comparing terms, it follows that $B = -\frac{\ln(1+\alpha_n)}{T}$ and $A = -(1+\alpha_n)\frac{\ln(1+\alpha_n)}{T}S$.

From **Eq. 3.9** we obtain $m'(0) = (1+\alpha_m)m'(T)$, resulting in:

$$C = \frac{\ln(1+\alpha_n)}{T} S \frac{\alpha_m - \alpha_n}{\alpha_m}$$

Together,

$$m'(a) = \frac{\ln(1+\alpha_n)}{T} S \left(\frac{\alpha_m - \alpha_n}{\alpha_m} - (1+\alpha_n)e^{-\frac{\ln(1+\alpha_n)}{T}a}\right)$$

With $M' = \int_0^T da\, m'(a)$ we obtain

$$M' = \ln(1+\alpha_n) S \frac{\alpha_m - \alpha_n}{\alpha_m} - \alpha_n S$$

With $D = M' + N'$ and the insertion of **Eq. 3.14** we obtain the final result:

$$D = \ln(1+\alpha_n) S \frac{\alpha_m - \alpha_n}{\alpha_m}$$

This equation has solutions for $D > 0$ if $\alpha_m < 0$ and $\alpha_n > 0$. If $\alpha_n < 0$ then N' would be negative, and if $\alpha_m > 0$ then M' would be negative, which are both not allowed.

3.6.2 Average number of proliferating cells

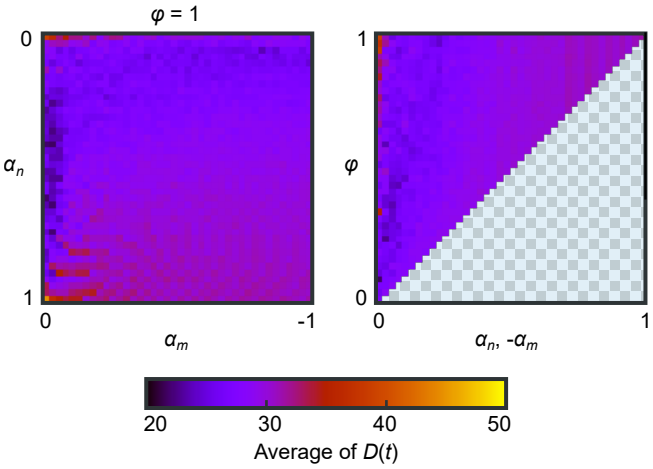


Figure S1 • Average of $D(t)$ for a 10^5h simulations of each parameter set
The initial D was set according to **Eq. 3.3**. The figure uses the same layout as panel **(b)** of **Fig. 3**.

3.6.3 Frequency of even and odd clone sizes over time

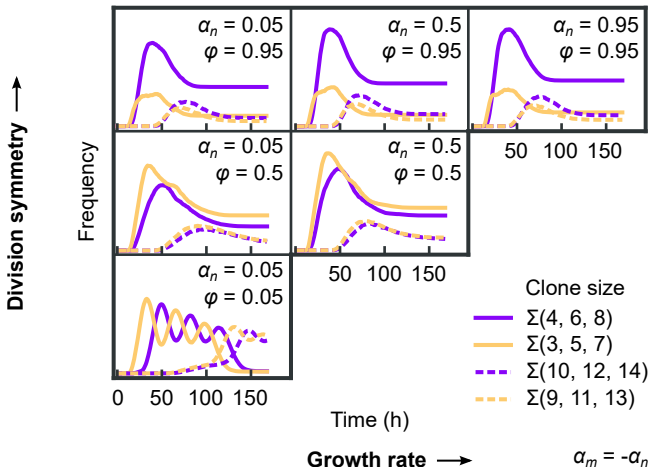


Figure S2 • Clone sizes plotted over time for the entire system, for the sums of the given even and odd clone sizes. The evolution of the clone size depends mainly on the division symmetry, not on the growth rate.

3.6.4 Clone size scaling functions

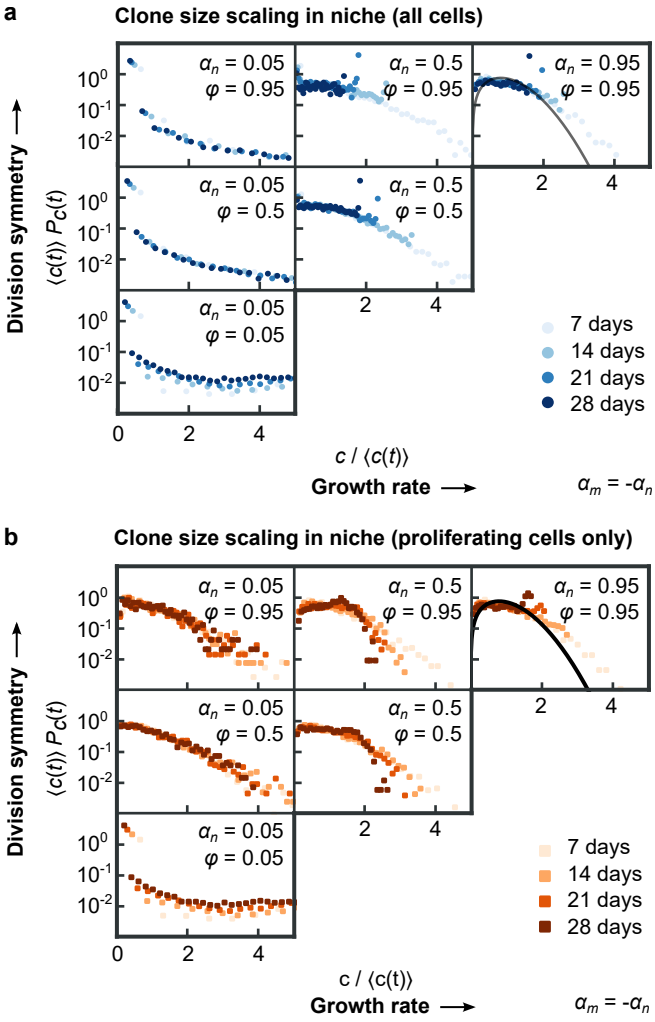


Figure S3 • Scaling function for various configurations
(a) Scaling functions that include all cells in the niche compartment. **(b)** Scaling functions that include only the proliferating cells in the niche compartment.

4 | Mother cells control daughter proliferation

Published as:
Huelsz-Prince, G., **Kok, R.N.U.**, Goos, Y.J., Bruens, L., Zheng, X., Ellenbroek, S.I., Van Rheenen, J., Tans, S.J., & Van Zon, J.S. (2022). *eLife*, 11, e80682.

Abstract

During renewal of the intestine, cells are continuously generated by proliferation. Proliferation and differentiation must be tightly balanced, as any bias towards proliferation results in uncontrolled exponential growth. Yet, the inherently stochastic nature of cells raises the question how such fluctuations are limited. We used time-lapse microscopy to track all cells in crypts of growing mouse intestinal organoids for multiple generations, allowing full reconstruction of the underlying lineage dynamics in space and time. Proliferative behavior was highly symmetric between sister cells, with both sisters either jointly ceasing or continuing proliferation. Simulations revealed that such symmetric proliferative behavior minimizes cell number fluctuations, explaining our observation that proliferating cell number remained constant even as crypts increased in size considerably. Proliferative symmetry did not reflect positional symmetry, but rather lineage control through the mother cell. Our results indicate a concrete mechanism to balance proliferation and differentiation with minimal fluctuations, that may be broadly relevant for other tissues.

4.1 | Introduction

Most adult organs and tissues constantly renew themselves by replacing old and damaged cells, while retaining their structure [70]. Theory indicates that this homeostasis requires a precise balance between proliferating and non-proliferating cells, as even a slight systematic bias towards producing proliferating cells yields uncontrolled exponential cell growth [73, 75–77, 90, 91]. Moreover, the exponential nature of proliferation also readily amplifies fluctuations in the numbers of proliferating cells, which can lead to stochastic cell overgrowth or depletion in absence of additional control mechanisms [78, 92]. How cell proliferation is balanced despite fluctuations has remained challenging to test in direct experiments, given the difficulties of following this process

in time.

The mammalian intestine has become an important model system to study the mechanisms of tissue renewal and homeostasis [57, 70]. The proliferating stem cells that sit at the base of intestinal crypts generate rapidly dividing transit-amplifying (TA) cells that in turn replenish the absorptive and secretory cells populating the lining of intestinal villi. Paneth cells positioned at the crypt bottom provide short-range signals that affect the proliferative and undifferentiated state of intestinal stem cells [93–95]. Originally, it was proposed that one or a few stem cells generated all differentiated cells by strictly asymmetric cell divisions [96, 97], thus directly ensuring a constant stem cell pool. Subsequent studies rather suggested that individual cells stochastically and independently cease to divide or not [6, 18, 76]. In this 'population asymmetry' model, in principle one stem cell daughter could remain proliferative by staying adjacent to a Paneth cell, while the other daughter exits the stem cell niche, differentiates, and stops proliferating. However, such asymmetric outcome is no longer guaranteed. Instead, proliferation and differentiation are balanced more indirectly, by averaging these stochastic events across the total stem cell population.

Observations of neutral drift, in which the offspring of a single cell randomly take over the stem cell population of intestinal crypts [6, 18, 76] established the stochastic nature of stem cell proliferation that distinguishes the population asymmetry model from the earlier division asymmetry model. However, approaches used thus far do not follow the underlying cell divisions and lineages in time. Proliferation symmetry between sister cells and its role in homeostasis of the intestinal epithelium have so far only been inferred indirectly, typically by quantifying the clone size distributions at a certain time point. Hence, we also lack insight into the fluctuations in the number of proliferating cells and the mechanisms that control them.

Here, we developed an alternative approach: we employed time-lapse microscopy and single-cell tracking of all cells in crypts of mouse intestinal organoids [98], thus providing complete lineage trees, division dynamics, and cell movement, and combine it with mathematical modelling and intravital imaging of the mouse intestine. Surprisingly, we found that most cell divisions ($> 90\%$) were symmetric in proliferative outcome, producing daughter cells that either both continued to proliferate or both ceased proliferating. Proliferation was symmetric even when one daughter neighbored a Paneth cell, the source of proliferative signals in the crypt, while the other did not. Hence, proliferation was not independent between sisters, but rather controlled through the lineage by the mother. Our data and simulations explained not only how this behavior achieves homeostasis, but moreover, that it constitutes a near-optimal strategy to minimize fluctuations in the number of proliferating cells. Consistently, despite their large size increases over multiple generations in crypts of various size, the number of proliferating cells was notably constant in time and exhibited sub-Poissonian fluctuations, indicating a precise balance between proliferative and non-proliferative sister pairs. Finally, by measuring clone size distributions in mice, we showed that stem cell divisions in vivo reproduced the strong symmetry in proliferative behavior between sister cells seen in organoids. As cell proliferation in many tissues follows inherently stochastic 'population asymmetry' mechanisms [6, 75, 99–101], we conjecture that high symmetry in proliferative behavior, controlled through the lineage, may be a more general mechanism to limit proliferation fluctuations.

4.2 | Results

4.2.1 Single-cell tracking of complete crypts in growing intestinal organoids

To examine the dynamics of individual cells within crypts during growth, we used organoids with a H2B-mCherry nuclear reporter (**Fig. 1a**, Video S1) and performed confocal 3D time-lapse microscopy for up to 65 hours at a time resolution of 12 minutes. Cell division events were distinguished by the apical displacement of the mother cell nucleus, followed by chromosome condensation and separation, and finally, basal migration of the daughter cell nuclei (**Fig. 1b**), consistent with epithelial divisions [102]. Custom-written software (**Chapter 2**) was used to track every cell within organoid crypts by recording their nuclei positions in 3D space and time (**Fig. 1c**) and reconstruct lineage trees containing up to six generations (**Fig. 1f**, **Fig. S1**).

To study the cell lineages along the crypt surface, we 'unwrapped' the crypts: First, we annotated the crypt axes at every time point, then projected every cell position onto the surface of a corresponding cylinder, which we then unfolded (**Fig. 1d, e**). This allowed us to visualize the cellular dynamics in a two-dimensional plane defined by two coordinates: the position along the axis and the angle around the axis. We found that lineages starting close to the crypt bottom typically continued to proliferate and expand in cell number, while those further up in the crypt contained cells that no longer divided during the experiment, consistent with stem cells being located at the crypt bottom and terminal differentiation occurring higher up along the crypt axis (**Fig. 1f**). Lineages were also terminated by the death of cells, as observed by their extrusion into the lumen. This fate occurred more often in some lineages compared to others, and the frequency did not depend strongly on position along the crypt-villus axis. At the crypt bottom we also observed a small number of non-dividing cells, suggestive of terminally differentiated Paneth cells. Indeed, these cells typically exhibited the larger cell size and granules typical of Paneth cells. Finally, a small fraction of cells could not be tracked during the experiment, as they moved outside of the field of view, or their fluorescence signal was degraded due to scattering in the tissue.

4.2.2 Control of cell proliferation in organoid crypts

To systematically study proliferation control, we classified cells as proliferating when they divided during the experiment. Cells were classified as non-proliferating when they did not divide for $> 30\mathrm{h}$ or were born $> 60\,\mu\mathrm{m}$ away from the crypt bottom, as such cells rarely divided in our experiments (See Materials and Methods for details). A smaller fraction of cells could not be classified. Some cells were lost from tracking (7%, N=2880 cells). These cells were typically located in the villus region (**Fig. 1f**) and therefore likely non-proliferating. For other cells, the experiment ended before a division could be observed or excluded based on the criteria above (27%). Such cases were particularly prominent in the last 15 h of each time-lapse data set (22%). To analyze proliferation control, we therefore excluded all cells born $< 15\mathrm{h}$ before the end of each experiment, thereby reducing the fraction of unclassified cells to 10%.

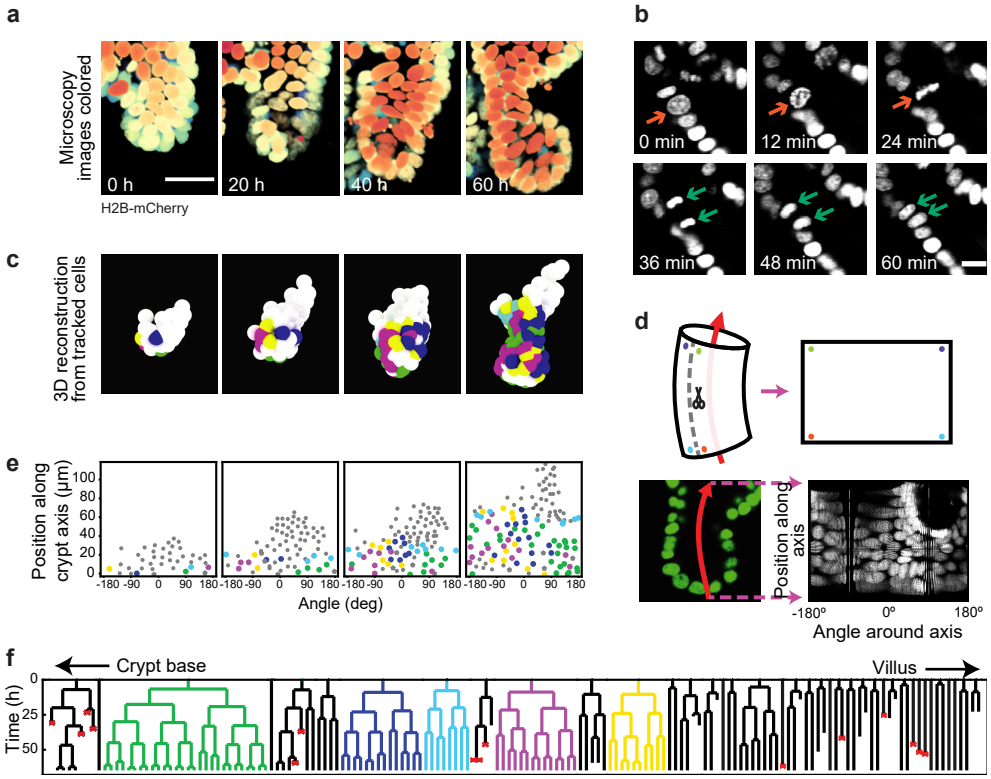


Figure 1 • Time-lapse imaging and single-cell tracking of intestinal organoid crypts

(a) 3D reconstruction of an organoid expressing an H2B-mCherry reporter to visualize individual nuclei. Shown here is the crypt region, with nuclei colored by their depth along the optical axis. Scale bar is 25 µm. **(b)** Snapshots of a cell division event in a crypt. Cell divisions are distinguished by the apical migration of the nucleus followed by chromosome condensation (red arrows). After mitosis, the nuclei of the two newly born cells are displaced basally (green arrows). Scale bar is 10 µm. **(c)** 3D reconstruction of a crypt growing in time using the positions of tracked nuclei. Colors represent cells that belong to the same lineage. **(d)** Illustration of crypt unwrapping. After the crypt-villus axis is annotated (red arrow), tracked cell positions are projected onto the surface of a bent cylinder. The cylinder is then unfolded and its surface is mapped onto a two-dimensional plane defined by the distance along the axis and the angle around the axis. **(e)** Unwrapped representation of the crypt in panel c, where colors represent the same lineages. **(f)** Lineage trees of cells within the crypt in panel c and colored accordingly. Cells in the initial time point are ordered according to their distance to the crypt base. Red crosses indicate cell deaths and incomplete lines indicate cells that could not be accurately traced further due to insufficient fluorescence intensity or movement outside of the field of view.

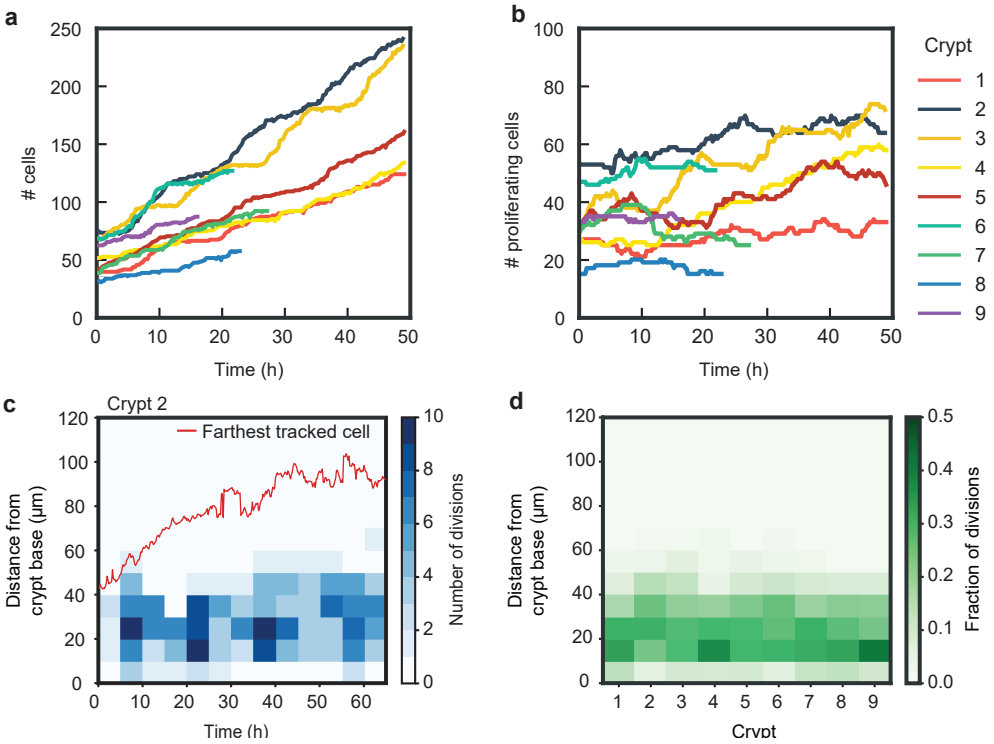


Figure 2 • Control of cell divisions in intestinal organoids

(a) Total number of cells born and **(b)** number of proliferating cells as a function of time for all cell lineages followed in nine tracked crypts. Cells that died were classified as non-proliferating. Note the different scale along the y-axis. Whereas total cell number increases, the number of proliferating cells remains approximately constant. The strongest increase in number of proliferating cells (∼2-fold) was seen in Crypts 3 and 4. **(c)** Number of divisions that occurred at different positions along the crypt axis as a function of time in a single tracked crypt. Red line corresponds to the position of the farthest tracked cell from the crypt base at every time point. Divisions occur in a compartment close to the crypt base, whose size remain constant over time. Apical displacement of the nuclei during mitosis results in few divisions occurring at less than 10 µm from the crypt base. **(d)** Fraction of divisions that occurred at different positions along the crypt axis for all tracked crypts, averaged over the full time course. The size of the proliferative region is similar between crypts, despite differences in the total number of divisions.

Using this classification procedure, we then quantified the total number of cells born in the tracked cell lineages for nine crypts and found a strong (∼4-fold) increase in time (**Fig. 2a**). In contrast, the number of proliferating cells remained approximately constant in time for most crypts (**Fig. 2b**). Two crypts (crypt 3 and 4 in **Fig. 2**) formed an exception with ∼2-fold increase in the number of proliferating cells, an observation that we discuss further below. We then estimated the exponential growth rate α for each crypt, by fitting the dynamics of total number of cells born and proliferating cell number to a simple model of cell proliferation (discussed in **Chapter 3** as the uniform model), where proliferating cells divide randomly into proliferating and non-proliferating cells. In this model, the number of proliferating and non-proliferating cells increase on average by α and $1-\alpha$ per cell division, respectively (Materials and Methods). Apart from Crypts 3 and 4, that displayed growth ($\alpha = 0.3$), the remaining crypts showed a low growth rate, $\alpha = 0.05 \pm 0.07$ (**Fig. S2a-c**), indicating that birth of proliferating and non-proliferating cells was balanced on average. We then quantified the magnitude of fluctuations in the number of proliferating cells, N. Calculations of birth-death models of cell proliferation show that, without any control, the standard deviation of the proliferating cell number grows in time without bounds as $\sigma_D \sim \sqrt{Nt}$ [92]. In models without exponential growth, with proliferating cells born at constant rate, fluctuations are reduced: they are constant in time and Poissonian, $\sigma_D \sim \sqrt{N}$ [103]. In models where exponential growth was controlled by homeostatic feedback loops, fluctuations were further reduced to sub-Poissonian: $\sigma_D < \sqrt{N}$ [78]. Using the same measures, we found here that most crypts exhibited sub-Poissonian fluctuations (**Fig. S2d**), implying the presence of a mechanism to limit fluctuations in proliferating cell number. Finally, we quantified the frequency of cell divisions along the crypt axis. Notably, divisions occurred in a region below 60 µm from the crypt base throughout the experiment, even as the crypts grew significantly (**Fig. 2c**), indicating that the size of the proliferative region was constant in time. Moreover, the proliferative region was found to have a similar size in all analyzed crypts (**Fig. 2d**), even though crypts varied both in size, as measured by diameter (30-50 µm, **Fig. S2e**), and number of proliferating cells (**Fig. 2b**). Overall, these results show that crypts by themselves are already capable of a specific form of homeostasis, namely maintaining a stationary number of proliferating cells that occupy a region of the crypt of constant size.

4.2.3 Symmetry of proliferative behavior between sister cells

To examine the origin of the observed balance between the birth of proliferating and non-proliferating cells, we first examined whether cell proliferation or cell death were correlated between sisters, for all observed sister pairs S_1 and S_2 (**Fig. 3a**). Strikingly, we found that the decision to divide or not was highly symmetrical between sisters. In particular, occurrences where one sister divided but the other not were rare (2%) compared to cases where both divided or stopped dividing (74%). This correlation was also apparent by visual inspection of individual lineages, as sisters showed the same division behavior (**Fig. 3b**, top). Indeed, if we ignore cell death, the fraction of pairs with symmetric proliferative outcome was high (97%) and could not be explained by sister cells making the decision to proliferate or not independently ($P < 10^{-5}$, bootstrap simulation, Materials and Methods). We also compared lineage dynamics

between all cousin pairs C_1 and C_2 (**Fig. 3a**). While we indeed found a significantly increased fraction (9%) of cousin pairs with asymmetrical division outcome, i.e. C_2 dividing and C_1 not, compared to sister pairs (2%), this fraction was still low, indicating that symmetric outcomes also dominated for cousins.

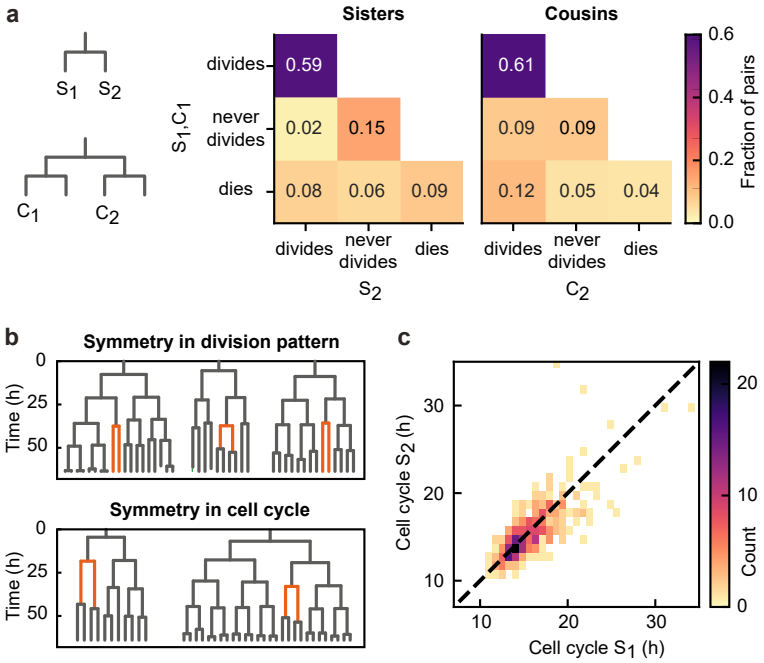


Figure 3 • Symmetry of lineage dynamics between sister cells
(a) Correlations in division patterns between sister (S_1,S_2) and cousin cells (C_1,C_2) (n=1004 and 1304 sister and cousin cell pairs). Most sister pairs show symmetrical outcomes, with most pairs consisting of sisters that either both divide or both cease dividing. Cell death occurs at low frequency and impacts sister cells asymmetrically. Symmetrical outcomes are still dominant for cousins, but the fraction of pairs that exhibited asymmetric proliferative outcomes (C_1 never divides, C_2 divides) was significantly increased in cousins compared to sisters ($P = 2.4{\cdot}10^{-7}$, Pearson's Chi Square Test). **(b)** Representative examples of measured lineages highlighting pairs of sister cells (orange) that differ in lineage dynamics from their more distant relatives (black), either in terms of proliferative behavior (top) or cell cycle duration (bottom). **(c)** Duration of sister cell cycles plotted against each other for pairs in which both sisters divided. Cell cycle duration is strongly correlated between sisters (R=0.80).

We found that symmetry between sisters did not only impact proliferation arrest, but also cell cycle duration: when a cell exhibited a longer-than-average cell cycle, this was typically mirrored by a similar lengthening of the cell cycle of its sister (**Fig. 3b**, bottom). Indeed, cell cycle duration was strongly correlated between sisters (R=0.8, **Fig. 3c**), even as cell cycle duration showed a broad distribution among tracked cells

(**Fig. S3**). In contrast, cell death was not symmetric between sisters, as the fraction of pairs where both cells died (9%) was smaller than the fraction of pairs where only a single sister died (14%, **Fig. 3a**).

When examining all sister pairs in our data set, pairs of dividing sisters (59%) outnumber pairs of non-dividing sisters (15%), which appeared at odds with the observation that in most crypts the number of proliferating cells remain approximately constant (**Fig. 2b**). This apparent mismatch was due to the exclusion of sister pairs where the proliferative state could not be classified in one sister or both (**Fig. S4**), as the majority of these unclassified cells were likely non-proliferating. Indeed, when we restricted our sister pair analysis to the cells of crypts with $\alpha \approx 0$ in **Fig. 2a, b** (Crypts 1,2,5,7-9) and furthermore assumed that all unclassified cells were non-proliferating, we found that now proliferating sister pairs (43%) are approximately balanced by non-proliferating sisters (40%, **Fig. S4**).

4.2.4 Symmetry between sisters minimizes fluctuations in a cell proliferation model

In principle, any combination of (a)symmetric divisions would yield a constant number of proliferating cells on average, as long as the birth of proliferating and non-proliferating cells is balanced. However, as we have shown in **Chapter 3**, dominance of symmetric divisions had a function specifically in controlling *fluctuations* in the number of proliferating cells.

We will reuse the notation from **Chapter 3**, with α defined as the growth rate and ϕ is defined as the asymmetry fraction (**Fig. 4a**). α ranges from -1 to 1, with all divisions resulting in non-proliferating daughters for $\alpha = -1$, and all divisions resulting in two proliferative daughters for $\alpha = 1$. ϕ ranges from 0 to 1. At $\phi = 1$, all divisions produce cells with the same proliferation outcome, either both non-proliferating or both proliferating, depending on the value of α. At $\phi = 0$, no divisions with the same proliferation outcome are allowed, which means that each division must produce one proliferating daughter, and one non-proliferating. As a result, the number of proliferating cells stays constant, and obtaining $\phi = 0$ is only possible for $\alpha = 0$.

For the uniform model, in which no space was defined, we found that maximizing the number of asymmetric divisions resulted in least number of fluctuations (**Fig. 4b, c**). However, this was not the case for the two-compartment model, which is a model where the space was divided into a niche compartment of fixed size and a differentiation compartment of unbounded size. Now, the optimal scenario was to let the niche compartment have a maximal growth rate ($\alpha_n = 1$), and the differentiation compartment a minimal growth rate ($\alpha_d = -1$). This required *minimizing* asymmetric divisions (**Fig. 4d-f**).

Similar low fluctuations were found for a broader range of α_d, provided that $\alpha_n \approx 1$. When we quantified the magnitude of fluctuations versus the average number of proliferating cells, we found that fluctuations for the suboptimal scenarios 1-3 (**Fig. 4d-f**) were larger than those expected for a Poisson birth-death process (**Fig. S2d**). In contrast, fluctuations in the optimal scenario 4 were similar to the low fluctuations we

observed experimentally.

Our simulations also provided an intuitive explanation for this global minimum. A bias $\alpha_n = 1$ is only reached when the birth of non-proliferating cells in the niche compartment, by symmetric or asymmetric divisions, is fully avoided. In this limit, all cells in this compartment are proliferating, meaning that fluctuations in the niche compartment are entirely absent, with the only remaining fluctuations due to cells ejected from the niche compartment that subsequently divide in the differentiation compartment. Consistent with this explanation, we found that fluctuations in N increased when more symmetric divisions in the niche compartment generated non-proliferating daughters (scenarios 1 and 2). Finally, we note that the well-established neutral drift model of symmetrically dividing stem cells in a niche of fixed size [6, 76] fails to reproduce the high symmetry in proliferation we experimentally observe between sister cells (**Fig. S5e**), indicating that the size constraint of a niche is by itself not sufficient to generate this symmetry. In conclusion, our simulations show that the dominance of symmetric divisions we observed experimentally might function to minimize fluctuations in cell proliferation.

4.2.5 Symmetry of proliferation is independent of Paneth cell distance

Our results raised the question how the strong symmetry in proliferative behavior between sister cells is generated. Stem cell maintenance and cell proliferation is control by signals such as Wnt and EGF, that in organoids are locally produced by Paneth cells [94, 104]. The symmetry between sister cells could therefore be explained by these sisters having a similar position relative to Paneth cells, leading them to experience a near identical environment in terms of proliferative signals. Alternatively, the proliferative behavior of sister cells could be controlled through the lineage, by their mother. In this case, symmetric proliferative behavior would even be seen in sisters that differ in position relative to Paneth cells. To differentiate between these two scenarios, we performed lysozyme staining after time-lapse imaging to retrospectively identify Paneth cells in our tracking data. Using crypt 'unwrapping' (**Fig. 1d, e**), we calculated for each cell and each time point the link distance δ to the closest Paneth cell (Materials and Methods, **Fig. S6a, b**, Video S2), i.e. the number of cells between the cell of interest and its closest Paneth cell, allowing us to examine proliferative behavior as function of distance to Paneth cells.

Paneth cell-derived Wnt ligands form gradients that only penetrate 1-2 cells into the surrounding tissue [104], suggesting that the steepest gradient in proliferative signals is found in close proximity to the Paneth cell. We therefore selected all dividing mother cells directly adjacent to a Paneth cell ($\delta_M = 0$) and examined their daughters. These sister pairs varied in Paneth cell distance ($\delta_{1,2} \approx 0 - 2$, **Fig. 5a**, **Fig. S6c**), with differences between sisters ($\delta_1 \neq \delta_2$) seen in 42% of pairs. We classified each sister pair as asymmetric in outcome, when only one sister continued proliferating, or symmetric (**Fig. 5b**). In the latter case, we distinguished between symmetric pairs where both sisters divided and those were both stopped proliferating. We found that most daughters cells divided again, even though a small fraction ceased division even in close proximity

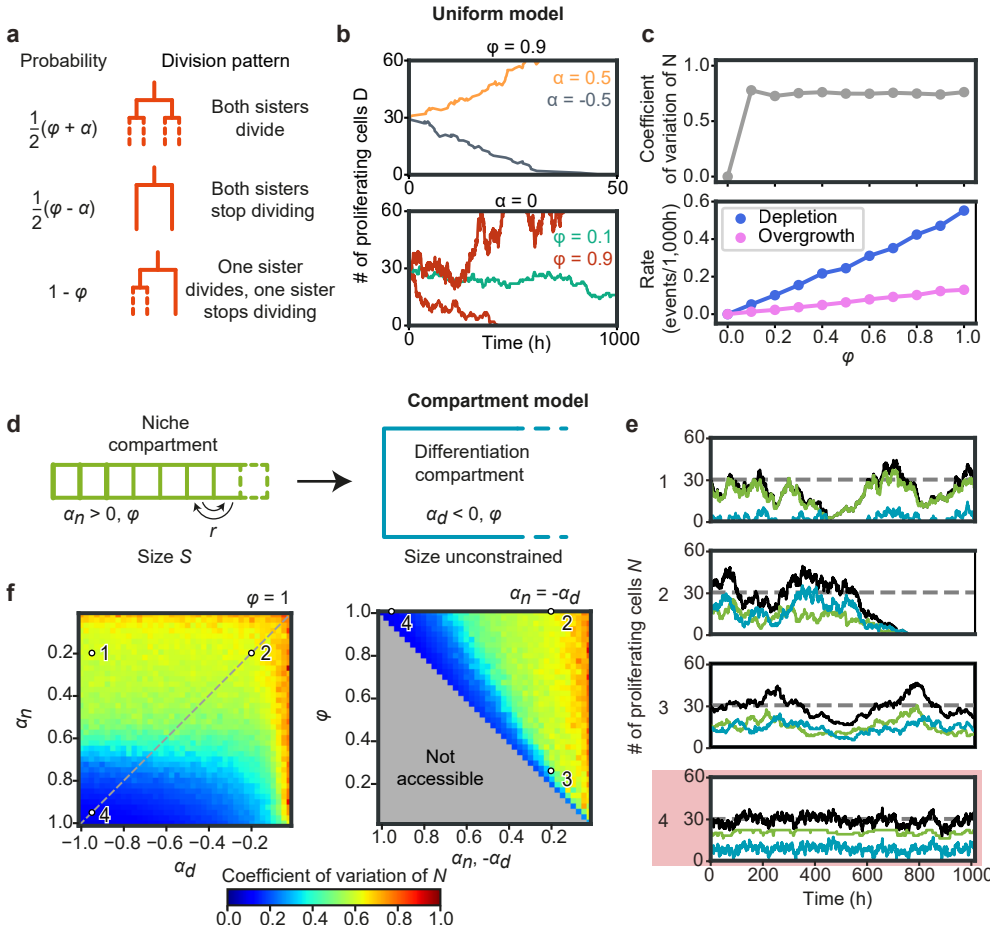


Figure 4 • Cell number fluctuations in stem cell models

(a) Division types in the 'uniform' stem cell model. **(b)** Number of proliferating cells, N, as function of time for different values of α (top) and ϕ (bottom). **(c)** Coefficient of variation (standard deviation divided by mean) of N as a function of ϕ (top panel) and the probability of depletion ($N = 0$, blue) or overgrowth ($N \geq 150$, pink) events (bottom panel), for the 'Uniform' model with $a = 0$. **(d)** 'Compartment' model. Cells divide according to panel a, but now the tissue is divided in a niche compartment n, with $\alpha_n > 0$, and a differentiation compartment d, where $\alpha_d < 0$. Both compartments have the same ϕ. In the niche compartment, the total number of cells cannot exceed S, so that upon cell division the distalmost cell moves into the differentiation compartment. **(e)** Number of proliferating cells as a function of time in the niche (green) and differentiation compartment (blue). Each panel's number refers to the parameter sets shown in panel f. The parameter set with lowest fluctuations is outlined in red. **(f)** Coefficient of variation of N. Left panel shows the effects of varying the growth rates of both compartments when all divisions are symmetric ($\phi = 1$) and right panel of varying the degree of symmetry when both compartments have opposite growth rates ($\alpha_n = -\alpha_d$, dashed line in top panel). Fluctuations are minimized for $\alpha_n, -\alpha_d = 1$ and $\phi = 1$, i.e. only symmetric divisions.

to Paneth cells (**Fig. 5b**, **Fig. S6d**) However, whether cells divided or not was fully symmetric between sister pairs, even when one cell remained adjacent to a Paneth cell ($\delta_1 = 0$) while the other lost contact ($\delta_2 > 0$). This also held for the few pairs where Paneth cell distance differed most between sisters ($\delta_1 = 0$, $\delta_2 = 2$).

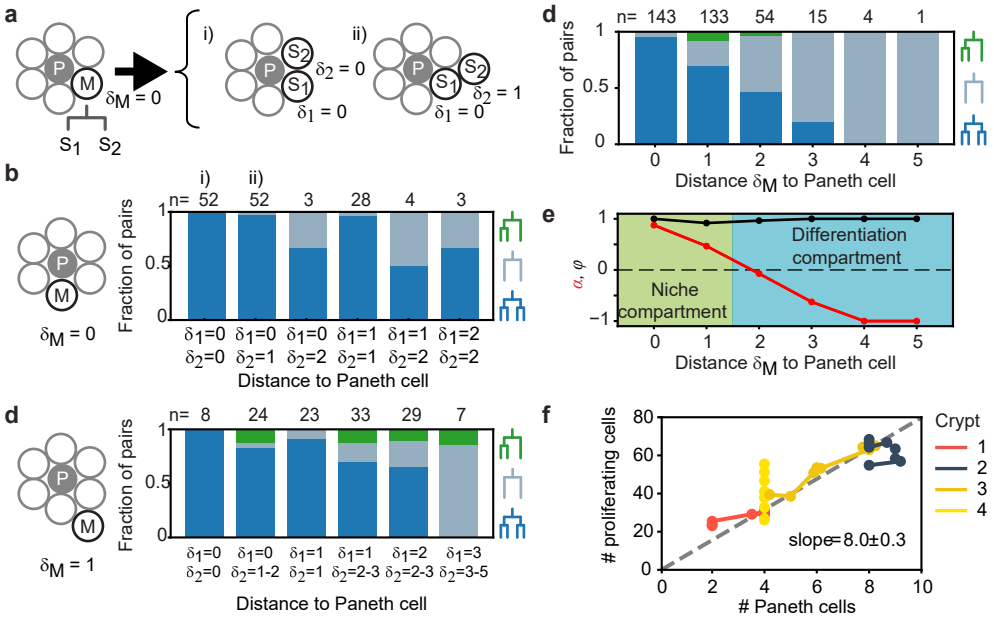


Figure 5 • Impact of Paneth cell distance on proliferation
(a) Dependence of proliferation on contact with Paneth cells. We examined cases where a mother cell (M) that touched a Paneth cell (P) divided into sister cells, S_1 and S_2, that either retained or lost Paneth cell contact. Here, the link distances δ_1 and δ_2 represents the number of cells between each sister and its closest Paneth cell. **(b)** Probability that both cells divide (blue), neither cell divides (orange) or only a single cell divides (green) for all sister pairs S_1 and S_2 of which the mother touched a Paneth cell ($\delta_M = 0$). Sister pairs exhibited full symmetry in proliferative behavior, even when distance to the Paneth cell differed between sisters ($\delta_1 \neq \delta_2$). **(c)** Same as panel b but for a mother cell positioned one cell away from the Paneth cell ($\delta_M = 1$). More daughter cells cease proliferation. While the fraction of pairs where only one sister divides increases, most sisters exhibit symmetric behavior. **(d)** Probability of each division pattern as function of Paneth cell distance of the mother cell. **(e)** Proliferative bias α and degree of symmetry ϕ as a function of Paneth cell distance. The observed values of α define a proliferative niche (green, $\alpha \approx 1$) and non-proliferative differentiation (blue, $\alpha < 0$) compartment, with the former corresponding approximately to the first two rings of cells surrounding the Paneth cell. **(f)** Number of proliferating cells as function of Paneth cell number. Time-courses for individual crypts were divided into 5-hour intervals (markers), for which average cell numbers were calculated. Apart from Crypt 4, Paneth cell number correlated well with number of proliferating cells, even when Paneth cell number increased in time due to divisions. Dashed line is a linear fit to the data.

When we instead examined mother cells that just lost contact with a Paneth cell ($\delta_M = 1$), we found that their offspring stopped proliferating more frequently (**Fig. 5c**).

While here we did find a substantial fraction of sister pairs with asymmetric outcome, for most pairs the outcome was still symmetric (92% of pairs), even for pairs that differed considerably in relative distance to the Paneth cell. Sister pairs with asymmetric outcome occurred more frequently for pairs with different Paneth cell distance ($\delta_1 \neq \delta_2$). For these pairs, however, the non-proliferating cell was the sister closest to the Paneth cell about as often as it was the more distant (5 and 3 pairs, respectively), indicating that position relative to the Paneth cell had little impact on each sister's proliferative behavior. Overall, these results show that the symmetry of proliferative behavior between sisters did not reflect an underlying symmetry in distance to Paneth cells, thus favoring a model where this symmetry is controlled by the mother cell.

4.2.6 Paneth cells control proliferative bias

Even though the proliferative behavior of sisters was not explained by their relative Paneth cell distance, we found that the bias towards proliferating daughters was clearly reduced when the Paneth cell distance of the mother increased (**Fig. 5b, c**). Our simulations showed that both division symmetry and proliferative bias are important parameters in controlling fluctuations in the number of proliferating cells, with fluctuations minimized when most divisions are symmetric ($\phi \approx 1$), biased strongly towards producing two proliferating daughters in one compartment ($\alpha \approx 1$) and two non-proliferating daughters in the other ($\alpha \approx -1$). To compare our experiments against the model, we therefore measured the frequency of each division class as function of the mother's Paneth cell distance, averaging over all positions of the daughter cells (**Fig. 5d**). Overall, cells had a broad range of Paneth cell distances ($\delta = 0{-}10$). Close to Paneth cells ($\delta \leq 1$), most divisions generated two proliferating daughters, while further away ($\delta > 1$), the majority yielded two non-proliferating cells. Asymmetry was rare and only occurred for $\delta = 1{-}2$. No divisions were seen for $\delta > 5$. When we used these measured frequencies to calculate α and ϕ as a function of Paneth cell distance (**Fig. 5e**), we found good agreement with the parameter values that minimized fluctuations in the model, with a niche compartment of strong proliferation close to Paneth cells ($\alpha = 0.67$, $\delta \leq 1$) and a non-proliferative compartment beyond ($\alpha = -0.67$), while almost all divisions were symmetric ($\phi = 0.98$).

The compartment model also predicted that the number of proliferating cells increases linearly with size S of the niche compartment. Above, we observed that the number of proliferating cells differed between crypts (**Fig. 2b**). We therefore examined whether variation in number of proliferating cells between crypts could be explained by differences in Paneth cell number, in those crypts where we identified Paneth cells by lysozyme staining (Crypts 1-4). For the crypts that maintained a constant number of proliferating cells in time (Crypts 1-2), we found that differences in number of proliferating cells were well explained by differences in Paneth cell number. Moreover, in crypts with increasing number of proliferating cells (Crypts 3-4), we found that for Crypt 3 this change could be explained by an increase of Paneth cell number, due to cell divisions that generated Paneth cell sisters. In Crypt 4, however, proliferating cells increased in number without apparent Paneth cell proliferation. This crypt appeared to undergo crypt fission [105] at the end of the experiment, suggesting that during fission cell proliferation is altered without concomitant changes in Paneth cell number. For

Crypts 1-3, the relationship between number of proliferating and Paneth cells was well-fitted by a linear function (**Fig. 5f**), consistent with the compartment model. The fitted slope of this line indicates that one Paneth cell maintains ∼8 proliferating cells. This agrees with the observation in **Fig. 5d** that divisions are strongly biased towards proliferation only for cells within the first and second ring of cells around each Paneth cell. Taken together, these results show that Paneth cells control proliferation by tuning the proliferative bias of divisions that are otherwise symmetric in proliferative outcome.

4.2.7 In vivo lineage tracing confirms symmetric proliferative behavior of sister cells

Finally, we asked whether the symmetry of proliferative behavior between sister cells was also observed in intestinal (stem) cells *in vivo*. Studying lineage dynamics with the high spatial and temporal resolution we employed here is currently impossible *in vivo*. However, we found that clone size distributions, which can be measured *in vivo*, exhibited a clear signature consistent with symmetric divisions. Specifically, clone size distributions of lineages generated by the compartment model showed that enrichment of even-sized clones depended strongly on a high frequency of symmetric divisions (**Fig. 6a**). When we quantified clone size distributions for our organoid lineage data, by counting the number of progeny of each cell at the end of a 40 h time window, while sliding that window through our ∼60 h data set, we indeed found that even clone sizes were strongly enriched compared to odd clone sizes (**Fig. 6b**). Both for organoid and model data, we still observed odd clone sizes even when virtually all divisions were symmetric in proliferative behavior. This reflected variability in the cell cycle duration, with odd clone sizes typically resulting from symmetric divisions where one daughter had divided, but the other not yet.

To measure clone size distributions in the small intestine in vivo, we stochastically induced heritable tdTomato expression in Lgr5+ stem cells using $Lgr5^{\text{EGFP-ires-CreERT2}}$; $R26_{\text{LSL-tdTomato}}$ mice. We activated Cre-mediated recombination by tamoxifen and examined tdTomato expression after 60 hours, similar to the timescale of our organoid experiments, and imaged crypts with 3D confocal microscopy. Cre-activation occurred in one cell per ∼10 crypts, indicating that all labeled cells within a crypt comprised a single clone. Indeed, we typically found a small number of tdTomato+ cells per crypt, of which most also expressed Lgr5-GFP (**Fig. 6c**). We then counted the number of tdTomato+ cells per crypt to determine the clone size distribution and found a clear enrichment of even clone sizes (**Fig. 6d**), with the overall shape of the distribution similar to that measured in organoids. Overall, these results indicated a dominant contribution of divisions with symmetric proliferative outcome also in the lineage dynamics of Lgr5+ stem cells in vivo.

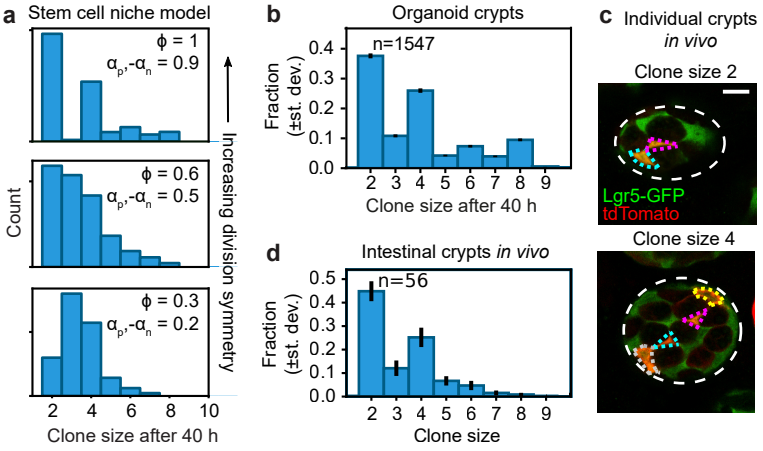


Figure 6 • Clone size distributions reveal symmetry of proliferative behavior

(a) Clone size distributions calculated for the Compartment model, for different degrees of division symmetry ϕ. Top panel corresponds to parameters that minimize fluctuations in number of proliferating cells. For high division symmetry, $\phi \approx 1$, even clone sizes are enriched compared to odd clone sizes. **(b)** Clone size distributions calculated for the lineage data obtained in organoids, for a sliding window of 40 h. Even clone sizes are enriched, consistent with the observed dominance of divisions with symmetric proliferative outcome. Error bars indicate standard deviation calculated using a bootstrapping approach (Materials and Method). **(c)** Examples of individual crypts found in vivo, displaying clone size 2 (top) and 4 (bottom). Crypts are viewed from the bottom, with individual cells belonging to a tdTomato+ clone (red) outlined. Scale bar is 10 µm. **(d)** Clone size distributions measured in vivo 60 h after induction of Cre-mediated recombination in small intestinal crypts of *Lgr5*$^{\text{EGFP-ires-CreERT2}}$; *R26*$_{\text{LSL-tdTomato}}$ mice (n=160 crypts). Even clone sizes 2 and 4 are enriched compared to odd clone sizes 3 and 5. Error bars indicate standard deviations calculated using a bootstrapping approach.

4.3 | Discussion

Self-renewing tissues exhibit homeostasis at multiple levels, such as overall tissue morphology, total cell number and the relative frequency of different cell types. To prevent exponential growth or tissue atrophy, the birth of each proliferating cell must be balanced by the loss of another through terminal differentiation. Experiments in a range of systems indicate that this is achieved through 'population asymmetry', with each cell making the decision to proliferate or not in a stochastic manner and this balance only achieved averaged over the entire population [6, 70, 75, 100]. However, our simulations showed that even though 'population asymmetry' ensures a constant pool of proliferating cells on average, its inherently stochastic nature can cause strong fluctuations in proliferating cell number, even resulting in their full depletion (**Fig. 4**d-f). This raises the question how these fluctuations are controlled.

We addressed this by a combined experimental and theoretical approach. We tracked all cell movements and divisions in the crypts of growing intestinal organoids, to reconstruct the full lineage of these crypts up to six generations (**Fig. 1**). These data showed that the number of proliferating cells in most organoid crypts was approximately stationary, with small, i.e. sub-Poissonian fluctuations in their number, while non-proliferating cells were born at a constant rate (**Fig. 2**, **Fig. S2**), an indication of homeostatic control of cell proliferation that also limits fluctuations. That intestinal organoids exhibited homeostasis is notable, as organoid culture completely lacks surrounding tissue, such as the mesenchyme, that provides key signals regulating stem cell fate and proliferation [93, 95], and shows that this form of homeostasis is inherent to the epithelium itself.

Our simulations showed that the fluctuations in proliferating cell number depended strongly on the relative proportion of divisions with symmetric proliferative outcome (either two proliferating or two non-proliferating daughters) and asymmetric outcome (one proliferating and one non-proliferating daughter), with small, sub-Poissonian cell number fluctuations only seen when most divisions had symmetric outcome (**Fig. 4**, **Fig. S2**). So far, the relative contribution of these three divisions patterns in the intestine could only be inferred indirectly from static measurements, leading to conflicting results [6, 76, 83, 106]. Here, we used direct measurements of cell dynamics in time to unambiguously identify the proliferative state of successive generations of cells. These measurements show that virtually all cell divisions (¿90%) showed symmetric proliferative behavior, generating either two proliferating or two non-proliferating sisters (**Fig. 3**). Clone size distributions calculated based on our measured lineage data in organoids showed that this symmetry in proliferative behavior between sister cell gave rise to an enrichment of even clone sizes (**Fig. 6**). Using short-term lineage tracing experiments in the mouse small intestine, we found that single Lgr5+ stem cells also gave rise to more even-sized than odd-sized clones, indicating that divisions that are symmetric in proliferative behavior indeed also dominate stem cell proliferation in vivo.

The symmetry in proliferative behavior we observe between sister cells could arise because both cells experience a highly similar environment, in terms of proliferative signals, or rather indicate control of cell proliferation through the lineage, by the mother

cell. The current models of stem cell dynamics in the intestinal crypt favor a strong role for position relative to the stem cell niche, formed in organoids by Paneth cells, and a minor role, if any, for control of cell proliferation through the lineage [6, 18, 76]. We found that sister cells exhibited symmetric proliferative behavior, even when sisters differed in distance to Paneth cells (**Fig. 5**), the sole source of proliferative Wnt signals in intestinal organoids [94]. This result implies control of proliferation by the mother cell rather than by each daughter's position in the stem cell niche. Our simulations provided a potential function for the predominance of divisions with symmetric proliferative outcome. When the tissue was subdivided into compartments of low and high cell proliferation, with the latter resembling the stem cell niche, we found that fluctuations in the number of proliferating cells were virtually eliminated, provided that cell divisions were symmetric, with all divisions generating two proliferating daughters in the niche compartment and two non-proliferating daughters outside (**Fig. 4**). Consistently, in our experiment we found that frequency of mother cells generating two proliferating rather than two non-proliferating daughters decreased with the mother's distance to the closest Paneth cell (**Fig. 5**). Taken together, our results suggest a model where differences in proliferative behavior emerges in the cell lineage over at least two generations: while a mother cell division generates two daughters with the same proliferative behavior, these daughters might subsequently generate granddaughters that differ in proliferative behavior, depending on each daughter's position relative to the Paneth cells. This is consistent with our observation that the symmetry of proliferative behavior between cousins is reduced significantly compared to sisters (**Fig. 3**).

In **Chapter 3**, we used mathematical modelling to explore the dependence of fluctuations in cell proliferation on the degree of symmetry in cell division outcome. In that chapter, we arrived at a two-compartment model. By comparing the simulations to the experimental data of this chapter, we showed that the simulations matched key features of our experiments (**Fig. 4d-f**). The simulations reproduced the observed low, sub-Poissonian fluctuations in number of proliferating cells (Figure 2-figure supplement 1D), but only when division symmetry was high, as we also observed experimentally. In contrast, high symmetry increased fluctuations in a spatially uniform stem cell model (**Fig. 4b, c**), while a standard neutral drift model of a stem cell niche [6, 76] failed to reproduce the observed symmetry in outcome (**Fig. S5**). Our model also reproduced the observed correlation in proliferative state between cousin cells (**Fig. S5**), explaining it as arising from closely-related cells having similar location in the tissue and therefore similar probability of leaving the stem cell niche. Finally, it predicted the observed division of the tissue in a compartment where most divisions generated proliferating cells (close to Paneth cells) and one where divisions mostly generated non-proliferating daughters (away from Paneth cells) (**Fig. 5**). However, the simplified nature of our model also poses limits. First, the observed transition from proliferating to non-proliferating daughter cells was more gradual than predicted by the model, indicating that each divisions proliferative outcome depended on space in a more complex manner than captured by the model. Second, the existence of compartments and the degree of symmetry in division outcome is imposed externally by the model rules. It will be interesting to examine whether simple mathematical models can explain how these properties emerge from the internal cellular states, long-range signaling pathways and local cell-cell interactions involved in intestinal homeostasis [57, 70].

Precise control of cell proliferation is key to homeostasis. It has been proposed that cells may sense cell density, either by chemical signals or mechanical cues, and decrease cell proliferation (known as contact inhibition) if the cell number is too high [78, 90, 107], thereby ensuring homeostasis of and minimize fluctuations in the number of proliferating cells. Here, we provide a mechanism that achieves this without explicit sensing of cell density. Instead, it relies on the dominance of divisions symmetric in proliferative behavior of the daughter cells, coupled with the organization of a tissue in a proliferative niche (stem and transit amplifying cell) compartment, and a non-proliferative differentiation compartment. Such an organization is found widely, e.g. in the skin, hair follicles, testis among others [89]. In all these tissues, homeostasis of and minimizing fluctuations in the number of proliferating and differentiated cells must be essential. Hence, we speculate that the model we propose here, which exploits proliferative symmetry between sister cells to minimize fluctuations, is conserved more broadly and relevant to diverse tissue systems.

4.4 | Materials and methods

4.4.1 Organoid line

H2B-mCherry murine intestinal organoids were a gift from Norman Sachs and Joep Beumer (Hubrecht Institute, The Netherlands).

4.4.2 Organoid culturing

Organoids were cultured in 24-well cell culturing plates, which were placed in an incubator at 37 °C and 5% CO_2. The organoids were seeded in 40 µL gel consisting of basement membrane extract (BME2, Trevingen) mixed with Advanced DMEM/F-12 in a 2:1 to 3:1 volume ratio. The gel was overlaid with 0.4 mL growth medium. This medium consisted of Advanced DMEM/F-12 (Life Technologies) with the following added ingredients:

- murine recombinant epidermal growth factor (EGF 50 ng/mL, Life Technologies)
- murine recombinant Noggin (100 ng/mL, Peprotech)
- human recombinant R-spondin 1 (500 ng/mL, Peprotech)
- n-Acetylcystein (1 mM, Sigma-Aldrich)
- N2 supplement (1x, Life Technologies)
- B27 supplement (1x, Life Technologies)
- Glutamax (2 mM, Life Technologies)
- HEPES (10mM, Life Technologies)
- Penicillin/Streptomycin (100 U/mL 100 µg/mL, Life Technologies)

The organoids were passaged every week, after which they were refed twice during the next seven days: once 2 to 3 days after passaging and once 4 to 5 days after passaging.

During refeeding, the growth medium was replaced. During passaging, the organoids were mechanically broken up into smaller organoids, to prevent the organoids from growing too large and therefore dying. The passaging was carried out as follows.

The organoids from two wells were collected into a 5 mL Eppendorf tube using 2 mL of Advanced DMEM/F-12 medium. The organoids were broken up by pipetting up and down with a glass pipette, of which the opening was narrowed using a flame. An additional 2 mL of Advanced DMEM/F-12 was added and the tube was centrifuged for 5 minutes at 800 rpm (320 rcf). Afterwards, the supernatant was aspirated. A second centrifugation round was carried out: 4 mL new Advanced DMEM/F-12 medium was added, and the tube was centrifuged for an additional 5 minutes at the same speed. The supernatant was again removed, after which the organoids were seeded in gel on wells of a new culturing plate. After leaving the plates in an incubator for 20 to 30 minutes, the growth medium was added.

4.4.3 Time-lapse imaging

Mechanically dissociated organoids were seeded in imaging chambers one day before the start of the time-lapse experiments. Imaging was performed using a scanning confocal microscope (Nikon A1R MP) with a 40x oil immersion objective (NA = 1.30). 30 z-slices with 2 µm step size were taken per organoid every 12 minutes. Experiments were performed at 37°C and 5% CO2. Small but already formed crypts that were budding perpendicularly to the objective were selected for imaging. Imaging data was collected for 3 independent experiments.

4.4.4 Fluorescent staining

After time-lapse imaging, organoids were fixed with 4% formaldehyde (Sigma) at room temperature for 30 minutes. Next, they were permeabilized with 0.2% Triton-X-100 (Sigma) for one hour at 4°C and blocked with 5% skim milk in TBS at room temperature for one hour. Subsequently, organoids were incubated in blocking buffer containing primary antibody (rabbit anti-lysozyme 1:800, Dako #A0099) overnight at 4°C, and then incubated with secondary antibody (anti-rabbit conjugated to Alexa Fluor©405 1:1,000, Abcam #ab175649) at room temperature for one hour. Afterwards, they were incubated with wheat germ agglutinin (WGA) conjugated to CF©488A (5 µg/ml Biotium) at room temperature for two hours, followed by incubation with RedDot™1 Far-Red Nuclear Stain (1:200, Biotium) at room temperature for 20 minutes. Finally, organoids were overlaid with mounting medium (Electron Microscopy Sciences). The procedure was performed in the same imaging chambers used for time-lapse imaging in order to maintain organoids in the same position. Imaging was performed with the same microscope as previously described. Note that WGA stains both Paneth and Goblet cells, but the lysozyme staining allowed the unequivocal distinction between them.

4.4.5 Single-cell tracking

Cells were manually tracked by following the center of mass of their nuclei in 3D space and time using custom-written image analysis software. Each cell was assigned a unique label at the start of the track. For every cell division, we noted the cell labels of the mother and two daughter cells, allowing us to reconstruct lineage trees. We started by tracking cells that were at the crypt bottom in the initial time point and progressively tracked cells positioned towards the villus region until we had covered all cells within the crypt that divided during the time-lapse recording. We then tracked at least one additional row of non-dividing cells positioned towards the villus region. Cell deaths were identified either by the extrusion of whole nuclei into the organoid lumen or by the disintegration of nuclei within the epithelial sheet. Only crypts that grew approximately perpendicularly to the imaging objective and that did not undergo crypt fission were tracked. During imaging, a fraction of cells could not be followed as they moved out of the microscope's field of view or moved so deep into the tissue that their fluorescence signal was no longer trackable. Because these cells were predominantly located in the villus region, where cells cease division, this likely resulted in underestimation of non-proliferating cells. Data was discarded when a large fraction ($> 25\%$) of the cells in the crypt move out of the imaged volume.

4.4.6 Classifying cell state

To classify cells as either proliferating or non-proliferating, we followed the following procedure. Defining proliferating cells was straightforward, as their division could be directly observed. As for non-proliferating cells, we applied two criteria. First, cells were assigned as non-proliferating when they were tracked for at least 30 hours without dividing. This was based on our observation that cells cycle times longer than 30 hours were highly unlikely ($P = 7.1 \cdot 10^{-7}$, from fit of skew normal distribution, **Fig. S3**). However, we were not able to track all cells for at least 30 hours, as cells moved out of the field of view during the experiment or, more frequently, because they were born less than 30 hours before the end of the experiment. In this case, we defined a cell as non-proliferating if its last recorded position along the crypt axis was higher than 60 µm, as almost no divisions were observed beyond this distance (**Fig. 2**). Finally, cells were assigned as dying based on their ejection from the epithelium, while the remaining unassigned cells were classified as undetermined and not included in the analysis.

We tested the accuracy of this approach as follows. In data sets of ¿60 hours in length, we selected the subset of all cells for which we could with certainty determine proliferative state, either because they divided or because they did not divide for at least 40 hours. We then truncated these data sets to the first 40 hours, which reduced the number of cells whose proliferative state we could identify with certainty, and instead determined each cells proliferative state based on the above two criteria. When we compared this result with the ground truth obtained from the > 60 hour data sets, we found that out of 619 cells, we correctly assigned 141 cells as non-proliferative and 474 as proliferative. Only 4 cells were incorrectly assigned as non-dividing, whereas they were seen to divide in the > 60 hour data sets.

Estimation of significance of symmetric divisions To estimate whether the experimentally observed fraction of sisters with symmetric division outcome could be explained by sister deciding independently to proliferate further or not, we used a bootstrapping approach. In our experimental data, we identified $n = 499$ sister pairs, where the proliferative state of each cell was known and excluding pairs where one or both sisters died. In this subset of sisters the probability of a cell dividing was found to be $p = 0.79$. For $N = 105$ iterations, we randomly drew n sister pairs, which each cell having probability p to be proliferative and 1-p to cease proliferation. For each iteration, we then calculated the resulting symmetry fraction ϕ. This resulted in a narrow distribution of ϕ with average ± st.dev. of 0.67 ± 0.02, well-separated from the experimentally observed value of $\phi = 0.97$. In particular, none of 105 iterations resulted in a value $\phi \geq 0.97$, leading to our estimated P-value of $\mathrm{P} < 10^{-5}$. Overall, this means that the high fraction of sisters with symmetric division outcome reflects correlations in sister cell fate.

4.4.7 Estimation of crypt growth rate

To estimate an effective growth rate from the time dynamics of the total cell number and number of proliferating cells N for each individual crypt, we used the 'Uniform' model as defined in the main text. Here, each generation the number of proliferating cells increases by α_N, and the number of non-proliferating cells, M, changes by $(1-\alpha)N$, where α is the growth rate with $-1 \leq \alpha \leq 1$. For α sufficiently close to zero, the resulting dynamics of the number of proliferating and non-proliferating cells, N and M, is given by $dN/dt = \alpha/TN$ and $dM/dt = (1-\alpha)/TN$, where T is the average cell cycle duration. Solving these differential equations yields $N(t) = N(0)\exp(\alpha t/T)$ and $U(t) = M(0) - (1-\alpha)/\alpha D(0) + D(0)/\alpha \exp(\alpha t/T)$ for the total number of cells, where $U = N + M$. We then fitted $U(t)$ and $N(t)$ to the experimental data in **Fig. 2a, b**, using a single value of the fitting parameter α for each crypt and the experimentally determined value of $T = 16.2\mathrm{h}$.

4.4.8 Crypt unwrapping

At every time point the crypt axis was manually annotated in the xy plane at the z position corresponding to the center of the crypt, since tracked crypts grew perpendicularly to the objective. Three to six points were marked along the axis, through which a spline curve $s(r)$ was interpolated. Then, for each tracked cell i we determined its position along the spline by finding the value of r that minimized the distance d between the cell position $\vec{x}_i$ and the spline, i.e. $d(r_i) = \min_r |\vec{s}(r) - \vec{x}_i|$. At each time point, the bottommost cell of the crypt, i.e. that with the lowest value of r_i, was defined as position zero. Thus, the position along the axis p_i for cell i was defined as $p_i = r_i - \min_i(r_i)$. To determine the angle around the axis θ_i for cell i, we considered a reference vector u pointing in the direction of the imaging objective, given by $\vec{u} = (0, 0, -1)$, and the vector $\vec{v}_i = \vec{x}_i - \vec{s}(r_i)$ defined by the position of the cell $\vec{x}_i$ and the position of minimum distance along the spline $\vec{s}(r_i)$. Then, the angle is given by $\theta_i = \mathrm{acos}((\vec{u} \cdot \vec{v}_i)/uv)$.

4.4.9 Distance to Paneth cells

To estimate the distance between cells we used the following approach. For each cell at each time point we found the five closest cells within a 15 µm radius, which became the edges in a graph representation of the crypt (**Fig. S6**). These values were chosen because a visual inspection revealed an average nucleus size of 10 µm and an average of five neighbors per cell. This graph was then used to define the edge distance of a cell to the nearest Paneth cell. At every time point during the lifetime of that cell, the minimum number of edges required to reach the nearest Paneth cell was recorded. The edge distance is then defined as the number of edges minus one. For example, a neighbor cell of a Paneth cell (1 edge) has a distance of zero. When the edge distance of a cell to a Paneth cell varied in time, we used the mode of its distance distribution, i.e. the most frequently-occurring value, as recorded during its lifetime.

4.4.10 In vivo clonal tracing

All experiments were carried out in accordance with the guidelines of the animal welfare committee of the Netherlands Cancer Institute. $Lgr5^{\text{EGFP-ires-CreERT2}}$; $R26_{\text{LSL-tdTomato}}$ double heterozygous male and female mice (Bl6 background) were housed under standard laboratory conditions and received standard laboratory chow and water ad libitum prior to start of the experiment. 60 hours before sacrifice, mice received an intraperitoneal injection with 0.05 mg tamoxifen (Sigma, T5648; dissolved in oil) resulting in maximally 1 labeled cell per ∼10 crypts. After sacrifice, the distal small intestine was isolated, cleaned and flushed with ice cold PBSO, pinned flat and fixed for 1.5 hours in 4% PFA (7.4 pH) at 4°C. The intestine was washed in PBT (1% Tween-20) for 10 min at 4°C after which it was cut into pieces of ∼2 cm and transferred to a 12 well plate for staining. The pieces were permeabilized for 5 hours in 3% BSA and 0.8% Triton X-100 in PBSO and stained overnight at 4°C using anti-RFP (Rockland, 600-401-379) and anti-GFP (Abcam, ab6673) antibodies. After 3 times 30 min washes at 4°C in 0.1% Triton X-100 and 0.2% BSA in PBSO, the pieces were incubated with Alexa fluor Donkey anti rabbit 568 (Invitrogen, A10042) and Alexa fluor Donkey anti goat 488 (Invitrogen, A11055) secondary antibodies overnight at 4°C. After an overnight wash in PBT, the pieces were incubated with DAPI (Thermo Fisher Scientific, D1306) for 2 hour and subsequently washed in PBS for 1 hour at 4°C. Next, the intestinal pieces were cleared using 'fast light-microscopic analysis of antibody-stained whole organs' (FLASH) described in Messal et al. [108]. In short, samples were moved to an embedding cassette and dehydrated in 30%, 75%, 2x 100% MetOH for 30 min each at RT. Subsequently, samples were put into MetOH in a glass dish and immersed in methyl salicylate diluted in MetOH: 25%, 75%, 2x 100% methyl salicylate (Sigma-Aldrich) 30 min each at RT protected from light. Samples were mounted in methyl salicylate in between two glass coverslips and images were recorded using an inverted Leica TCS SP8 confocal microscope. All images were collected in 12 bit with 25X water immersion objective (HC FLUOTAR L N.A. 0.95 W VISIR 0.17 FWD 2.4 mm). Image analysis was carried out independently by two persons. Afterwards, all discrepancies between both datasets were inspected, resulting in a single dataset. Each biologically stained cell was annotated once in the 3D-dimensional image. Different cells in the

same crypt were marked as belonging to the same crypt, which is necessary to calculate the clone size for that crypt. Only crypts that were fully visible within the microscopy images were analyzed.

Uncertainty estimation in clone size distributions In organoids, the clone sizes are measured by calculating the number of offspring the cell will have 40 hours later. This calculation is performed for every hour of the time lapse, up to 40 hours before the end. In vivo, clone sizes are measured once per crypt, as we cannot view the dynamics over time. To estimate the uncertainty in our clone size distribution, both in organoids and in vivo, we use a bootstrapping approach. We denote the total number of clones observed as N. We then used random resampling with replacement, by drawing N times a random clone from the data set of observed clones, to construct a new clone size distribution. We ran this procedure 100 times, each run storing the measured fraction of clones sizes. As a result, for every clone size we obtained a distribution of fractions, which we used to calculated the standard deviation of the fraction, as a measure of sampling error.

4.4.11 Computational model

Simulations follow the approach of **Chapter 3**, except for one difference: the proliferation compartment is now a one-dimensional space. If a cell divides, the topmost cell is pushed out of the compartment, into the differentiation compartment. Neighbor cells in the proliferation compartment can rearrange at a rate of r, as described below.

For each cell division i, with time since the previous cell division in the entire simulation Δt_i, we drew the number of cell rearrangements from a Poisson distribution with mean $(r \cdot S)\Delta t_i$, where r is the rearrangement rate per cell. We then implemented each individual rearrangement by randomly selecting a cell at position $j \in (0, S-1)$ and swapping it with the cell at position $j+1$.

The model had six parameters, of which three (α_n, α_d, ϕ) were systematically varied in our simulations. The remaining parameters were constrained by the experiments. We picked the niche size S so that the total number of proliferating cells was 30, corresponding to the typical number of dividing cells observed in the experiments, through a procedure outline in the main text. We obtained the average cell cycle duration T, as well as its distribution, from the data in **Fig. S3**. Finally, we obtained the rearrangement rate r from the observed (a)symmetry in proliferative fate observed between cousin cells. For a 'well-mixed' niche compartment, $r \cdot T \gg 1$, cousin pairs showed asymmetric outcome as often as symmetric outcome (**Fig. S5**), in contrast to our experimental observations (**Fig. 3a**). In contrast, for infrequent cell rearrangement, $r \cdot T = 1$, cells expelled from this compartment close together in time are also closely related by lineage, leading to correlations in division outcome between cousins that reproduced those observed experimentally (**Fig. S5**, **Fig. 3a**).

4.5 | Supplementary Information

File S1. Replication package. All cell lineage data, simulation code and data analysis scripts used to generate the figures have been deposited in Zenodo.

Download at
https://doi.org/10.5281/zenodo.7197573

Video S1. Tracking cell position and lineage. (Left) 3D reconstruction of an organoid expressing an H2B-mCherry reporter to visualize individual nuclei. Shown here is the crypt region, with nuclei colored by their depth along the optical axis. (Right) 3D reconstruction of a crypt growing in time using the positions of tracked nuclei. Colors represent cells that belong to the same lineage.

Download at
https://elifesciences.org/articles/80682/figures#fig1video1

Video S2. Crypt unwrapping and Paneth cell distance. (Left) 3D reconstruction of an organoid expressing an H2B-mCherry reporter to visualize individual nuclei. Shown here is the crypt region, with nuclei colored by their depth along the optical axis. (Right) 3D reconstruction of a crypt growing in time using the positions of tracked nuclei. Colors represent cells that belong to the same lineage.

Download at
https://elifesciences.org/articles/80682/figures#fig5video1

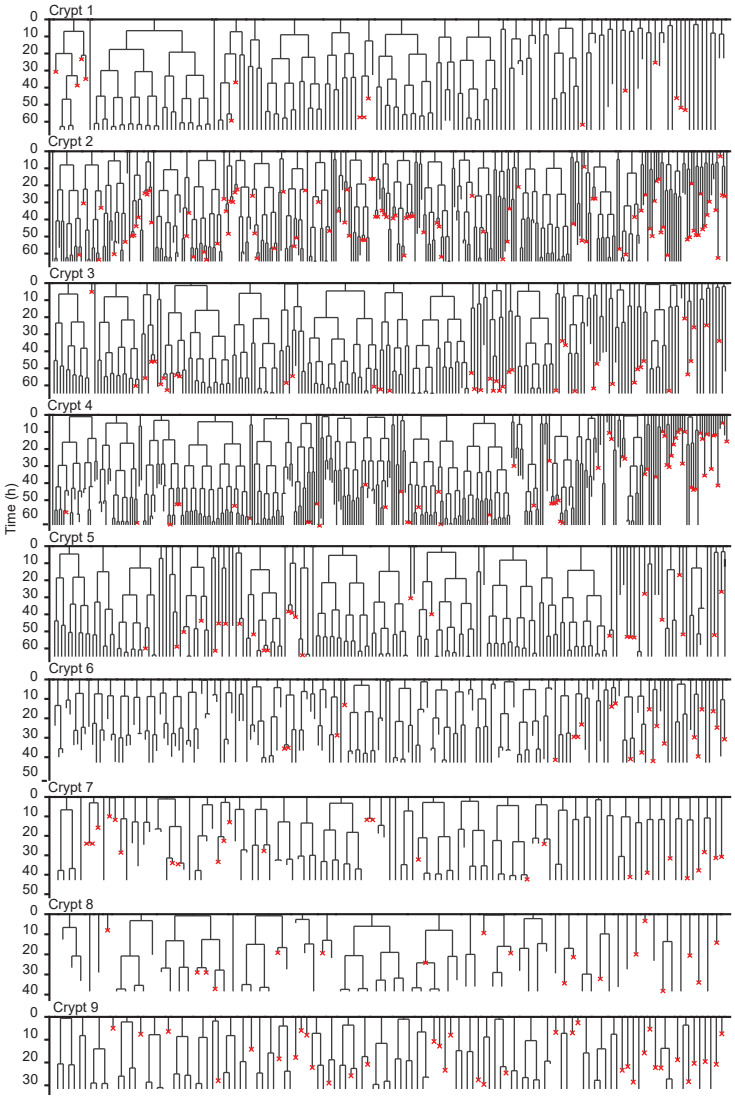


Figure S1 • Lineage trees of all tracked cells

Lineage trees of cells tracked in nine different crypts. Cells in the initial time point are ordered according to their distance to the crypt base. Red crosses indicate cell deaths and incomplete lines indicate cells that could not be accurately traced further due to insufficient fluorescence intensity or movement outside of the field of view.

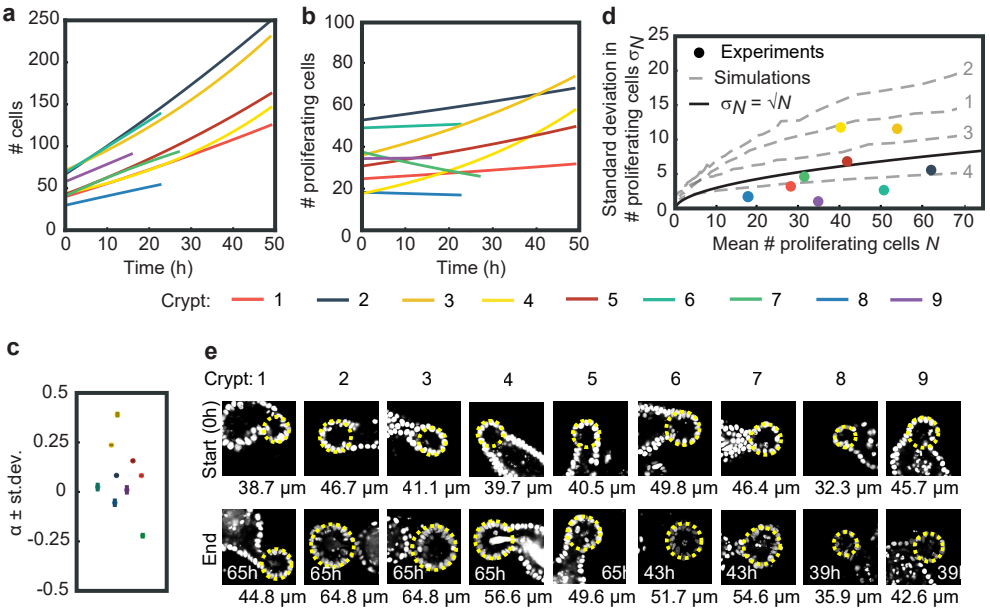


Figure S2 • Crypt growth and heterogeneity

(a), **(b)** Fits of a simple cell proliferation model (Materials and Methods) to the experimental data in Figure 2A,B. Each fit to the total number of cells (panel a) and number of proliferating cells (panel b) yields a single fitted value of the exponential growth rate α. **(c)** Fitted values of α, with the color of each marker corresponding to the crypts in panel a and b. Two crypts exhibited significant growth (crypts 3 and 4, $\alpha \geq 0.25$) while one crypt show decay in proliferating cell number (crypt 7, $\alpha = -0.25$). For most of the remaining crypts, α deviated significantly from zero, but otherwise the fitted growth was low. For this reason, the reported average growth rate in the main text excluded crypts 3,4 and 7. Including crypt 7 yields $\alpha = 0.000.1$. Error bars are the standard deviation in the fit of α. **(d)** Standard deviation of the number of proliferating cells, σ_D, versus the average, D, for each crypt, with color indicating crypts as in panel a and b. Black line shows the relationship expected for a Poisson birth-death process, $\sigma_D = D$. For crypts that exhibited growth (crypt 3 and 4), σ_D was larger than a Poisson process, while other crypts showed fluctuations close to or below what was expect for a Poisson process, implying that fluctuations were limited by some form of homeostatic control. Grey dashed lines show the same analysis for simulations of the 'Compartment' model (**Fig. 4d**), with individuals lines corresponding to the different scenarios 1-4 (indicated in grey) in **Fig. 4e, f**. For the suboptimal scenarios 1-3, fluctuations are larger than Poissonian, while the optimal scenario 4 reproduced the low fluctuations observed experimentally. **(e)** Fluorescence images of analyzed crypts 1-9 at the start and end of the experiment. Duration of the experiment is indicated in each 'end' panel. Diameter of the crypt at the start and end of the experiment is calculated by overlaying a circle (yellow line) over the z-slice at the center of the crypt. Overall, crypts display heterogeneity both in initial crypt size and growth of crypt diameter.

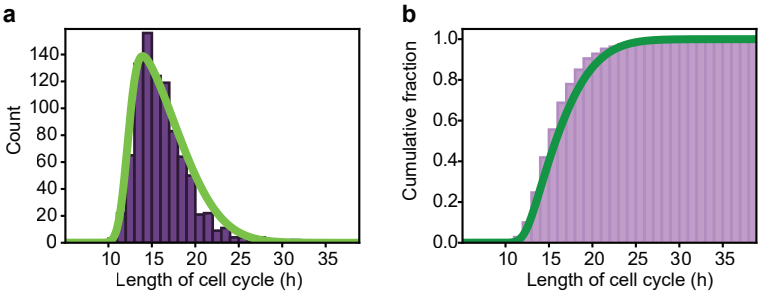


Figure S3 • Crypt growth and heterogeneity
(a) Histogram of cell cycle times of all tracked dividing cells. Green line represents a skew normal distribution fitted to the data, with parameters: shape = 6.1, location = 12.2 h, and scale = 5.2 h. **(b)** Cumulative version of the histogram of panel a.

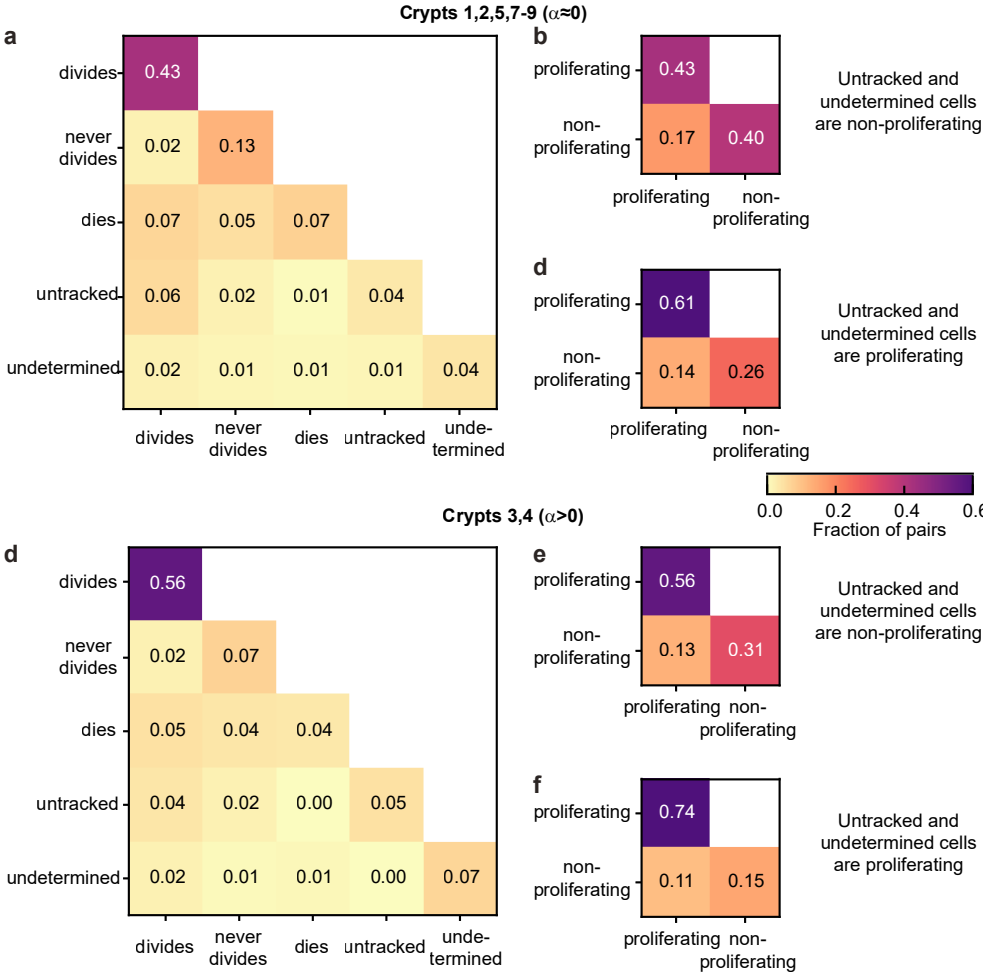

Crypts 1,2,5,7-9 (α≈0)
a
divides
never divides
dies
untracked
undetermined
0.43
0.02 0.13
0.07 0.05 0.07
0.06 0.02 0.01 0.04
0.02 0.01 0.01 0.01 0.04
divides
never divides
dies
untracked
unde-termined
b
proliferating
non-proliferating
0.43
0.17 0.40
proliferating
non-proliferating
Untracked and undetermined cells are non-proliferating
d
proliferating
non-proliferating
0.61
0.14 0.26
proliferating
non-proliferating
Untracked and undetermined cells are proliferating
0.0 0.2 0.4
Fraction of pairs
Crypts 3,4 (α>0)
d
divides
never divides
dies
untracked
undetermined
0.56
0.02 0.07
0.05 0.04 0.04
0.04 0.02 0.00 0.05
0.02 0.01 0.01 0.00 0.07
divides
never divides
dies
untracked
unde-termined
e
proliferating
non-proliferating
0.56
0.13 0.31
proliferating
non-proliferating
Untracked and undetermined cells are non-proliferating
f
proliferating
non-proliferating
0.74
0.11 0.15
proliferating
non-proliferating
Untracked and undetermined cells are proliferating

Figure S4 • Sister pair divisions patterns and proliferation control

(a) Correlations in division patterns between sister pairs (S_1,S_2), for all cells born at least 15 h prior to the end of the experiment in crypts that exhibited low growth rate $\alpha \approx 0$ in **Fig. S2b, c** (Crypts 1,2,5,7-9). Sister pair fractions are also included for cells that were lost from cell tracking (untracked) or cells that could be tracked but where the experiment ended before a division could be observed or excluded (undetermined). **(b)** Fraction of sister pairs with proliferating and/or non-proliferating cells, under the assumption that untracked and undetermined cells are non-proliferating. This assumption matches our observation that untracked cells are located almost exclusively in the villus, where cells are less likely to proliferate, and that undetermined cells older than 15 h are most likely non-proliferating according to **Fig. S3**. Cells that died were classified as non-proliferating, as in **Fig. 2a, b**. Under this assumption, the fraction of symmetrically proliferating (43%) and non-proliferating sisters (40%) is approximately equal, consistent with constant number of proliferating cells and low growth rate α in **Fig. 2a, b** and **Fig. S2b, c**. **(c)** Fraction of sister pairs with proliferating and/or non-proliferating cells, under the assumption that all untracked and undetermined cells proliferate. Under this assumption, which is not supported by our experiments, the fraction of symmetrically proliferating (61%) and non-proliferating sisters (26%) is still relatively close to balanced, implying that the balance in proliferation observed in panel b is likely not impacted strongly if a small fraction of untracked or undetermined cells would have been proliferating. **(d)** Same as panel a but for crypts that exhibited an increase in number of proliferating cells ($\alpha > 0$, Crypts 3,4). **(e)**, **(f)**) Same as panels b and c but for Crypts 3 and 4. Symmetrically proliferating sisters (56%) outnumber non-proliferating sisters (31%) even under the assumption that all untracked and undetermined sisters are non-proliferating, consistent with the growth in proliferating cells observed in these crypts.

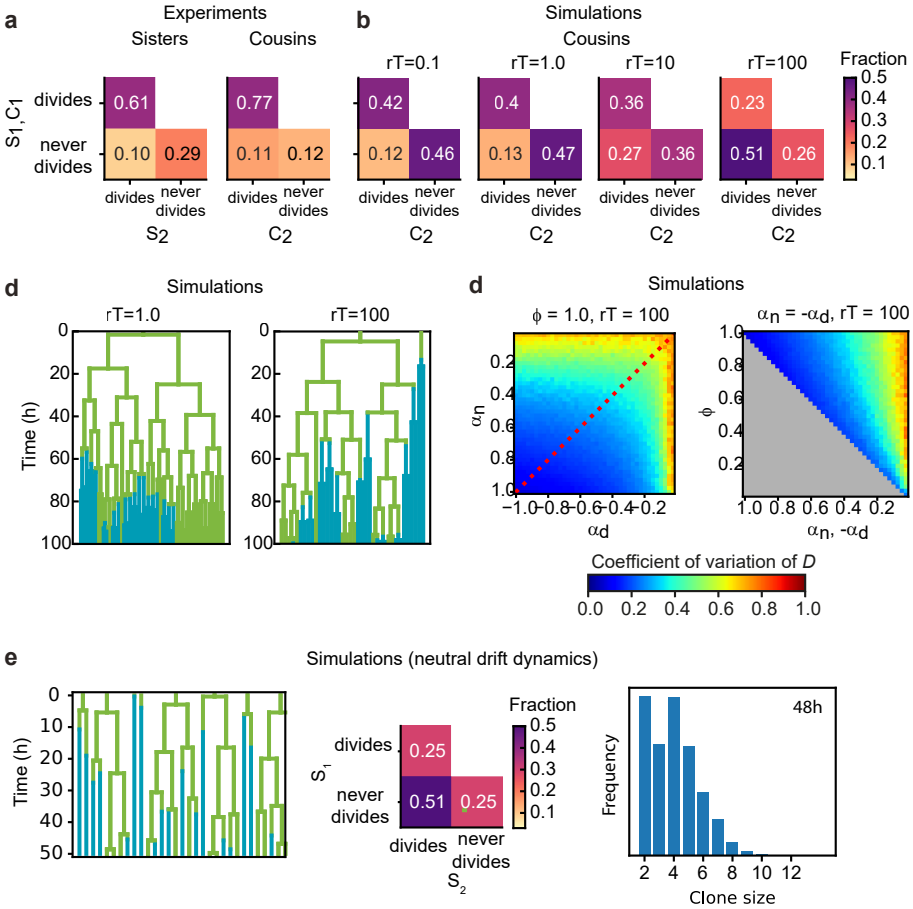

a
Experiments
Sisters
Cousins
S1,C1
divides
never divides
0.61
0.10
0.29
0.77
0.11
0.12
S2
C2
b
Simulations
Cousins
rT=0.1
rT=1.0
rT=10
rT=100
Fraction
0.42
0.12
0.46
0.4
0.13
0.47
0.36
0.27
0.36
0.23
0.51
0.26
C2
d
Simulations
rT=1.0
rT=100
Time (h)
d
Simulations
ϕ = 1.0, rT = 100
αn = -αd, rT = 100
αn
αd
ϕ
αn, -αd
Coefficient of variation of D
e
Simulations (neutral drift dynamics)
Time (h)
S1
S2
0.25
0.51
0.25
Fraction
48h
Frequency
Clone size

Figure S5 • Dependence of lineage dynamics on cell rearrangements

(a) Experimentally observed correlations in division patterns between sister pairs (S_1,S_2, n=499) and cousin cells pairs (C_1,C_2, n=560), for cell pairs where neither cell died. **(b)** Correlations in division pattern between cousin cells (C_1,C_2) for simulations of the two-compartment model, for increasing cell rearrangement rate r. Here, the product rT, where T is the average cell cycle duration, indicates the approximate number of rearrangements experienced per cell cycle. For a 'well-stirred' proliferative compartment, $rT \gg 1$, asymmetric lineage outcome is more prevalent that symmetric outcome for cousin cells. However, for infrequent cell rearrangements, $rT \approx 1$, asymmetric lineage outcome for cousin cells is less prevalent that symmetric outcome, with a frequency that resembles that seen experimentally in panel a. Data is averaged over 50 independent simulations, with $\alpha_n = -\alpha_d = 0.95$, $\phi = 0.95$., and the average number of proliferating cells $D = 30$. **(c)** Example lineage trees obtained from simulations for different rT. Line color indicates whether the cell was in the proliferative (green) or non-proliferative compartment (blue). For $rT = 1$, closely related cells remain largely together in space, causing ejections from the proliferative compartment to be correlated within lineages. As a consequence, not only sisters, but also cousins and more distantly related cells keep proliferating or cease proliferation together. For $rT = 100$, this correlation within lineage trees is lost. **(d)** Coefficient of variation of the total number of proliferating cells D for simulations with a 'well-stirred' proliferative compartment, $rT = 100$. Left panel shows the effects of varying the growth rates of both compartments when all divisions are symmetric and right panel of varying the degree of symmetry when both compartments have opposite growth rates. Overall, the coefficient of variation shows a similar dependence on parameters as seen for the simulations with infrequent cell rearrangements, $rT = 1$, shown in **Fig. 4** in the main text. **(e)** Simulations of a neutral drift dynamics model for $N = 30$. This model is similar to the 'well-stirred' two-compartment model for the case where all cells are proliferative in the niche compartment ($\alpha_n = 1$, $\phi = 1$), and no cells are proliferative in the differentiation compartment ($\alpha_n = -1$, $\phi = 1$). However, the main difference is that in the neutral drift model, cells switch from proliferative to non-proliferative immediately upon exiting the niche compartment. This results in half of all divisions being asymmetric, as every cell has an independent chance of 50% to exit the niche and thus halt proliferation.

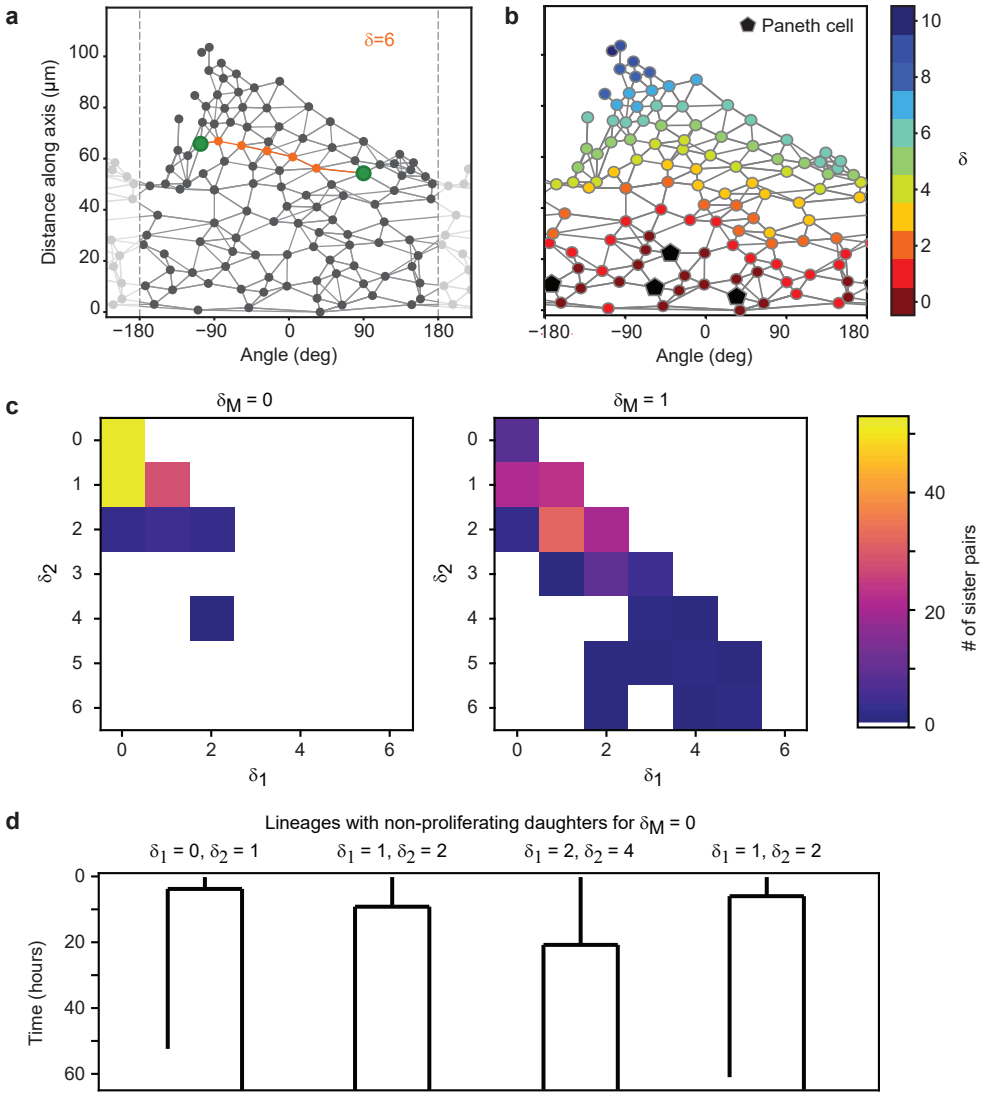


Figure S6 • Proliferation dynamics as function of Paneth cell distance
(a) After crypt unwrapping and determination of neighbors, the link distance δ between two cells (green markers) is defined as the number of edges that constitute the shortest path connecting the two cells (indicated in orange, resulting in this case in a neighbor distance of $\delta = 6$. **(b)** Link distance δ of each cell to its closest Paneth cell, shown for a single time point. **(c)** Distribution of Paneth cell distances δ_1 and δ_2 of sister pairs generate from a mother cell either directly touching a Paneth cell ($\delta_M = 0$) or one cell removed from a Paneth cell ($\delta_M = 1$). **(d)** Examples of cell lineages for rare sister pairs that ceased proliferation, even though their mother was touching a Paneth cell ($\delta_M = 0$). For each sister pair, the distances δ_1 and δ_2 after division are given.

5 | Loss of Paneth cell contact starts a Wnt differentiation timer in the intestinal crypt

In preparation for publication.
Kok, R.N.U., Zheng, X., Clement, T.V.M., Spoelstra, W.K., Tans, S.J., & Van Zon, J.S. (2024).

Abstract

Wnt signaling is essential for development and tissue maintenance. In the small intestine, Paneth cells at the crypt bottom secrete Wnt, which maintains the proliferation of neighboring stem cells. When stem cells lose contact with Paneth cells, they lose Wnt signaling and start differentiation. Two competing hypotheses exist for how Wnt signaling is lost after stem cells lose contact with a Paneth cell: either cells experience less Wnt signaling the further they move away from Paneth cells, or Wnt signaling is lost passively upon each cell division by dilution of the amount of bound Wnt per cell. Here, we show that loss of contact with a Paneth cell starts a timer that controls stem cell differentiation. Cells quickly reduce their Wnt signaling once contact with the Wnt-producing Paneth cells is lost, even in absence of cell growth and division. If Paneth cell contact is not regained in time, cells differentiate even if the distance towards the nearest Paneth cell is short. The timer model thus provides an explanation for the onset of stem cell differentiation in which cell migration is unnecessary.

5.1 | Introduction

Wnt signaling plays an essential role in embryonic development, adult tissue maintenance and cancer [109–111]. In the adult small intestine, canonical Wnt signaling is required for stem cell maintenance, and overstimulating Wnt causes hyperproliferation of the epithelium [28, 93, 112, 113]. The size of the stem cell zone in the intestine is known to be regulated by the Wnt gradient along the crypt-villus axis. Wnt is maximally expressed in the stem cell region at the crypt base and gradually decreases along the crypt domain [104, 113, 114]. Therefore, a spatial model has often been assumed.

In this model, once a stem cell has moved a particular distance away from the source of Wnt signaling, it will start to differentiate. Such a model is appealing, as it provides an intuitive explanation for cellular differentiation. However, the time dynamics of the maintenance of the Wnt gradient and the regulation of cell differentiation have not been tested.

There are several indications of the insufficiency of such a simple spatial model. First, Farin et al. [104] showed that Wnt signaling only affects direct neighbor cells, thereby preventing any long-range signaling. The authors looked at intestinal organoids, where Paneth cells are the only source of Wnt [94, 115–117]. They proposed an alternative model, in which after loss of contact with Paneth cells, Wnt dilutes by every cell division. Thus, preventing cell divisions keeps Wnt signaling active. However, the authors did not track Wnt signaling over time, and left open the question of how differentiation is controlled in time.

Second, the spatial gradient of Wnt does not explain the origins of Paneth cells. Paneth cells are the only differentiated cell type commonly found next to stem cells at the bottom of the crypt. As a spatial model requires Paneth cell differentiation to happen outside the stem cell region and higher in the crypt, it was assumed that all Paneth cells were born higher in the crypt and then move back into the stem cell region at the bottom [57]. However, that such a detour is not required for differentiation was recently shown by Böttcher et al. [118]. In their work, the authors have shown that Paneth cells and enteroendocrine cells originate directly from stem cells, and that this process happens low in the crypt. It is unclear how this observation fits into the spatial model of differentiation.

Here, we directly look at Wnt dynamics over time. We use intestinal organoids of a recently developed mouse model with an *Axin2* reporter [13], which is an important and widely expressed Wnt target gene [119]. In our work, we combine live cell tracking of these organoids with end-point antibody staining for the most common intestinal cell types. We use the resulting time lapse movies to observe when cells lose and gain contact with other cells, when the *Axin2* expression of the cells is changed, and which cell type the cells assume. In addition, we test for causality by ablating individual Paneth cells using a laser. We show that once contact with Paneth cells is lost, cells rapidly decrease their Wnt target gene production, even before a division occurs.

Our findings indicate that Wnt signaling follows a timer model, where the amount of Wnt on the cell surface decreases rapidly after contact with Paneth cells is lost. Once Wnt signaling becomes low enough, cells would differentiate. This model provides a robust mechanism for differentiation: small lapses in Paneth contact will not cause accidental differentiation, but longer losses of contact will make stem cells differentiate, even if the stem cell does not move any further away from the Paneth cells.

5.2 | Results

5.2.1 Characterizing Wnt signaling in time and space

To quantify Wnt signaling dynamics, we measured expression of the Wnt target *Axin2*, a well-established reporter of active Wnt signaling [13, 119], in small intestinal organoids using confocal time-lapse microscopy (**Fig. 1a**). We employed a recently developed *Axin2*$^{P2A\text{-}rtTA3\text{-}T2A\text{-}3xNLS\text{-}SGFP2}$ reporter, where nuclear SGFP2 is expressed under control of the endogenous *Axin2* promoter through a multi-cistronic targeting cassette [13]. In this reporter, SGFP2 expression level increased with induced activation of Wnt signaling and exhibited a spatial gradient along the crypt-villus both in the small intestine and in intestinal organoids, indicating that nuclear SGFP2 level was a reliable proxy for Wnt signaling activity [13].

We followed *Axin2* expression in time in individual cells of *Axin2*$^{P2A\text{-}rtTA3\text{-}T2A\text{-}3xNLS\text{-}SGFP2}$; *Rosa26*^{mTmG} intestinal organoids, which carry an additional tdTomato membrane marker (**Fig. 1a**). We used the nuclear SGFP2 signal to manually track > 140 cells in three organoids, and quantified *Axin2* expression by integrating the total nuclear SGFP2 fluorescence signal of a cell (**Fig. S1**, Materials and Methods). Paneth cells, which form the only Wnt source in organoids [94, 115–117], were identified as crypt-based cells with low *Axin2*/SGFP2 expression [13]. We found that the average SGFP2 signal decreased with each cell's distance to the nearest Paneth cell (**Fig. 1b**). SGFP2 expression was high only for each Paneth cell's first- and second-degree neighbors and quickly fell to background level beyond. This agrees with earlier observations of a short-ranged Wnt gradient, reaching only 1-2 cell diameters beyond each Paneth cell [104].

5.2.2 *Axin2*/SGFP2 decreases after loss of contact with Paneth cells

Using the membrane marker to delineate cell boundaries, we noticed that SGFP2-expressing cells that touched Paneth cells often lost contact subsequently (**Fig. 1c**). We therefore examined how SGFP2 levels depended on Paneth cell contact. First, we quantified SGFP2 levels in selected cells that were first in sustained (> 1 hour) contact with Paneth cells, and then remained out of contact for > 2 hours (**Fig. 1d**). We found that all cells showed an overall increase of SGFP2 while still in Paneth cell contact. However, most cells (6/8) showed a decrease in SGFP2 level that initiated with 1-2 hour after loss of Paneth cell contact. Next, we examined whether these marked changes in *Axin2*/SGFP2 accumulation were more broadly observed in our dataset. For this, we divided all measured SGFP2 trajectories into intervals of constant Paneth cell contact, i.e. either touching or fully detached, and excluding cell division, and then measured both the SGFP2 level and its slope averaged over the interval (**Fig. 1e**). We found that cells in Paneth cell contact exhibited higher average SGFP2 level than cells not in contact ($P = 2.952 \cdot 10^{-8}$; two-sided T-test assuming equal variance), and that their SGFP2 increases over time, as evidenced by a positive slope. In contrast, cells that are detached from Paneth cells but still have a high (> 1.5) average SGFP2 level generally

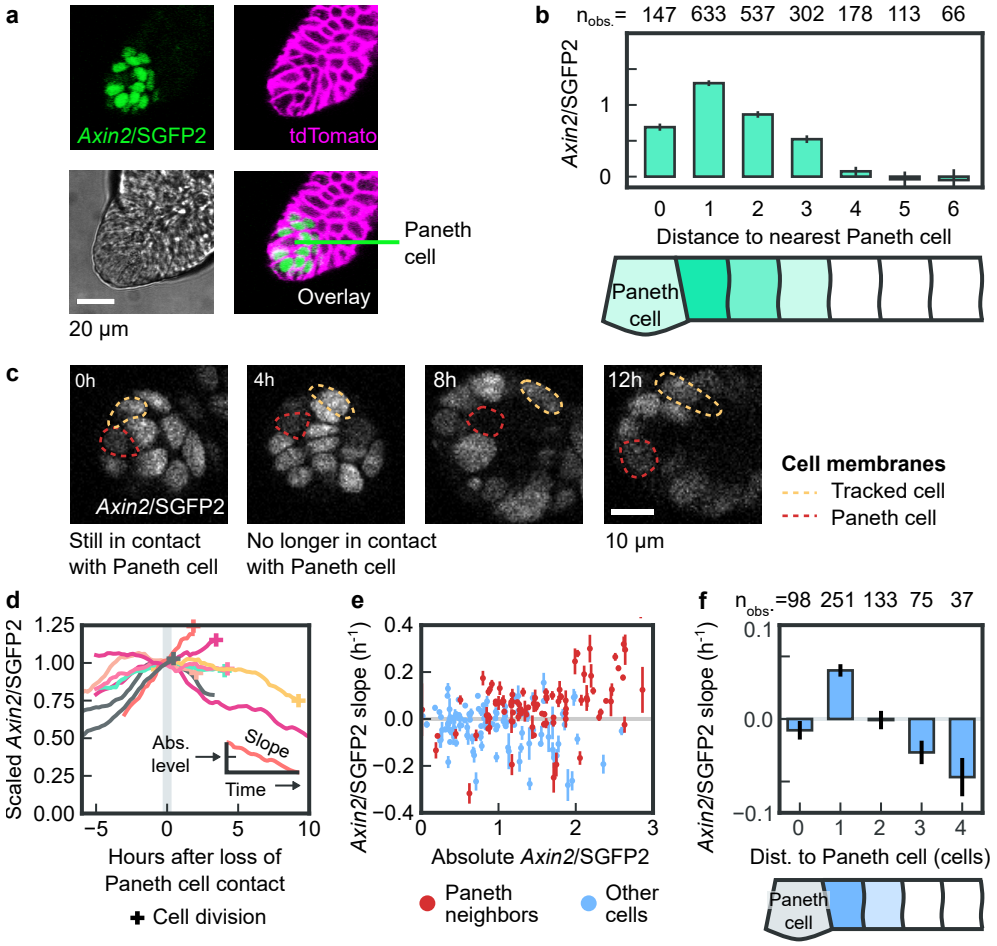


Figure 1 • Influence of Paneth contact on *Axin2*/SGFP2 expression

(a) *Axin2*/SGFP2 expression (green) in the crypt domain of an intestinal organoid that also carried a tdTomato membrane marker (magenta). Scale bar: 20 µm. **(b)** *Axin2*/SGFP2 level as function of distance to the nearest Paneth cell. Error bars are standard error of the mean (S.E.M.) **(c)** Image sequence of a single cell (dashed yellow line, indicating membrane outline) losing contact with a Paneth cell (red dashed line). **(d)** *Axin2*/SGFP2 levels of cells that were initially in Paneth cell contact, and then lost contact for at least two hours. *Axin2*/SGFP2 levels are normalized by value at time of contact loss. Crosses indicate trajectories that end with cell division. The slope of trajectories (change in *Axin2*/SGFP2 over time) was positive while in Paneth cell contact and generally negative when contact was lost. **(e)** *Axin2*/SGFP2 level averaged of individual trajectories plotted versus each trajectory's slope. Color indicates cells that were in contact (red) or detached from Paneth cells (blue). Cell trajectories were split at cell divisions, and at moments where cells lose or gain contact with Paneth cells. Error bars are the standard error around the linear fit. **(f)** Slope of *Axin2*/SGFP2 trajectories as function of distance to the closest Paneth cell. Error bars are standard error.

exhibit a negative slope, corresponding to decreasing SGFP2 level. Consistently, when we quantified slope as function of distance to the closest Paneth cell (**Fig. 1f**), we found SGFP2 only increased for only for cells in direct Paneth cell contact, while the slope was close to or below zero for more distant cells. Overall, these results indicate that direct Paneth cell contact is required for maintaining *Axin2*/SGFP2 expression. Moreover, loss of contact results in a decline in *Axin2*/SGFP2 expression that is rapid (1-2 hour) on the timescale of cell division ($\sim$20 hour) and therefore appeared to occur independently of Wnt dilution due to cell division.

5.2.3 Paneth cell ablation induces loss of *Axin2*/SGFP2 expression

To directly test that physical Paneth cell contact is required for *Axin2*/SGFP2 expression in its neighbors, we used 800 nm laser pulses to ablate individual Paneth cells [120] and examined the impact on SGFP2 level of the surrounding cells. For this, we selected crypts with 1-2 Paneth cells, so that almost all cells were only in contact with a single Paneth cell, if at all. We found that a 50ms 800nm laser pulse resulted in rapid destruction of the targeted Paneth cell (**Fig. 2a**), with subsequent resealing of the epithelium observed within 30 minutes. Strikingly, Paneth cell ablation resulted in a strong decrease in SGFP2 level in its direct neighbors, within 12 hours, with a 56% decline of SGFP2 level in Paneth cell neighbors (**Fig. 2b**). In contrast, in control organoids where we performed a mock ablation, SGFP2 levels in neighbors remain approximately constant. Interestingly, we found that Paneth cell ablation had no effect on its second-degree neighbors, as SGFP2 levels showed a similar decrease both upon ablation and in control organoids (**Fig. 2c**). This is consistent with our observation above that direct Paneth cell contact is required for maintaining *Axin2*/SGFP2 expression and suggests that Paneth cell-secreted Wnt proteins do not spread substantially beyond its first-degree neighbors.

The observed decrease in *Axin2*/SGFP2 expression upon Paneth cell ablation could be due to passive dilution of cell-bound Wnt by cell growth and division, in line with the dilution model, or instead could be explained by a decrease in Wnt signaling that is independent of cell growth. To differentiate between these two scenarios, we performed Paneth cell ablation in organoids exposed to the cell division inhibitor Palbociclib [121]. As expected, Palbociclib fully inhibited cell division, as evidenced by a lack of increase in cell number compared to control organoids (**Fig. S2a**). Moreover, cells also failed to increase in volume, as measured using the *tdTomato* membrane marker (**Fig. S2b**), indicating that Palbociclib also inhibited cell growth. We found that, upon exposure to Palbociclib, Paneth cell ablation resulted in a 35% decrease in SGFP2 level (**Fig. 2d**), whereas mock-ablated organoids exposed to Palbociclib showed no such decrease. Overall, this indicates that the loss of *Axin2*/SGFP2 expression following Paneth cell ablation is independent of cell growth and division.

Changes in Wnt signaling likely manifest themselves in concomitant changes in *Axin2* production rate. Given that, in our experiments, *Axin2* and SGFP2 are transcribed as a single polycistronic mRNA molecule, the production rate of SGFP2 is similar to that of *Axin2* by design. Therefore, assuming simple, linear transcription and translation

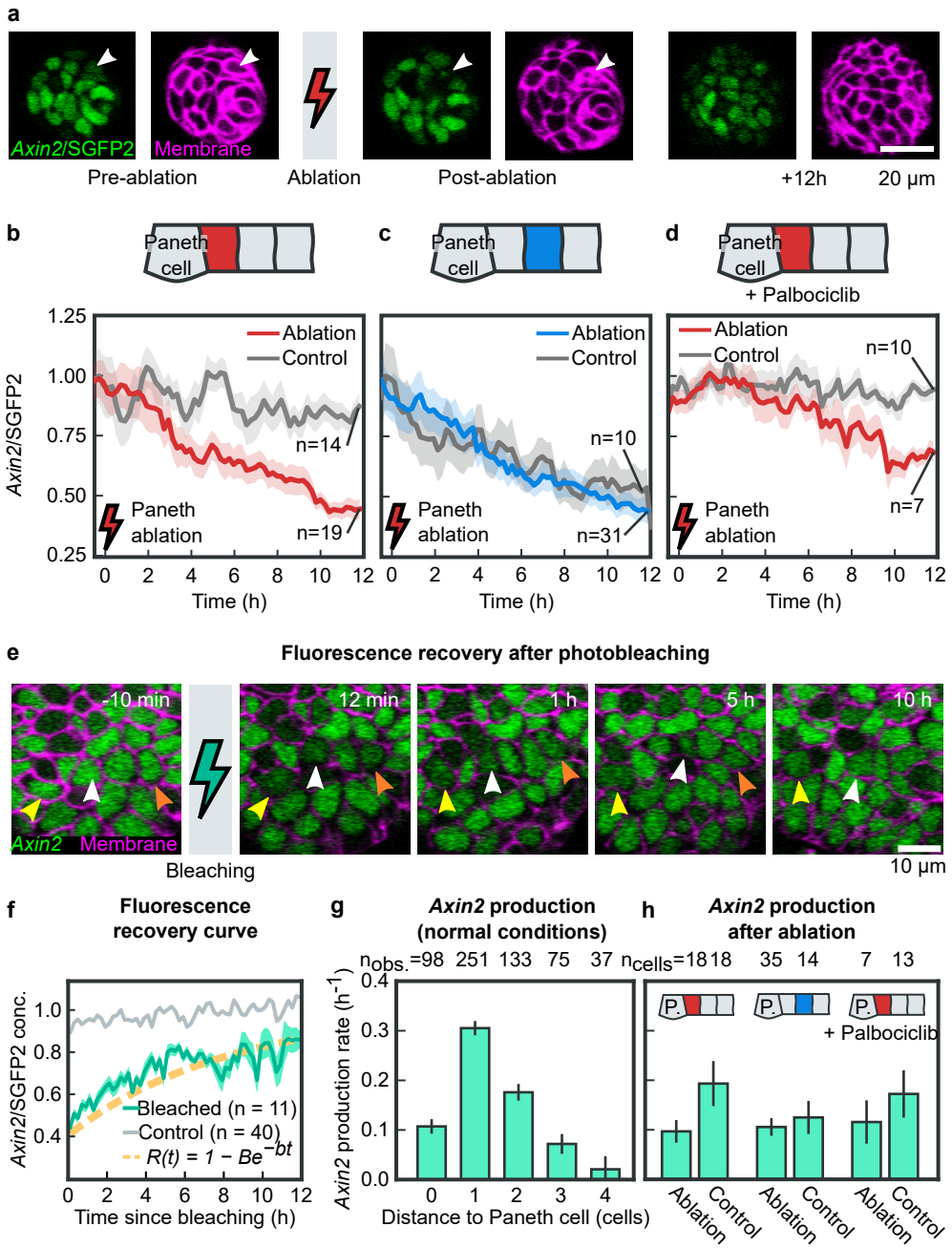

a
Axin2/SGFP2
Membrane
Pre-ablation
Ablation
Post-ablation
+12h
20 µm
b
Paneth cell
Ablation
Control
n=14
n=19
Paneth ablation
Axin2/SGFP2
Time (h)
c
Paneth cell
Ablation
Control
n=10
n=31
Paneth ablation
Time (h)
d
Paneth cell
+ Palbociclib
Ablation
Control
n=10
n=7
Paneth ablation
Time (h)
e
Fluorescence recovery after photobleaching
10 min
12 min
1 h
5 h
10 h
Axin2
Membrane
Bleaching
10 µm
f
Fluorescence recovery curve
Axin2/SGFP2 conc.
Bleached (n = 11)
Control (n = 40)
R(t) = 1 − Be−bt
Time since bleaching (h)
g
Axin2 production (normal conditions)
nobs.=98 251 133 75 37
Axin2 production rate (h−1)
Distance to Paneth cell (cells)
h
Axin2 production after ablation
ncells=18 18 35 14 7 13
P.
+ Palbociclib
Ablation
Control

Figure 2 • Impact of Paneth cell ablation on *Axin2*/SGFP2 expression

(a) Targeted photoablation of a Paneth cell (white arrow) by a 50ms 800 nm laser pulse. The ablated Paneth cell is removed from the epithelium in 0.5 hour after ablation, while *Axin2*/SGFP2 expression (green) has decreased significantly in the neighboring cells after 12 hours. Magenta: membrane-bound tdTomato. **(b)** SGFP2 level in first-degree Paneth cell neighbors following Paneth cell ablation (red) or in mock-ablated organoids (grey). **(c)** As (b), but for second-degree neighbors. Results for (b) and (c) were averaged over three independent (mock) ablation experiments. **(d)** As (b), but for organoids exposed to Palbociclib (red) to inhibit cell growth and division or in control organoids without Palbociclib (grey). Results were averaged over two independent (mock) ablation experiments. **(e)** Fluorescence recovery of SGFP2 (green) in individual bleached cells (indicated by colored arrows). **(f)** SGFP2 fluorescence recovery curve (green), compared to neighboring non-bleached cells (grey). Lines indicate mean, while shaded area is S.E.M. Red dashed line is a fit to the recovery curve with inferred SGFP2 degradation rate d. **(g)** Inferred *Axin2* production as a function of the distance to the nearest Paneth cell. Paneth cell distance and *Axin2* production rate were calculated for subsequent four-hour time windows. Error bars are S.E.M. **(h)** *Axin2* production rate inferred for first-degree Paneth cell neighbors (left), second-degree neighbors (middle) and first-degree neighbors in Palbociclib-treated organoids, both in Paneth cell-ablated and mock-ablated organoids. Paneth cell ablation reduced *Axin2* production rate to minimal levels, independent of cell growth and division.

dynamics, the Wnt-dependent *Axin2* production rate $r(t)$ can be inferred from the observed SGFP2 fluorescence signal $S(t)$ as $r(t) = \frac{d}{dt}S(t) + d \cdot S(t)$, provided that the SGFP2 degradation rate d is known (see Materials and Methods). This equation shows that directly measuring production rate would be more informative than measuring the slope of the SGFP2 fluorescence signal, as we did above, as for a given production rate the SGFP2 slope can either be negative or positive, depending on the current SGFP2 level. We then measured the degradation rate d using Fluorescence Recovery After Photobleaching (FRAP). We bleached SGFP2 in individual cells at different positions relative to Paneth cells and measured recovery of SGFP2 fluorescence over time (**Fig. 2e, f**), with the observed recovery over $\sim$8 hours consistent with a degradation rate of $d = 0.14/\mathrm{hour}$ (Materials and Methods).

Using the measured SGFP2 degradation rate, we then estimated the *Axin2* production rate r based on the SGFP2 intensity and its slope, averaged over 4 hours. As expected, r was highest in direct Paneth cell neighbors, and decreased for second-degree neighbors (**Fig. 2g**). For more distant cells, the inferred *Axin2* production rate is minimal. When we next inferred the production rate for the ablation experiments, we found that upon Paneth cell ablation r was reduced (**Fig. 2h**), to a level comparable with the production rate in the second-degree neighbors. Palbociclib-treated organoids showed similar r as control organoids, indicating that *Axin2* production rate, and hence Wnt signaling, did not depend strongly on cell growth and division.

5.2.4 Loss of Paneth cell contact drives decreasing *Axin2*/SGFP2 expression and differentiation

The loss of Wnt signaling in stem cells that are displaced away from Paneth cells is assumed to trigger cell differentiation [122] but how changes in Wnt signaling level drive loss of stemness and subsequent differentiation remains poorly understood. We therefore adapted an approach we developed recently [123] to link *Axin2*/SGFP2 expression dynamics with simultaneous measurement of cell type and differentiation state. Briefly, we performed time-lapse imaging of $Axin2^{P2A\text{-}rtTA3\text{-}T2A\text{-}3xNLS\text{-}SGFP2}$; $Rosa26^{mTmG}$ organoids and subsequently performed multiple rounds of antibody staining to identify the main intestinal cell types (Stem, Paneth, enteroendocrine and goblet cells, and enterocytes) (**Fig. 3a**, Materials and Methods). Using both the nuclear SGFP2 signal and the tdTomato membrane marker, we manually tracked cells in time, allowing to connect the endpoint cell type determined by antibody staining to the SGFP2 dynamics measured by time-lapse imaging.

We first characterized the *Axin2*/SGFP2 level for each cell type (**Fig. 3b**), averaging the SGFP2 signal over the time interval between the birth of each cell and the end of imaging. As expected, we found the highest SGFP2 level for stem cells and intermediate levels for Paneth cells expression [13], while fully differentiated goblet cells or enterocytes showed no SGFP2. For these cell types, *Axin2*/SGFP2 level correlated with distance to Paneth cells (**Fig. 3c**), with stem cells forming first- or second degree neighbors of Paneth cells, while goblet cells and especially enterocytes located further away. Surprisingly, rare enteroendocrine cells showed higher SGFP2 than goblet cells or enterocytes. We also found a population of cells that were not stained by markers for stem cell or differentiated fate. These cells showed intermediate SGFP2 levels and were often found as first- or second-degree Paneth cell neighbors, similar to stem cells. Based on our observations here and our previous work [123], we identify these cells as transit-amplifying (TA) cells, that are between stem cells and differentiated cells in the differentiation hierarchy. Overall, these measured *Axin2*/SGFP2 expression levels for each cell type agreed well with single-cell RNA sequencing data [124] (**Fig. S3**), both for *Axin2* itself and for other intestinal Wnt target genes important for differentiation (LGR5, SOX9, MYC and ASCL2) [122].

Next, we selected cells in our data set that were initially in Paneth cell contact but were identified as different cell types at the end of the experiment. In this subset of cells, we only found stem cells (high SGFP2) and TA cells (intermediate SGFP2), indicating that full differentiated to goblet cells or enterocytes required > 17 hours. Interestingly, all cells exhibited high SGFP2 expression at the start of the experiment, consistent with their position as first-degree Paneth cell neighbors, but cells that assumed TA fate showed a marked decrease in SGFP2 expression, in contrast to cells with stem cell fate (**Fig. 3d**). We then quantified for each cell the fraction of time it was in direct Paneth cell contact, averaged over a 4 hour time window. Strikingly, we observed that prospective TA cells exhibited reduced Paneth cell contact well before SGFP2 levels decreased relative to prospective stem cells (**Fig. 3d**). This is consistent with our observation above that *Axin2*/SGFP2 expression decreased in response to loss of direct Paneth cell contact, and indicates that this loss of contact directly drives differentiation.

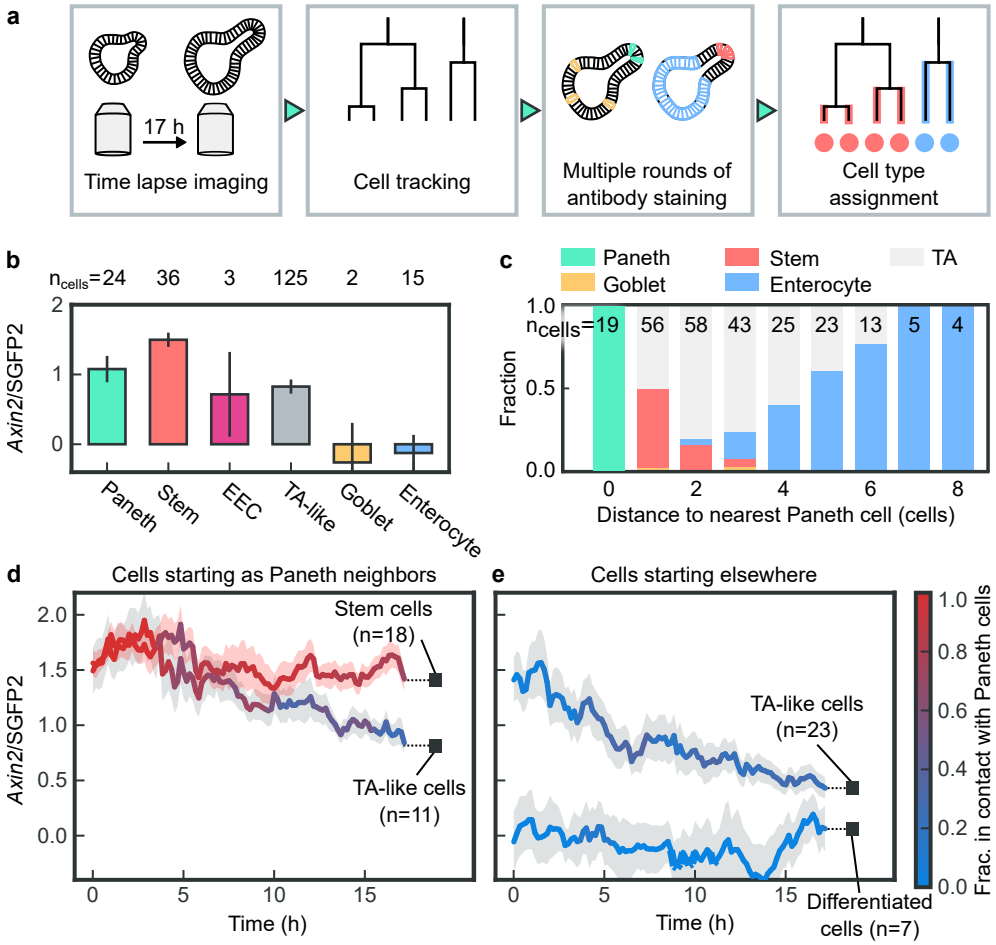


Figure 3 • *Axin2*/SGFP expression during cell differentiation

(a) Workflow for linking *Axin2*/SGFP2 dynamics to cell differentiation. **(b)** Average *Axin2*/SGFP2 expression level for different intestinal cell types. Transit amplifying (TA) cells exhibit intermediate *Axin2*/SGFP2 expression, between stem cells (high SGFP2) and fully differentiated goblet cells and enterocytes (no SGFP2). Error bars are S.E.M. **(c)** Position of different cell types relative to the closest Paneth cells, as observed at the time-lapse imaging end point. Cells with the lowest SGFP2 expression (goblet cells and enterocytes) are furthest away from Paneth cells. **(d)** *Axin2*/SGFP2 level in time for cells that were in Paneth cell contact at the start of the experiment. SGFP2 trajectories are averaged over subsets of cells that have the same cell type as observed at the end of the experiment, either stem cell, TA cell or differentiated cell (goblet cell or enterocyte). Trajectories are colored according to the average fraction of time cells were in direct Paneth cell contact. Shaded area is S.E.M. Loss of Paneth cell contact precedes decrease in SGFP2 level for first-degree Paneth cell neighbors that eventually assumed TA fate. **(e)** Same as (d) but for cells that were initially detached from Paneth cells. The panel shows how TA cells started with an *Axin2*/SGFP2 level close to that of stem cells, but ended at a level close to differentiated cell types.

We also examined *Axin2*/SGFP2 expression in selected cells that were initially located further away from Paneth cells (second degree neighbor or more distant, **Fig. 3e**). At the end of the experiments, these cells were either differentiated cells (goblet cells or enterocytes) or TA cells. We found that all differentiated cells showed constant low SGFP2 level throughout the experiment. TA cells initially exhibited intermediate SGFP2 levels, similar to the newly differentiated TA cells in **Fig. 3d**, but SGFP2 expression subsequently decreased strongly. Interestingly, even though at the end of the experiments SGFP2 levels in these TA cells was almost as low as for differentiated cells, they stained negative for any markers of goblet of enterocyte fate, indicating that expression of these markers requires prolonged absence of Wnt signaling. Overall, these measurements point to a specific differentiation sequence: first, stem cells lose direct Paneth cell contact, which drives a decrease in Wnt signaling, loss of stem cell markers and assumption of TA cell fate, with expression of differentiation markers only apparent after *Axin2*/SGFP2 expression was lost for $>$ 10 hours.

5.2.5 Recovery of *Axin2*/SGFP expression by regaining of Paneth cell contact

Our results thus far showed that direct Paneth cell contact is required for maintenance of stem cell fate. An unexpected observation in this light was that only 48% of the cells adjacent to Paneth cells were stem cells, with 50% exhibiting TA fate (**Fig. 3c**) (and 2% were goblet cells). We confirmed that TA cells adjacent to Paneth cells exhibited a lower inferred *Axin2* production, compared to stem cells at the same location (**Fig. 4a**), indicating that they received a comparatively lower Wnt signal. The crypts we examined typically contained multiple Paneth cells, allowing for the possibility that only cells in contact with multiple Paneth cells receive sufficient Wnt to maintain stem cell fate, while cells in contact with a single Paneth cell receive too little and assume TA fate. However, cells that contact one Paneth cell displayed similar *Axin2*/SGFP2 levels and inferred *Axin2* production rate as cells adjacent to two Paneth cells (**Fig. 4b**), indicating that a single Paneth cell is able to provide Wnt at saturating levels in its direct neighbors.

Above we showed that loss of Paneth cell contact was prerequisite for assuming TA fate (**Fig. 3d**). At the same time, we found that, upon loss of Paneth cell contact, cells frequently regained Paneth cell contact due to cell rearrangement (**Fig. 4c**), with periods of 1-5 hours observed between subsequent Paneth cell contacts. This raised the alternative hypothesis that TA cells adjacent to Paneth cells were cells that lost Paneth cell contact previously, causing a transition from stem to TA fate, only to subsequently regain contact. To further study this hypothesis, we selected (fragments of) cell trajectories that exhibited a single change in Paneth cell contact and examined its impact on *Axin2*/SGFP2 expression. Cells exhibited comparatively low SGFP2 level and production rate 2-10 hours prior to regaining Paneth cell contact (**Fig. 4d**), comparable to values seen for TA cells (**Fig. 3b**, **Fig. 4a**). Shortly (2-4 hours) after regaining Paneth cell contact, we found that whereas SGFP2 levels remained low, SGFP2 production rate had increased significantly, while cells that had regained Paneth cell contact for $>$10 hours showed SGFP2 levels and production rate comparable to that of stem cells (**Fig. 3b**, **Fig. 4a**). Finally, for cells that lost Paneth cell contact for $>$2 hours, SGFP2

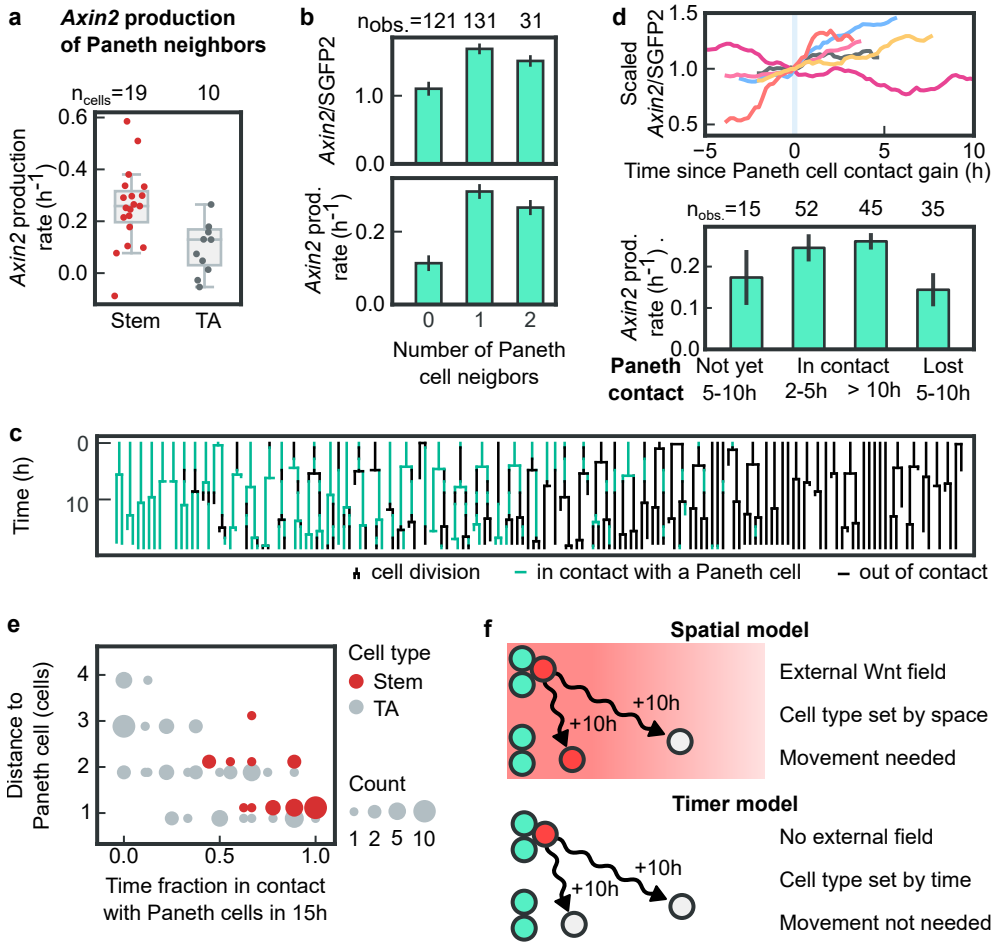


Figure 4 • *Axin2*/SGFP2 expression as function of Paneth cell contact

(a) Inferred *Axin2*/SGFP2 production rate for cells in direct Paneth cell contact, averaged over 4 hours prior to the end of the experiment. For cells younger than 4 hours, we included data from their mother cell. TA cells exhibit lower production rate than stem cells ($P = 0.0048$; two-sided t-test assuming equal variance). **(b)** *Axin2*/SGFP2 level and production rate as function of the number of adjacent Paneth cells. Contact was measured over 4 hour time window. Cells with one or two Paneth cell neighbors exhibit similar SGFP2 level and production rate ($P = 0.29$; two-sided t-test assuming equal variance). **(c)** Lineage trees of tracked cells, with line color indicating Paneth cell contact. Lineages are ordered by increasing averaged Paneth cell contact. Cells often lose and subsequently regain Paneth cell contact. **(d)** *Axin2*/SGFP2 production rate for cells that experience a change in Paneth cell contact. The top panel shows that most cells (5/6) gaining contact with Paneth cells stop their decrease in *Axin2*/SGFP2 fluorescence, or start an increase. The bottom panel shows the change in production. **(e)** Dot plot of distances to Paneth cells and the fraction of the last 15 hours that a cell has been in contact with a Paneth cell. **(f)** Cartoon contrasting the spatial gradient model of Wnt signaling with our timer model.

level and especially SGFP2 production rate was significantly lower again. The increase in SGFP2 level and production rate upon regaining Paneth cell contact was also visible in individual cell traces (**Fig. 4d**). When we then examined the history of Paneth cell contact for stem and TA cells that were found adjacent to Paneth cell at the end of the experiment, we indeed found that the fraction of time cells were in direct Paneth cell contact was significantly higher for stem cells than for TA cells ($P = 3.3 \cdot 10^{-10}$, **Fig. 4e**). This is consistent with the hypothesis that TA cells adjacent to Paneth cells have only recently regained Paneth cell contact. More generally, these results indicate that loss of Wnt signaling, as evidenced by decreasing *Axin2*/SGFP level and production rate, is not irreversible, but can be recovered by regaining Paneth cell contact through cell rearrangements.

5.2.6 A timer model of Wnt signaling

Wnt signaling is generally assumed to generate spatial cell fate patterns by acting as spatial gradients, either through diffusion of Wnt or transport of Wnt by cell movement [104]. In such gradient models, cell type is determined by the cell's spatial position within the gradient (**Fig. 4f**), with differentiation occurring as cells move away from the Wnt source. Our results, however, point to a fundamentally different model for Wnt-induced patterning in the intestine. In this model, Wnt signaling is induced only in cells that are in direct Paneth cell contact and maintenance of Wnt signaling requires continuous contact with Paneth cells (**Fig. 1**, **Fig. 2**). Once Paneth cell contact is lost, Wnt signaling directly falls (**Fig. 1**, **Fig. 2**), acting as a timer that drives the transition first towards TA fate and subsequently fully differentiated cell types (**Fig. 3**). However, regaining Paneth cell contact causes an immediate recovery of Wnt signaling (**Fig. 4**), effectively resetting the timer. As a consequence, and in contrast to the gradient model, cell fate in this timer model is not determined by each cell's spatial position, but rather by the time since last Paneth cell contact (**Fig. 4**), with the well-known spatial subdivision of the intestinal epithelium into a regions containing stem cells, TA cells or differentiated cells emerging through the action of cell rearrangement and displacement away from Paneth cells.

5.3 Discussion

In this paper, we have observed a rapid decline of the *Axin2*/SGFP2 signal for cells not in contact with Paneth cells. This result shows that our data may not be compatible with preexisting models of Wnt signaling, namely the spatial Wnt diffusion model and the division dilution model. For the spatial diffusion model, we have shown that Paneth cells are indeed not able to sustain Wnt pathway activity in their second-degree neighbors [104], which makes our data incompatible with this model. For the division dilution model, it would be expected that the level of *Axin2*/SGFP2 should stay constant until the next cell division. However, we saw a rapid decline in the level of *Axin2*/SGFP2, which instead points towards degradation of Wnt signaling.

Instead, our results point towards a model that we named the timer model. In this

model, Wnt signaling has a high turnover rate, requiring a continued supply of new Wnt proteins from Paneth cells. Once contact with Paneth cells is lost, a cell will no longer receive Wnt proteins. Therefore, the loss of contact with a Paneth cell starts a timer for loss of stemness.

Recent work showed that epithelia are complex environments where cells can lose and regain contact with other cells multiple times [125], and that specifically for the small intestinal epithelium cells can move both forwards and backwards along the crypt-villus axis [126]. In this environment, our timer model would work well. Small lapses of Paneth cell contact would not disturb stem cell fate, but longer lapses of Paneth cell contact, thereby resulting in low Wnt pathway activation, would inevitably induce stem cell differentiation [122].

While we have shown that *Axin2* expression decreases upon loss of contact with Paneth cells, we have not investigated the molecular basis. In recent years, it has become clear that Wnt signaling in embryos is not static, but shows dynamic patterns including contextual adaptations, transient activation and even complex oscillations [127]. In cells, Wnt signaling is carefully controlled by both antagonists and potentiators, as well as by positive and negative feedback loops [128]. The rapid loss of Wnt signaling seen in this paper could be implemented using a negative feedback loop. The Wnt target genes Rnf43 and Znrf are one example of a negative feedback loop, as they cause endocytosis and destruction of the Wnt receptor. *Axin2* itself is another example, as it is part of the β-catenin destruction complex [129].

The timer model is different from the spatial diffusion model, as in that model all cells would first need to move away from the crypt to differentiate. Paneth cells would then move back to the crypt after differentiating [130]. However, this required movement contradicts recent experimental evidence that cells can already differentiate in the crypt [118, 123]. Our proposed timer model does not require such long-range movement for differentiation, as differentiating cells can stay close to the source of Wnt signaling. Likewise, the timer model is also different from the division dilution model, in which the Wnt signaling level of a cell can only decrease due to passive dilution by cell divisions, instead of changing continuously. The division dilution model raises the question of what would make cells differentiate, given that substantial changes of Wnt signaling would take several cell cycles. Farin et al. speculated that cells could use a Wnt signal that is coming from only one side of the cell as a cue to differentiate [104]. In the timer model, such cell asymmetry models would not be needed, where large changes in Wnt signaling can happen substantially faster.

5.4 | Materials and Methods

5.4.1 Organoid line

Axin2$^{P2A\text{-}rtTA3\text{-}T2A\text{-}3xNLS\text{-}SGFP2}$; *Rosa26*^{mTmG} murine intestinal organoids were a gift from the Van Amerongen lab (Universiteit van Amsterdam, The Netherlands) and are described in [13].

5.4.2 Organoid culturing

Organoids were cultured as described in **Section 4.4.2** on page 79.

5.4.3 Organoid time lapse microscopy

To perform a time lapse experiment, we first seeded mechanically dissociated organoids in 40 µL gel in four-well chambered cover glass (Cellvis). This was performed one day before the start of the time lapse experiment. To prevent the gel from solidifying immediately, seeding was performed on a cold block. The imaging well was put in a fridge until the organoids had sedimented down, as observed using light microscopy. Typically, this sedimentation process took about five to ten minutes. Afterwards, the chambered cover glass was put in the incubator. After leaving the plates in an incubator for 20 to 30 minutes, the growth medium was added.

A Leica confocal microscope was used for time lapse imaging. In this setup, a controlled closed environment was used to maintain the temperature at 37 °C, CO_2 at 9% and high humidity. The CO_2 concentration was kept at 9% instead of 5% due to leakiness of the system. A 40x water immersion objective (Leica) was used for image acquisition. Organoids were imaged every 12 minutes for up to 20 h. Z-stacks were made at 2 µm intervals, with a field of view of 233 by 233 µm. Z-ranges were specified for each organoid, with a maximum range of 70 µm. The following excitation wavelengths and laser intensities were used: 488 nm (3% or 1% intensity), 552 nm (1% intensity), 638 nm (1% intensity). Detection was done at 493-547 nm and 582-633 nm.

An imaging well typically contains over 50 organoids, of which we only imaged a small number. We selected for small organoids, with only one to three crypts, that we close to the bottom of the imaging plate.

The nuclei were made visible using SirDNA. This allowed us to track *Axin2*-negative cells more easily, as otherwise we would need to rely only on the membrane-located tdTomato signal. When organoids in the imaging plate had generated buds (early crypt structures), the growth medium was removed and replaced by new growth medium supplemented with 1 µM SirDNA (Spirochrome) and 10 µM Verapamil (Spirochrome). This was then left to incubate for $\sim$3 hours before starting imaging.

5.4.4 Antibody staining

After time lapse imaging, organoids were immediately fixed using 4% formaldehyde in PBS. After fixation, the organoids were permeabilized using 0.2% Triton-X. Then the organoids were sequentially stained with several rounds of primary and secondary antibody and dyes (**Table 5.1**). Before antibody staining, organoids were blocked in 5% skim milk for 1 h. Primary antibodies were diluted in 5% skim milk and were incubated on the organoids for 48 h. The organoids were then washed with PBS, and incubated with secondary antibodies diluted in TBS, for 1 h. If applicable, the organoids were then incubated with Wheat Germ Agglutamin (WGA) in PBS for 2 h. The organoids were washed of excess secondary antibody and/or dye before imaging. Imaging was

done using the same Leica confocal microscope setup as for time lapse imaging. Using these combinations of antibodies and dyes mentioned before, different cell types can be identified within the intestinal organoid (**Table 5.2**).

Table 5.1 • Used antibodies and dyes.

Antibody/Dye	Concentration	Manufacturer reference	Research Resource Identifier
rabbit anti-OLFM4	1:500 dilution	Bioke D6Y5A	RRID:AB_2650511
rabbit anti-human lysozyme	1:800 dilution	Dako #A0099	RRID:AB_2341230
mouse anti-human Krt20	190 ng/mL	Dako M701929-2	
mouse anti-human ChrA	2.5 µg/mL	Santa Cruz #sc-393941	RRID:AB_2801371
donkey anti-rabbit	2 µg/mL	Abcam #ab175649	RRID:AB_2715515
donkey anti-mouse	4 µg/mL	Thermofisher #A31571	RRID:AB_162542
Wheat Germ Agglutamin	2.5 µg/mL	Biotium #29028-1	

Table 5.2 • Mapping from antibody staining to cell type.

	Stem cell	Paneth cell	Goblet cell	Enterocyte	Enteroendocrine cell
Olfm4	+				
Lysozyme		+			
WGA (dye)		+	+		
Krt20			+	+	
ChrA					+

5.4.5 Ablation of Paneth cells

Imaging for the was performed on a different microscope (Nikon A1R MP) with a 40x oil immersion objective (NA = 1.30). 30 z-slices with 2 µm step size were taken per organoid every 12 minutes. Experiments were performed at 37°C and 5-8% CO_2, achieved by using a stage-top incubator (Okolab).

Ablation was carried out using an 800 nm laser (tunable Ti:Sapphire laser, Mai Tai High Performance DeepSee, Spectra Physics; 100% power) for 40 to 60 ms, on a single point set at the center of the nucleus. Paneth cells were identified by their bright membrane-localized tdTomato signal and their low *Axin2*/SGFP2 signal amidst neighbor cells that have a high *Axin2*/SGFP2 signal (**Fig. 1a**). No antibody staining was performed after the time lapse.

5.4.6 Fluorescence recovery after photobleaching

The experiment was performed on the same microscope as the ablation experiment. No antibody staining or SirDNA was used. SGFP2 in three cells per crypt cells was partially bleached for 5 seconds at 20% laser power (Nikon LU-N4 laser unit, 488

nm laser). The remaining cells served as controls. Bleached and control cells were manually tracked for up to 12 hours and fluorescence intensity levels of GFP signal were measured until cell division, death or moving out of frame. For each individual recovery curve, we normalized the fluorescence intensity by the average intensity of the four images taken before bleaching, and we fitted the function

$$f(t) = 1 - Be^{-Bt}$$

to the individual recovery curves using `scipy.optimization.curve_fit` in Python with the fraction of bleached molecules B and the degradation rate of SGFP2 b as free fitting parameters. Bleached cells were excluded from analysis if (i) if the cell could only be tracked for less than five hours, (ii) their average SGFP2-fluorescence intensity in first hour was more than 65% (insufficient bleaching), (iii) if their maximum measured intensity was higher than 35% above their intensity prior to bleaching (indicating that the cell was not in homeostasis at the moment of bleaching) or (iv) if the measured recovery time $1/b$ was more than 21 hours, indicating that *Axin2* was either not produced or immeasurably low on the timescale of the cell cycle.

5.4.7 Automatic extraction of fluorescence values

For extracting fluorescence curves outside the FRAP experiment, a different, less labor-intensive method was used compared to drawing ellipses. In addition, instead of measuring the fluorescent concentration, this method measures the total fluorescence of a nucleus.

The centers of the nuclei in the crypt were manually tracked from the start to the end of the time lapses in OrganoidTracker (**Chapter 2**). (The automated tracking module of OrganoidTracker could not be used, as the cells do not contain a nuclear reporter.) Every pixel in the (3D) image was then assigned to the nearest nuclear center position, with a maximum distance of 7 µm (**Fig. S1**). The sum of the intensities of all pixels belonging to a cell is regarded as the total fluorescence of the cell.

When a trace of a single cell is shown in this paper, the measured values are averaged over the two hours before and the two hours after. When multiple cells are shown together, the *Axin2*/SGFP2 values of the cells are averaged, but no time averaging is performed.

To be able to compare signals from multiple organoids, we needed to set a consistent scale for each organoid. We wanted to have a scale where 1 represents the *Axin2*/SGFP2-level of a typical cell where Wnt signaling is on, and 0 for a cell that displays only background signal. Since we mostly track cells in the crypt, we defined the median SGFP2 signal of a cell as 1, with the idea that this value represents a typical cell in the crypt. For defining 0, we noticed that starting from the 4th-degree-neighbor of Paneth cells, in general no SGFP2 fluorescence was observed, and that the average measured signal was minimal. Therefore, we defined this average as 0. Since the background is not uniform across the organoid and across time, this caused some cells to have an intensity below zero. We did not cap these negative values to zero, as this would artificially increase the average signals.

5.4.8 Measuring SGFP2 production and degradation

We set up a simple production-degradation model to describe how the SGFP2 signal evolves over time:

$$\emptyset \underset{b}{\overset{f}{\rightleftharpoons}} G$$

Here, G is the measured concentration of SGFP2, which can be produced at rate f and degraded at rate b. This reaction equation results in the following differential equation:

$$\frac{dG}{dt} = f - b \cdot G(t) \tag{5.1}$$

We assume that the GFP degradation rate b is equal in all cases, but we allow the production rate f to vary over time for each cell. This setup allows us to measure b from a Fluorescence Recovery After Photobleaching (FRAP) experiment, which then allows us to measure f for all other cells. Because GFP and *Axin2* are produced together, this model allows us to measure the production rate of *Axin2* for all cells.

For the FRAP experiment, we bleached several cells in the bottom of the crypt (**Fig. 2e**). Cells here show a constant amount of GFP, as verified by the gray line in **Fig. 2f**. Therefore, for these cells $\frac{dG}{dt} = 0$, and the recovery of the bleached cells can be fit to $\frac{G(t)}{G_0} = 1 - Be^{-bt}$ to obtain the degradation rate b. Here, G_0 is the measured concentration of GFP before bleaching and B the fraction of bleached GFP.

5.4.9 Determining neighbors of cells

Neighbors of cells were manually annotated every 10 time points, which corresponds to every two hours. To do this, connections were established between all cells for which the membranes touched, as observed in the tdTomato channel.

5.4.10 Measuring *Axin2* production rate as a function of distance to the nearest Paneth cell

In **Fig. 2g**, as well as in Figure 4b and d, we measured the *Axin2* production rate as follows. We iterated over all time lapses, and for every time point for which we had annotated each cell's neighbors (which was done every two hours), we iterated over every cell present in that time point. For every cell, we either measured the distance to the nearest Paneth cell (**Fig. 2g**), or the number of Paneth cells the cell was currently in contact with (Figure 4b and d). We then looked at a 4-hour time window of the *Axin2*/SGFP2 intensities and measured the mean intensity and the slope. If we could not obtain a 4-hour time window for that cell, because the cell divided or we could not

track it for that long, we discarded the cell. Using **Eq. 5.1**, we then measured the *Axin2* production of the cell. Note that this method measures the same cell multiple times.

In **Fig. 2h**, we are not interested in sampling the organoid at different time points, but instead in tracking the original neighbors or second-degree neighbors of a Paneth cell. Therefore, we tracked the cells that were a neighbor at the start of the experiment. If a cell divided, the daughters were regarded as two new cells for the analysis. For all cells that we could track for at least 4 hours, we calculated the mean *Axin2*/SGFP2 intensity and intensity slope, from which we calculated a single *Axin2* production rate, still using **Eq. 5.1**. Note that we now obtain only a single *Axin2* production rate for each cell.

5.4.11 RNA sequencing data analysis

Single-cell RNA sequencing data of organoids cultured in control conditions [124] were obtained from the GEO database [131, 132] using accession number GSE92332. Normalization, feature selection and scaling of the UMI count table was done using the R package Seurat v4 [133].

5.5 | Supplementary Information

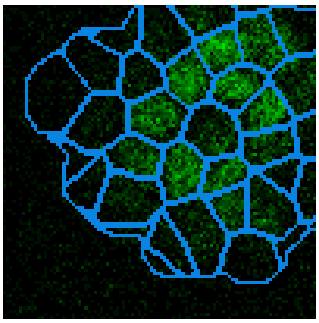


Figure S1 • Automatic measurement method of *Axin2*/SGFP2
Measurement of *Axin2*/SGFP2 signal. Every pixel is assigned to the closest nucleus center, with a maximum distance of 7 µm.

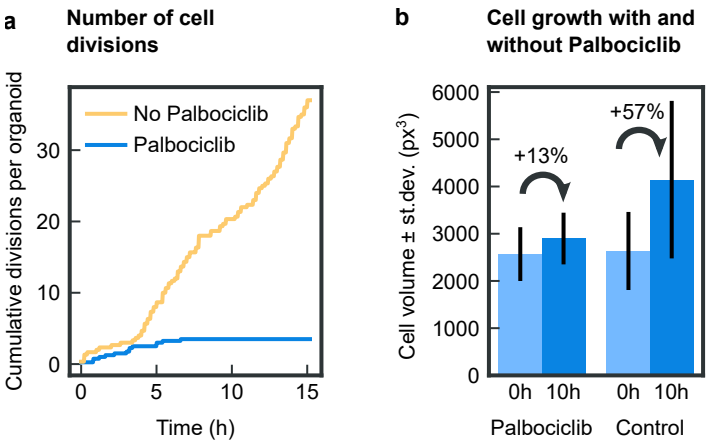


Figure S2 • Palbociclib treatment
(a) Cumulative number of cell divisions for control organoids and for organoids where Palbociclib was added to the medium several hours before imaging. **(b)** Effect of Palbociclib on cell growth. If a cell divides (which happens in the control only), the volume of the two daughter cells is summed.

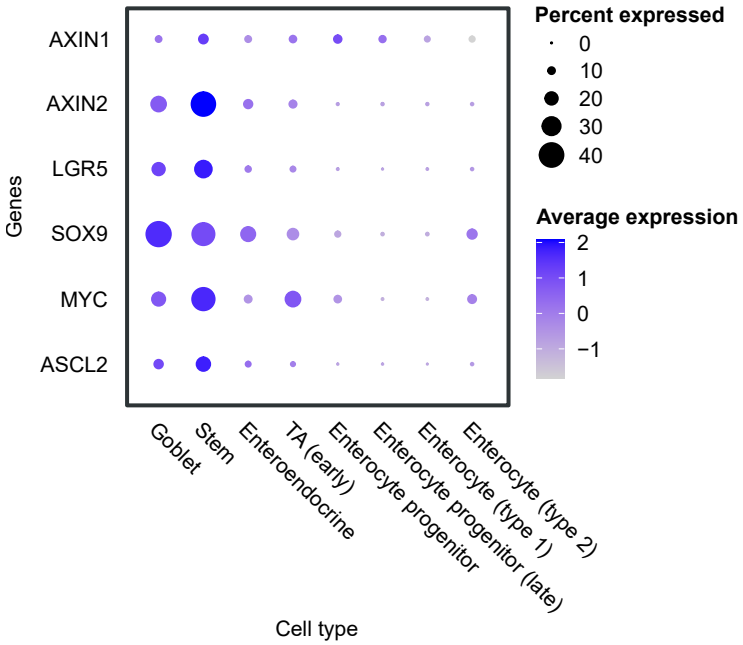


Figure S3 • Single-cell RNA sequencing
Gene expression of various genes in various cell intestinal types. AXIN1 is part of the β-catenin destruction complex and is used as a baseline for gene expression. The other genes are all Wnt target genes. The sequencing dataset contained no Paneth cells. All data in this figure is from the mouse intestinal organoid dataset of Haber et al. [124] (data accessible at NCBI GEO database [131], accession GSE92332).

6 | Label-free cell position and volume tracking in intestinal organoids

In preparation for publication.
Kok, R.N.U., Spoelstra, W.K., Betjes, M.A., Van Zon, J.S., & Tans, S.J. (2024).

Abstract

For cell tracking and segmentation in organoids, a fluorescent nuclear and membrane marker is required. However, incorporating markers is challenging, labor-intensive and limits imaging. Here, we show that for intestinal organoids, surprisingly the transmitted light images of this dense and complex three-dimensional tissue contain enough information to reconstruct the nuclei and membranes. Our method only requires the training of a 3D U-net network and should therefore be easily applicable.

6.1 | Main text

Organoids promise direct study of dynamic processes within tissues at the cellular level, which is key to issues including cell type specification, pathway activation, aberrant proliferation, and drug responses [134–136]. Realizing this potential requires imaging approaches that can follow cells in time and extract quantitative functional information. Using fluorescent labelling, it has recently become possible to track the closely packed and highly dynamic nuclei within 3D organoid architectures [135, 137] (**Chapter 2**). In principle, one can extend this approach, by adding fluorescent labels for key markers and the cellular envelope. The latter is relevant to delineating the cell volume beyond the nucleus, and hence to quantify cellular processes in the cytoplasm and at the membrane. However, this reliance on fluorescent labelling has important drawbacks: spectral overlap limits the number of colors which in turn impairs process quantification, phototoxicity degrades temporal resolution and cell health, in particular for multi-color imaging, while the genetic construction of multi-color lines is not readily accessible in all cases, even as technologies are rapidly improving.

Here, we present the LabelFreeTracker approach based on machine learning to overcome these general limitations. Machine learning has driven a rapid expansion of

microscopy image analysis. It has enabled automated tracking of cells in 3D organoid architectures, and cell feature identification from bright-field microscopy, by training on fluorescent images of these features. The latter has thus far only been achieved for isolated cells on 2D supports [1, 2, 138]. However, we surmised that combining these two ideas could allow one to use confocal bright-field microscopy to track the dynamics of cell envelopes and nuclei in 3D organoid tissues. This aim poses major challenges, however. First, due to scattering of the transmitted light by the organoid tissue structure, which typically measures several tens of microns, this imaging mode is considered ill-suited for cell identification within 3D tissue structures. Indeed, while bright field imaging is often used to visualize global organoid shapes [139], an amorphous texture is typically observed internally. Second, following cells in time introduces other imaging and analysis challenges. Specifically, lineage trees can be dramatically mixed up by minor tracking errors; when tracked cells are not identified even in just one frame, or cell identities are exchanged during division, which involves rapid movement out of the epithelial layer. Third, cells can be of many differentiated types in organoid cultures and undergo major change in shape and size as they develop in time, which poses additional identification challenges.

To address these issues, our method follows these steps. First, we defined a 3D variant of the standard U-net network [46] and trained it on matching bright-field and fluorescent images (**Fig. 1a**). We used 3D still images from time-lapse movies of growing organoids with crypt and villus domains, which allows covering all cell types and their morphological changes during growth, division, and any underlying differentiation process. This involved 3894 image pairs from 22 organoids for the nuclei and 1908 image pairs for the membranes. Critical to tracking cells though time are the nuclear positions, which we found are not well identified by these networks (see below). Instead, we first generated cell nuclear position training data (ground truth, **Fig. 1a**) with a previously developed neural network and the measured fluorescent nuclear images. Next, we used this data to train a third 3D U-net network that draws small dots at the nuclear positions based on bright field images (**Fig. 1a**, bottom). After this training phase, the three networks yielded three channels to identify the cell nuclei, envelopes, and nuclear positions in bright field movies (**Fig. 1b**, middle). Next, cells were linked in time using a min-cost flow solver and manual correction (**Fig. 1b**, right). This approach yields cellular lineage trees, movements, neighbor relations, as well as cell volume and shape in time. Specific fluorescent markers can hence be correlated with these features and allow quantification of expression levels or other cellular process for each cell in time.

Most notably, LabelFreeTracker correctly identified nuclei and membranes that were not visible by eye in the bright field images (**Fig. 2a-b**, triangles). In many regions, bright field either showed no nuclear or membrane-like features or did show strong false-positive features – for instance suggesting epithelial layer boundaries where they were not. Identification was accurate despite these confounding elements. Feature identification worked deep into the organoid tissue, allowing full 3D reconstruction of the key crypt and villi domains at the cellular level (**Fig. 2c**). Fidelity deteriorated for tissue depths beyond 30 to 40 µm, reflecting general imaging limitations (**Fig. S1**). Notably, the intensity of the predicted fluorescence levels correlated well with the measured level, both for 8-by-8-pixel areas (**Fig. 2d**) or for individual pixels (**Fig. 2e**). A key difference between the 3D U-net and visual inspection is that the former is inher-

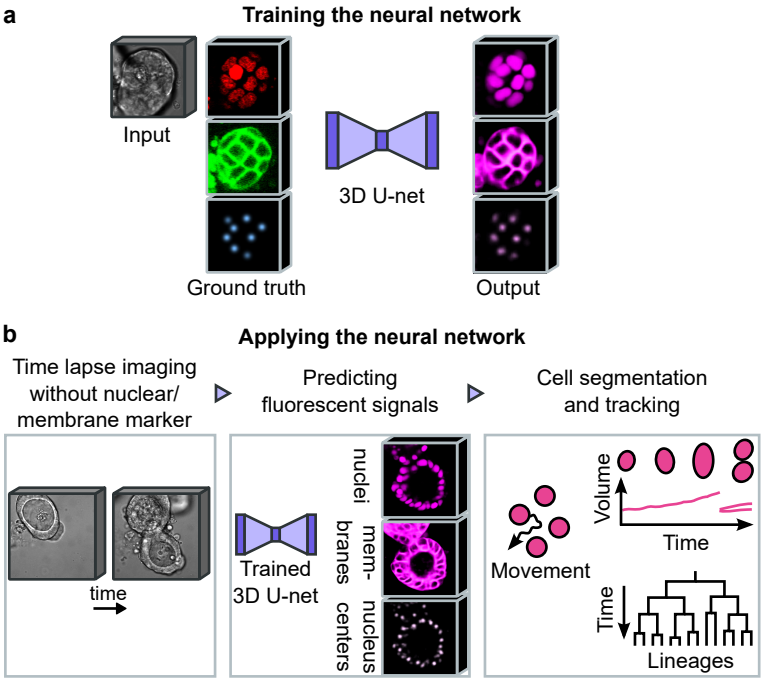


Figure 1 • Nucleus and membrane prediction overview
(a) Training of the neural network. This step requires organoids with a fluorescent nuclear and membrane marker, although the two markers can be in different organoid lines. **(b)** Intended application of the method: segmentation and tracking of organoids without a nuclear and/or membrane marker.

ently 3D, and can use information distributed across other Z-layers to identify features, whereas the human inspection is inherently 2D, and hence limited to information presented in a single Z-layer. Indeed, we found that information from other Z-layers was important to increasing the correlation coefficient for predicted and measured fluorescence levels (**Fig. S2**). Most notably, LabelFreeTracker correctly identified nuclei and membranes that were not visible by eye in the bright field images (**Fig. 2a-b**, triangles). In many regions, bright field either showed no nuclear or membrane-like features or did show strong false-positive features – for instance suggesting epithelial layer boundaries where they were not. Identification was accurate despite these confounding elements. Feature identification worked deep into the organoid tissue, allowing full 3D reconstruction of the key crypt and villi domains at the cellular level (**Fig. 2c**). Fidelity deteriorated for tissue depths beyond 30 to 40 µm, reflecting general imaging limitations (**Fig. S1**). Notably, the intensity of the predicted fluorescence levels correlated well with the measured level, both for 8-by-8-pixel areas (**Fig. 2d**) or for individual pixels (**Fig. 2e**). A key difference between the 3D U-net and visual inspection is that the former is inherently 3D, and can use information distributed across other Z-layers to identify features, whereas the human inspection is inherently 2D, and hence limited to information presented in a single Z-layer. Indeed, we found that information from other Z-layers was important to increasing the correlation coefficient for predicted and measured fluorescence levels (**Fig. S2**).

The third channel predicted nuclear positions. The percentage of nuclei identified in each frame was high and remained high up to an imaging depth of 30 µm (recall is 90.0%), after which the performance dropped (**Fig. 2f**). A large majority of predicted nuclei within 30 µm of the epithelial layer were indeed nuclei (**Fig. 2f**, precision is 91.6%). As shown in **Fig. S3**, the performance can vary depending on the quality of the images and the layout of the organoid. As a further control, we first counted nuclei using bright field with LabelFreeTracker, subsequently labelled these same organoids with nuclear labelling dye (**Fig. S4**), and found that the numbers correlated tightly (**Fig. 2g**, $R^2 = 0.92$).

LabelFreeTracker is useful for a wide range of applications, as illustrated with examples (**Fig. 3**). First, cell counting in growing organoids reveals growth heterogeneity between organoids (**Fig. 3a**, grey traces), and identifies drug-induced cell division arrest (**Fig. 3a**, blue traces). Second, cell tracking over notably long periods (over 60 hours) can directly quantify cell cycle times as about 15 hours for proliferative cells, show how some lineages remain proliferative (**Fig. 3b**, blue and red traces), while others transition to non-proliferative states (**Fig. 3b**, green trace). Moreover, the ability to correlate these features with others is key to the power of this approach, yielding a third example (**Fig. 3c**).

Here, the outer cell envelopes determined by LabelFreeTracker (see **Fig. 3d**) allow us to quantify expression levels of a fluorescent marker for *Axin2* [13], a Wnt target gene determining stemness [113]. The high *Axin2* of the red lineage matches with its proliferative nature (**Fig. 3c**, bottom), and its position in the crypt (**Fig. 3c**, top). Notably, the blue lineage shows a sharply decreasing *Axin2* signal, which may seem at odds with its proliferative nature, but is in fact consistent with its movement up the crypt-villus axis. The green lineage also shows decreasing *Axin2*, indeed in line with its non-proliferative nature. These and other data features, including the sub-cell-cycle

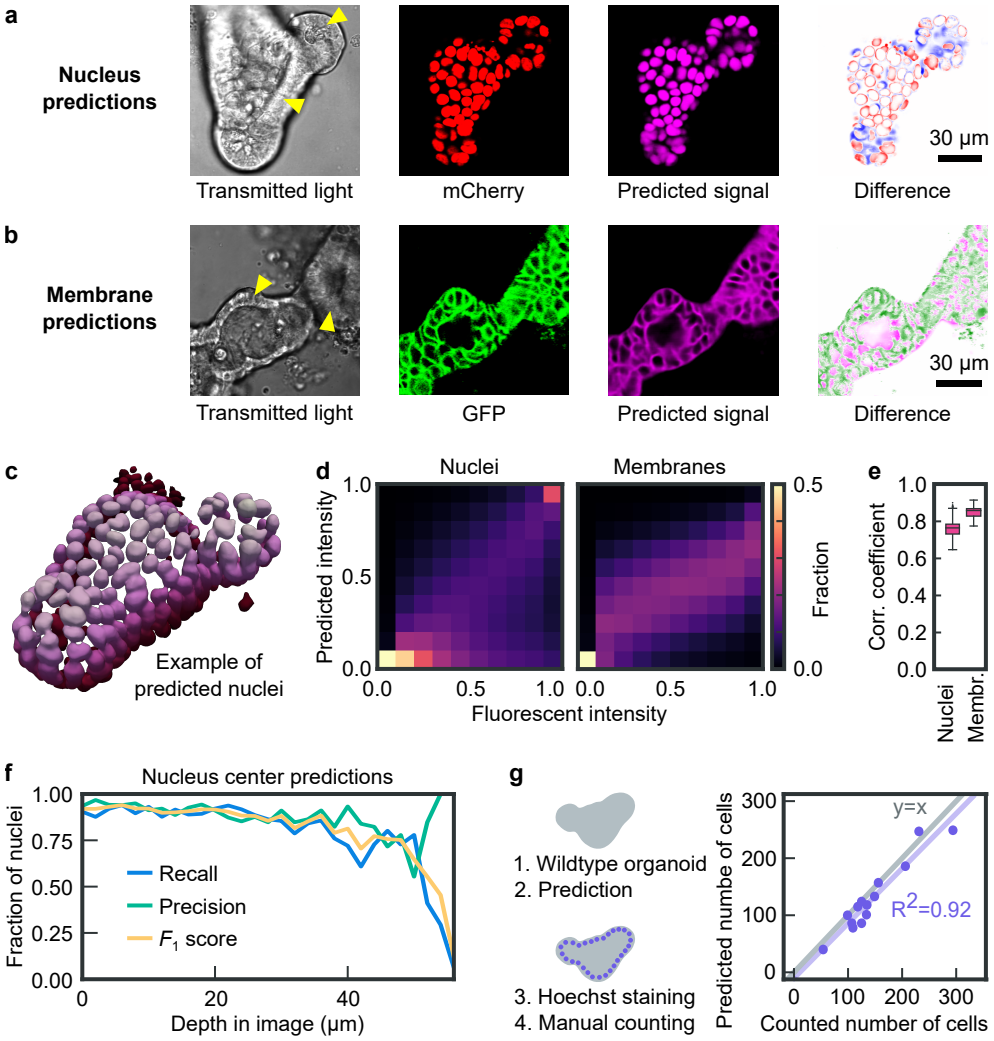


Figure 2 • Nucleus and membrane prediction results

(a) 2D slices of the transmitted light and the fluorescent and predicted nuclear signals, and their difference. **(b)** Like panel a, but for the membrane signal, and for a different organoid. **(c)** 3D example image of nuclei predicted from transmitted light. The picture shows the same organoid as panel a. **(d)** Pixel correlation plot, comparing the fluorescent and predicted signals after scaling down the images 8 times in the x and y direction. The nucleus comparison uses three organoids (327 images in total) while the membrane comparison uses two organoids (136 images). **(e)** Pearson correlation coefficients, calculated for all individual pixels in each image pair (fluorescent versus predicted). Same dataset as panel d. **(f)** Automatic nucleus detection results along the depth in the image, compared to manual annotations for 15 time points from 3 organoids. **(g)** Control with 14 wildtype organoids distributed over three wells. We first predicted the amount of nuclei, and verified the count by adding Hoechst staining afterwards.

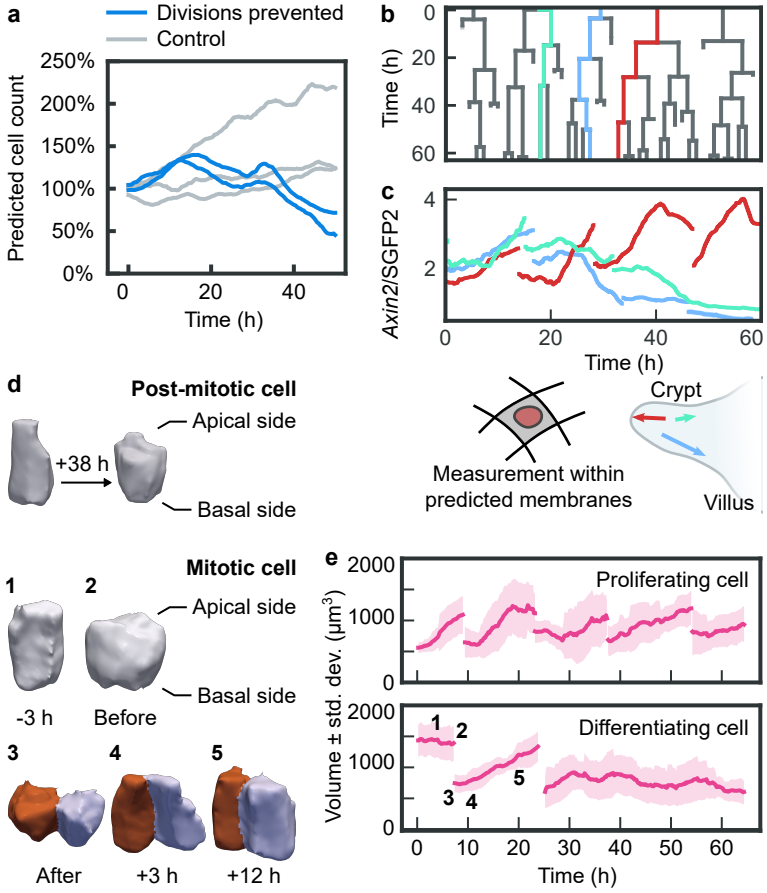


Figure 3 • Applications of label-free tracking in organoids
(a) Number of predicted cells in movies of growing wildtype organoids. **(b)** Selected lineage trees of a single organoid obtained using predicted nuclei and manual error correction. **(c)** *Axin2*/SGFP2 levels of two cells that were highlighted in panel b. Cells were tracked using a predicted nucleus signal, and segmented using a predicted membrane signal. The cell drawn in red remains in the crypt, the cell drawn in blue moves to the villus, and as a result it loses its *Axin2* signal. The cells were aligned such that they both divide at time zero. The measurement uses an 8-hour local average. **(d)** Predicted cell segmentation before. One cell is shown directly after its last cell division, another cell in between two cell divisions. For this last cell, the segmentation shows the rounding up of the cell during mitosis, and the growth on the apical side of the cell after mitosis. **(e)** Predicted cell volume. Pink line is a 6-hour local average. The numbers next to the line match the number of panel d.

dynamics, could have important functional implications that this method can address.

Fourth, determination of the outer cell envelope allows direct study of cell volume and shape (**Fig. 3d, e**), which has been challenging for cells embedded within epithelial tissues [140]. It also shows the fidelity of the approach, as it can only be successful if entire cell envelopes are determined on all sides, and for many successive points in time. The results show how cells rounds up during mitosis, and show narrow apical sides directly after, before growing again in volume (**Fig. 3d**). The results also evidence notably sudden volume growth arrests directly after the last division in a differentiating lineage (**Fig. 3e**, bottom). Interestingly, in proliferating lineages, the data shows appreciable variability in cell volume at birth (just after division), which can jeopardize cell size homeostasis if unregulated. Cell size homeostasis mechanisms have in recent years generated significant attention but have only been studied scarcely in tissues [140] due to the technical difficulties that we aimed to address here.

In conclusion, we have demonstrated how for the dense, 3D tissue of intestinal organoids it is possible to reconstruct the nuclear and membrane signals, and use those signals for tracking and segmentation. The method can be reproduced with little effort, as training data is generated by simply imaging fluorescent organoids, and the convolutional neural network is trained using standard techniques, for which we provide the software. We expect our method to provide a useful alternative for adding fluorescent nuclear and membrane markers.

6.2 | Materials and methods

6.2.1 Organoid culturing

Organoids were cultured as described in **Section 4.4.2** on page 79.

6.2.2 Organoid time lapse microscopy

To perform a time lapse experiment, we first seeded mechanically dissociated organoids in 40 µL gel in four-well chambered cover glass (Cellvis). This step was performed one to two days before the start of the time lapse experiment. To prevent the gel from solidifying immediately, seeding was performed on a cold block. The imaging well was put in a fridge for ten minutes to allow the organoids to sediment down. Afterwards, the chambered cover glass was put in the incubator to allow the gel to solidify. After leaving the plates in an incubator for 20 to 30 minutes, the growth medium was added.

Imaging was performed on a Nikon A1R MP microscope with a 40x oil immersion objective (NA = 1.30). 31 z-slices with 2 µm step size were taken per organoid every 12 minutes. Experiments were performed at 37 °C and 5-8% CO_2, achieved by using a stage-top incubator (Okolab). An imaging well typically contains over 50 organoids, of which we imaged about 20 per experiment. We selected for a diverse set of organoids, as our goal was to obtain a diverse training set. We set the imaging settings such that the transmitted light was never oversaturated or undersaturated. For the fluorescent

channels for imaging H2B and E-cadherin, we aimed to make all cells anywhere in the imaged part of the organoid appear as bright cells, at the cost of oversaturation elsewhere in the organoid.

6.2.3 Training data for nucleus and membrane predictions

For acquiring training data of nuclear fluorescence, we used H2B-mCherry / Lgr5-GFP murine intestinal organoids, which were a gift of Norman Sachs and Joep Beumer (Hubrecht Institute, The Netherlands). For acquiring training data of membrane fluorescence, we used H2B-mCherry / E-cadherin-GFP organoids, which were a gift of Daniel Krueger (Hubrecht Institute, The Netherlands). We did not use this line to create nucleus prediction training data, as the nuclear marker of this line is less bright compared to our other line.

We used E-cadherin as a membrane marker, even though it technically only marks cell-cell junctions. In practice, this is not a problem, as the intestinal epithelium is a tightly connected epithelium. As a result, often the entire cell surface is still visible in images of E-cadherin (**Fig. S1b**). We used this fluorescence marker because it produced a bright signal in our organoids, making it easy to image.

We have collected 3894 pairs of images of transmitted light/nuclei of 22 organoids in total. 13 organoids were newly imaged, the rest is from a previous chapter (**Chapter 2**). In addition, we have collected 1908 pairs of images of transmitted light/membranes of 25 organoids. We normalized the time lapse so that at all time points and all depths layers the fluorescent signal is saturated. Different normalizations were sometimes used for different depths and time points. We split the 512 by 512 by 32 pixel images in smaller patches of 128 by 128 by 16 pixels.

For many of our images, the organoids do not occupy the entire view, which caused a considerable amount of patches to show no fluorescence. To speed up the training process, we removed almost all of these patches. The criterium we used was that a patch was considered black if the brightest pixel found in the patch is lower than half of the highest fluorescence in the entire image. We still kept 5% of the black patches in the training set, so that the network still learns to not paint any fluorescence outside the organoid.

We first trained the network that predicted the nuclear signal from transmitted light, and then the network that predicted the membrane signal. For the network predicting the membrane signal, we used the trained network that predicts nuclei as a starting point, so that the cell membrane network can reuse cell detection knowledge of the nuclear network. We trained the network to minimize the mean squared error loss between the fluorescent signal and the predicted signal.

6.2.4 Setup of the fluorescence prediction neural network

The network was implemented in Tensorflow [64] 2.5.1. The network architecture follows a U-net type architecture, with a downwards block, an upwards block and skip

connections between the blocks. The input of the network consists of 3D grayscale images of 128 by 128 by 16 pixels. The downwards part of the network consists of four repeating units, with each unit consisting of 2 3D convolutions, maximum pooling and batch normalization. The upwards part consists of four repeating units, with each unit consisting of a transposed convolution, followed by two convolutions, followed by batch normalization. At the end of the network, a final convolution is done. The network is optimized using the Adam optimizer [141] with the mean squared error loss function.

The intensities of the transmitted light images are offset and multiplied such that the average of an entire 3D stack of a single time point is 0, and the standard deviation is 1. For the fluorescent images, the images were normalized from 0 to 1, with the background at zero. Normalization factors were chosen by hand for several slices during the time lapse, at both different times and depths. For other slices, the factors were calculated from the manually set factors by linear interpolation.

The only data augmentation that was done was horizontal and vertical flipping. Rotations and scaling data augmentations were omitted on purpose, as to not remove any details provided by individual pixels.

6.2.5 Correlating predicted and fluorescent images

The Pearson correlation coefficient was calculated for each image pair (fluorescent image and predicted image) separately, using the following equation:

$$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}}$$

Here, x is the fluorescent image, y the predicted image, x_i and y_i are the intensities in the image at pixel i, $\overline{x}$ and $\overline{y}$ is the mean intensity of the entire image and $\sum_i$ represents the sum over all pixels i in the image.

The above equation yields a single r value for each image. By calculating r for each image, we obtain the distributions that are displayed in **Fig. 2e** and **Fig. S2a-b**. The panels follow the standard boxplot settings of the software library Matplotlib [60] version 3.6.0, which are as follows. The box runs from the lower quartile value Q_1 to the upper quartile value Q_3 of the correlations coefficients, and the line is placed at the median. The lower whiskers extend up to the lowest data point that is higher on the y-axis than $Q_3 + 1.5 \cdot (Q_3 - Q_1)$, while the higher whiskers extend up the lowest data point that is higher on the y-axis than $Q_1 - 1.5 \cdot (Q_3 - Q_1)$. Any data point beyond the whiskers is displayed as a small dot.

The pixel correlation plots in **Fig. 2d** were created by first downscaling each image eight times in the x and y direction (not z), scaling the intensities of the images from 0 to 1, and then comparing the intensities. The counts are normalized such that each column sums to 1. The scaling was done to allow the predictions to have an offset of a few pixels compared to the fluorescent signal.

6.2.6 Predicting nucleus center positions

For training the nucleus center position network, we used pairs of transmitted light images, and images with bright spots at the center of the nuclei. Because transmitted light images are diverse in appearance, we needed a large training set. A manually created training set was therefore deemed unfeasible.

Instead, we reused the training set used for training nuclear signal prediction network. We first used OrganoidTracker (**Chapter 2**) to automatically find the nucleus centers in our training set, based on the fluorescent nuclei. We did not correct the nucleus center positions, which means that our training data will contain some errors. Then, we created images with bright Gaussian spots at the location of the nucleus centers.

$$G\left(\vec{x}\right) = e^{(\vec{x}-\vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{x}-\vec{\mu})}$$

Here, $\vec{x}$ represents the (x, y, z) coordinates of some pixel in the image and $\vec{\mu} = (\mu_x, \mu_y, \mu_z)$ is the center position of a nucleus in pixel coordinates. $(\vec{x} - \vec{\mu})^T$ denotes the transpose of $\vec{x} - \vec{\mu}$, which results in a row vector. The covariance matrix $\mathbf{\Sigma}$ has the same value for all nuclei, namely

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

with σ_x and σ_y being 4 px (1.28 µm), and σ_z being 1 px (equal to 2 µm in this case, as the resolution is lower in the z-direction).

We trained the neural network on the training set for 10 epochs. A large part of the images containing the nucleus centers was black, as there are no nucleus centers within a few pixels. Therefore, the network tended to just predict all pixels as black, and not draw any nucleus centers at all. We solved this issue by reducing the weight of black pixels by a factor of 4.

To test the network, we applied the network to three organoids. We first obtained a ground truth for five time points in each of these organoids, by manually annotating all cell centers in these time points. We selected the time points at 0%, 25%, 50%, 75% and 100% of the time lapse duration.

To obtain a performance metric, we used the following method to calculate the number of true positives (TP), false positives (FP) and false negatives (FN):

Any nucleus center detected by the neural network was assigned to the closest nucleus center from the ground truth, under the condition that the distance was no longer than 5 µm. Every nucleus center cannot have more than one assignment. Each successful assignment was a true positive. Then, any manually tracked nucleus center that was left with no assignments became a false negative. Finally, any nucleus center from the neural network that was left with no assignments was regarded as a false positive.

The precision is then calculated as $\text{precision} = \text{TP}/(\text{TP} + \text{FP})$ and the recall is calculated as $\text{recall} = \text{TP}(\text{TP} + \text{FN})$. The F_1 score is the harmonic average of the

precision and recall: $F_1 = 2 \cdot (\text{precision} \cdot \text{recall})/(\text{precision} + \text{recall})$.

6.2.7 Control experiment with wildtype organoids

To demonstrate that transmitted light-based prediction of nuclei works in completely label-free organoids, wildtype (WT) mouse intestinal organoids were mechanically disrupted and embedded in BME2 gel with growth medium. Two days after passaging, Z-stacks of label-free organoids were acquired. After the first image acquisition, cells were incubated for 15-25 minutes in live-cell staining medium (growth medium with Hoechst 34580 and 0.8 µg/mL Alexa 647 conjugated anti-mouse CD326 antibody), before a second Z-stack was acquired with the same imaging conditions as the label-free image acquisition.

6.2.8 Semi-automated cell tracking

We used the position detection network (see "Predicting nucleus center positions" on page 128) to generate a set of positions. Our next step was to link these positions over time.

As a first step, we estimated the chance $P(L|d)$, which is the chance of two cell positions in subsequent time points being of the same cell (i.e. having a link L), given the distance d between both positions. This function is expected to start at a high chance for $d = 0$, and then drop off towards zero for increasing d. To estimate this function, we used the tracking data of one organoid with a nuclear marker, that was tracked previously (**Chapter 2**). For every position, we recorded the distance to the nearest position in the previous time point, as well as the distance to any position that is at most twice as far. We also recorded whether that position represented the same cell, which allowed us to estimate the chance $P(L|d)$.

We noticed that the center positions of the predicted nuclei were less accurate than the center positions of fluorescent nuclei. This resulted in larger apparent cell movements in between time points, which made the function $P(L|d)$ underestimate the chance of two nucleus positions being of the same cell. To correct for this, we simulated errors in the nucleus center positions of our previously tracked organoid. We added $(\texttt{rand}() - \texttt{rand}()) \cdot 1.6\,\mu\text{m}$ to the x and to the y coordinate of each position, with $\texttt{rand}()$ a function that returns a uniformly distributed random number from 0 to 1. For the z coordinate, we choose a random integer from 1 to 6 (inclusive); if the number was 5, we added 2 µm (corresponding to one z layer) to the z coordinate, and if the number was 6, we subtracted 2 µm.

Having obtained a representation for $P(L|d)$, we used this as input for the track creation algorithm of Haubold et al. [49]. This resulted in tracks without divisions, as we did not have a method of calculating the chance of a giving cell undergoing a division. Using OrganoidTracker (**Chapter 2**), we manually corrected these tracks for five lineages, thereby adding the missing cell divisions.

6.2.9 Cell segmentation

The following procedure was carried out for each individual time point.

We start by loading the predicted membrane image, which is a 3D image. Next, we obtain a mask of the organoid, which is done as follows. This image is thresholded at 10% of the maximum intensity, separating the image in a foreground and background. In each xy plane of the image, the holes in the foreground are closed. Then, the largest object in the 3D image is regarded as the organoid. We make this organoid mask a bit smaller and smoother by eroding 4 pixels of this mask on the x and y axis.

To segment all individual cells, we perform a watershed segmentation within the organoid mask. The watershed landscape is simply the predicted membranes, and the seeds from which the watershed starts are formed by the nucleus centers. The 3D watershed algorithm was provided by the Mahotas software library [59].

Note that the predicted membrane signal creates a smooth border around the entire cell, which made it a good signal to use as a watershed landscape. Using the original E-cadherin-GFP signal as a watershed landscape would be more challenging, for two reasons. First, the fluorescent signal is more noisy. Second, the signal does not always form a complete border around the cells, as E-cadherin is a cell-cell adhesion protein. Therefore, the protein is absent at the basal side of the organoid.

6.2.10 Measuring cell volume

Cell volumes were measured in organoids with a nuclear marker, which were previously tracked using OrganoidTracker (**Chapter 2**). The organoids contained no membrane marker.

For measuring the cell volume, we segmented the cells using the predicted membrane signal (see above) and counted the number of pixels per cell. We used a six-hour moving average to smooth the volume over time.

6.2.11 Imaging and measuring *Axin2* reporter organoids

Axin2$^{P2A\text{-}rtTA3\text{-}T2A\text{-}3xNLS\text{-}SGFP2}$; *tetO-Cre*; *Rosa26*^{mTmG} murine intestinal organoids were a gift from the Van Amerongen lab (Universiteit van Amsterdam, The Netherlands) and were described earlier [13]. The organoids were passaged one day before the experiment, and were imaged over an entire weekend as described above.

We performed semi-automatic tracking on these organoids (see above), and corrected the tracks for five larger lineage trees (Figure 2b). We then segmented all cells at all time points (see above), and used the obtained masks to measure the total *Axin2*/SGFP2 signal of every cell.

We noticed that a specific type of cell segmentation error was causing issues with the GFP measurement. Sometimes, the membranes between two cells was missing, likely because the neural network could not locate it. This caused the volume of one cell to be

assigned to a neighbor of the cell. This error was largely eliminated by averaging the predicted membranes with a distance map, which was 0 at the center of the cell, and then increased linearly up to 1 at a distance of 7 µm. In this average, the membranes had a weight of 75%, and the distance map a weight of 25%.

The downside of adding the distance map to the watershed landscape is that for cells that are both large and stretched, it causes them to appear more spherical than they are. For measuring the intensity of a nuclear-localized signal, this is not an issue, but for measuring the volume of the cell it is. Therefore, we only used the distance map for measuring *Axin2*/SGFP2 (Figure 2c), and not for measuring the cell volume (Figure 2e).

The *Axin2*/SGFP2 signals were then normalized as follows. The lowest measured *Axin2*/SGFP2 signal in the time lapse was assumed to correspond to the background. This background, measured in intensity per pixel, was then subtracted from every measured *Axin2*/SGFP2 signal. The intensities where then multiplied by a factor such that the median intensity of the organoid, across all time points, was 1.

To reduce noise in the measured *Axin2* signal, we applied a moving average with an eight-hour time window.

6.3 | Supplementary Information

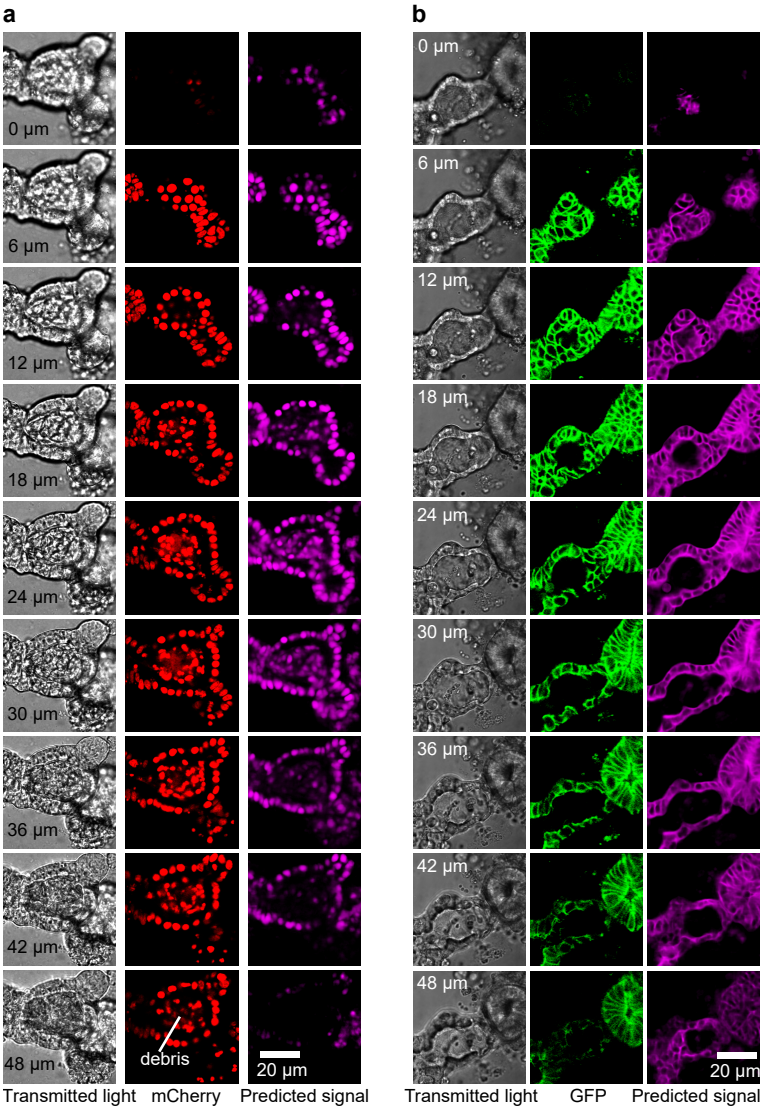


Figure S1 • Z-slices of two organoids with a hard-to-interpret transmitted light signal

The organoid on the left **(a)** contains a nuclear marker and the organoid on the right **(b)** contains a membrane marker. Zooming in on the image is recommended.

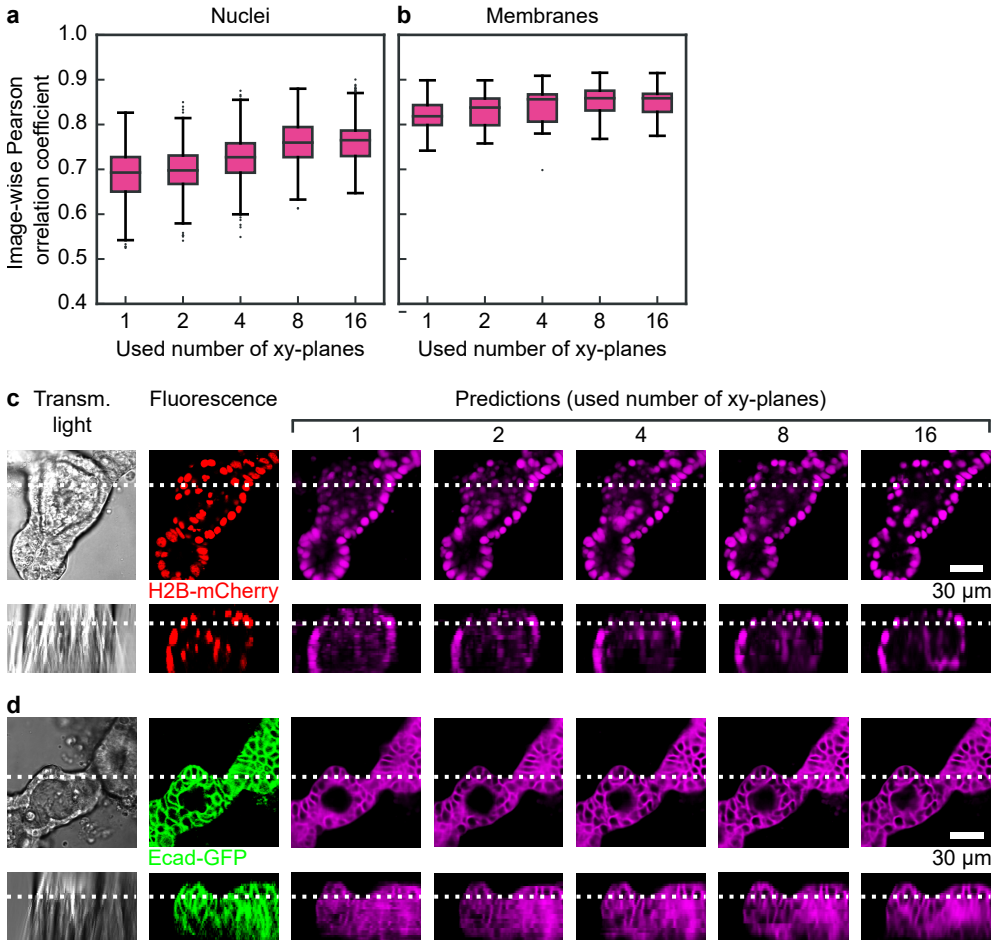


Figure S2 • Network performance for providing varying amounts of xy-planes
(a) Pearson correlation coefficient, calculated for each 3D image pair, comparing fluorescent and predicted nucleus images. **(b)** Like panel a, but for membrane images. **(c)** Example results for nucleus prediction, using various amounts of xy-planes. **(d)** Like panel c, but for membrane images.

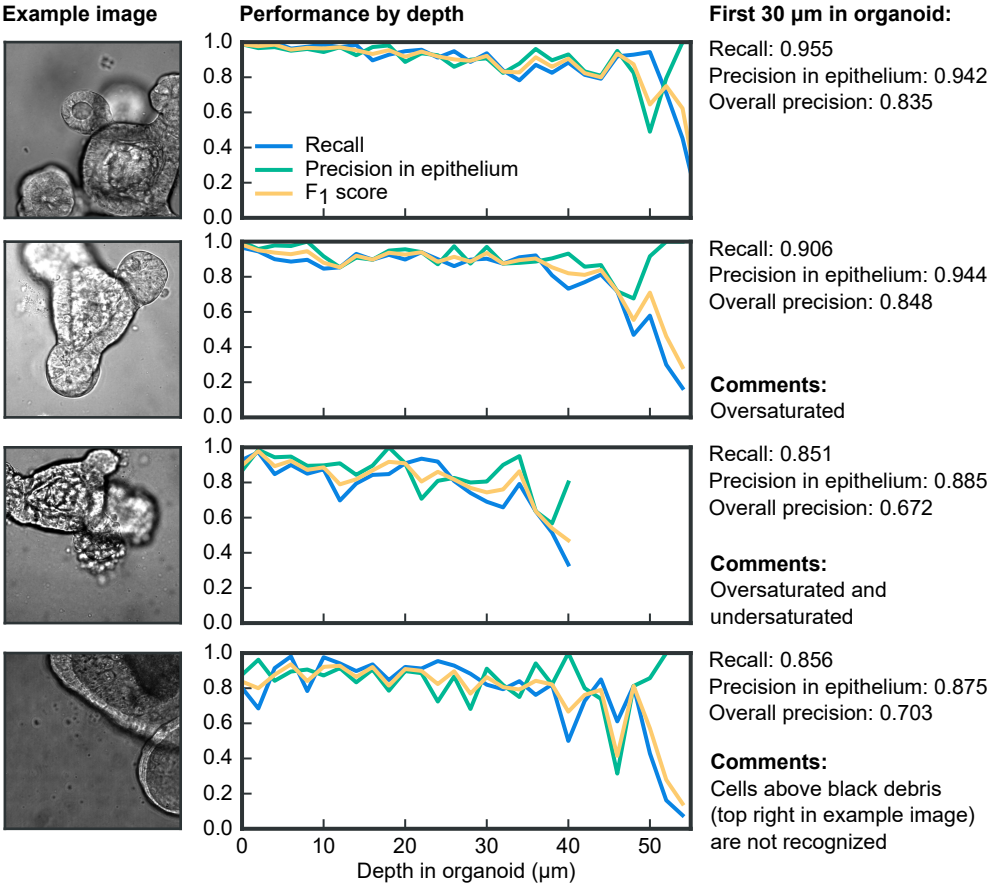


Figure S3 • Cell detection performance for different organoids

The performance varies for different imaging settings and for the amount of debris in the lumen.

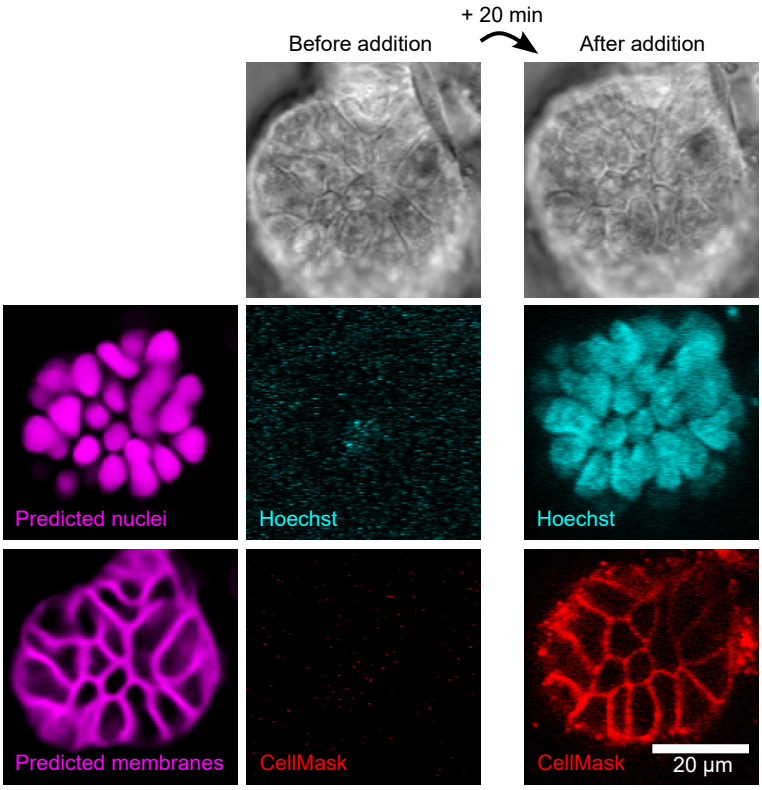


Figure S4 • Control with wildtype organoids
We performed this control to show that the neural network is independent of any trace that the fluorescent reporters might leave behind in the transmitted light images. First wildtype organoids were imaged, then Hoechst (nuclear) and CellMask (membrane) dyes were added and finally the organoids were imaged again.

7 | Conclusion and outlook

In this chapter, we briefly summarize the main discoveries included in this thesis. We look at the implications of our results, and suggest several concrete new experiments.

In this thesis, we have gained important insights in how homeostasis in the small intestines is maintained. The studies are made possible by a powerful cell tracking approach applied on intestinal organoids, where all cells in organoid crypts are followed over time. We first developed an automated cell tracker, which can be used to accurately find cells, and follow them over time (**Chapter 2**). This cell tracker uses a convolutional neural network with a graph-based trajectory optimization, an approach that is now becoming more common [134, 142]. We are still developing this cell tracker, and we expect the software to remain in use for a long time.

Next, we performed a theoretical study to find how stem cell self-renewal and differentiation are balanced (**Chapter 3**). Our goal was to find under which regime the fluctuations in the number of cells are minimized. Previously, it was assumed that asymmetric cell divisions (divisions where only one of the daughter cells is allowed to continue proliferating) lead to a system with the lowest fluctuations. However, this is only true for the case where cells cannot respond to their environment. For the common case of a stem cell niche of fixed size, cells within the niche should proliferate maximally, and outside the niche cells should proliferate minimally. This leads to a system containing two groups of cells with maximally different proliferative behavior, and without asymmetric divisions. We expect all tissues that use a stem cell niche for continuous self-renewal to follow this approach.

For the intestinal organoids, we confirmed using cell tracking that this was indeed the case. Intestinal organoids are often seen as chaotic structures that grow uncontrollably [143]. However, within a single crypt, the cells appear to follow a strict organization **Chapter 4**. The number of proliferating cells appears to be approximately constant, and this number appears to be dictated by the number of Paneth cells within the crypt. Moreover, organoids are indeed divided into two zones, the proliferative and non-proliferative zone, with a rather sharp boundary between those. While we could not track cells *in vivo* in mice, using a lineage tracing approach we obtained clone size distributions, which indeed carry the signature of a two-compartment system with symmetric divisions. Our results demonstrate that the intestinal stem cell niche is a

well-controlled environment, that follows an optimized strategy to maintain itself.

A question that remains is how the cells are able to reliably decide between proliferation and differentiation. It is known that Wnt signaling controls stemness [122]. Previously it was thought that Wnt signaling follows a spatial model, i.e. the further a cell resides from a source of Wnt, the lower its Wnt target gene activation would be. However, the spatial model is incompatible with recent experimental observations: Farin *et al.* [104] had shown that Wnt proteins do not spread over multiple cells, and Böttcher *et al.* [118] had shown that two cell types, namely Paneth and enteroendocrine cells, already start differentiating close to the source of Wnt proteins. Using cell tracking of a marker of Wnt signaling, and using laser ablation studies, we show that Wnt signaling operates using a timer model. Cells that lose contact with the source of Wnt signaling, Paneth cells in organoids, start decreasing their Wnt levels, like a hourglass losing sand. If a cell does not regain Paneth cell contact in time, it will differentiate. This timer model provides an attractive alternative for the often-assumed spatial model for cell fate, with contact time deciding cell fate, instead of spatial distance. We expect more signaling systems to follow such a model.

In the final chapter, we find that it is possible to track and segment cells using only transmitted light images, without needing a fluorescent signal. This is a surprising result, and it allows one to track organoids without a nuclear or membrane marker. Although this type of tracking is more error-prone than tracking using fluorescent markers, it is still possible to track cells for over 60 hours. In addition, the automatic segmentation provides a useful way to estimate the shape and volume of cells over time, and to measure signals within the cell.

The above discoveries help us in understanding intestinal organoid growth and homeostasis, and by extension, improve our understanding of the small intestines. With our approach, we propose several more questions that we could answer:

The role of the transit-amplifying compartment. In the theoretical model of **Chapter 3**, we simplified the small intestines into two compartments, one for stem cells and one for differentiated cells. However, most cell types do not immediately go from being a stem cell to being a fully differentiated cell. For example, enterocytes go through a variable number of cell cycles before maturing [123, 144], in which the cells are called transit-amplifying cells. This result is somewhat puzzling, as based on the model proposed in **Chapter 3**, one would expect cells to differentiate immediately, or at least have a fixed number of division rounds before they mature. To reconcile these observations with the model, we could work out several ways to include a transit-amplifying compartment in the model, and explore how the number of transit-amplifying cells could be regulated.

Interplay of Wnt and Notch signaling in cell type determination. In **Chapter 5** we observed that cells can start their differentiation close to Paneth cells, as long as they are not in sustained contact with Paneth cells. In addition, our lab [123] showed that all investigated cell types directly differentiate from stem cells. Previously, this was known only for Paneth and enteroendocrine cells [118]. In our observations, a

multipotent progenitor cell, like a secretory cell that can produce both goblet and enteroendocrine cells, does not exist. Instead, the eventual cell type of a progenitor cell is already set early on, and all of its offspring will have the same cell type.

Therefore, we have a strong suspicion that cell decide on their cell type next to the most far-out Paneth cell of a crypt. At this location, it would be Notch signaling would send cells towards secretory or absorptive fate [112]. To confirm this, one would need to construct a line with reporters for both Notch and Wnt signaling, and ideally also a nuclear marker. One would then need to make timelapse images of this line, ideally overweekend movies. One would first identify the moment of differentiation of each cell, which one could define as the moment where Paneth cell contact fell below 50% for the last 10 hours. Next, one would record Notch signaling around and before this time, and search for correlations with the eventual cell type, as determined by antibody staining. Such an experiment would show whether intestinal cell use neighbor inhibition through Notch signaling for controlling secretory cell numbers, as shown for various embryos [145], or whether secretory cell numbers are controlled using another method, like asymmetric cell divisions.

Generation of new Paneth cells using laser ablation. Paneth cells are the only cell type that can differentiate while Wnt signaling is high [112]. Since Paneth cells themselves produce Wnt proteins, on the first sight this seems like a mechanism that would result in exponential growth in the number of Paneth cells. However, creating Paneth cells also requires low Notch signaling. Since Paneth cells activate Notch signaling in their neighbors, Paneth cells effectively inhibit the creation of new Paneth cells, thus preventing exponential growth.

In vivo, Wnt is supplied too by mesenchymal cells, which do not supply Notch signaling. As a result, if a crypt at some point contains no Paneth cells, a number of stem cells will quickly differentiate into Paneth cells, thus restoring the health of the crypt [93]. Since intestinal organoids in standard medium have no other Wnt source than Paneth cells, we can expect that a crypt without Paneth cells will not survive.

However, we can expect one repair mechanism. Notch signaling requires cell-cell contact through membrane proteins, so Notch signaling is lost immediately once contact with a source of Notch signaling is lost. However, as we have seen in **Chapter 5**, Wnt signaling is reduced gradually in neighbor cells. Therefore, if a Paneth cell is violently ablated, for a few hours the original neighbor cells of the Paneth cell will experience low Notch signaling, but still high Wnt signaling. This should result in the formation of new Paneth cells. Therefore, the sudden death of a Paneth cell should promote the formation of new Paneth cells, thus providing a useful repair mechanism for the crypt. Using the laser ablation technique described in **Chapter 5**, we should be able to test whether this mechanism works.

Volume control in the intestinal epithelium. Using a cell membrane signal, it is possible to segment cells, which makes it possible to measure their volume. So far, little research has been done on cell volume control in animal cells. In bacteria, cell volume is easily regulated by the cell wall, but animal cells lack such a cell wall. Therefore,

they control their volume using osmosis and protein contents [146, 147]. Conflicting results have been obtained on how cells regulate their volume. Either cells grow to a certain size before they divide, or cells add a fixed volume during their cell cycle [140].

Especially in epithelia, regulating cell size is important, as otherwise holes could be created in the epithelium. We could test how the volume of cells changes along the epithelium. Are cells in the villus larger than cells in the crypt? Does the cell volume depend on the cell type? Do cells in environments with low crowding assume a larger volume, or does the epithelium only become thinner? Does cell volume play a role in determining cell fate? By combining laser ablation (**Chapter 5**) with cell volume measurements (**Chapter 6**), we should be able to investigate these questions.

Dankwoord

Allen, heel hartelijk dank voor de mooie tijd bij AMOLF! Ik ben echt blij dat ik hier heb kunnen werken, voor twee geweldige begeleiders en een hele stapel fantastische collega's. Om de herinneringen niet verloren te laten gaan, hier een schets van de sfeer:

Ik denk dat de oorzaak klein is, en dat het blaft.

Burak, over waarom Jeroen niet op AMOLF kan zijn. Jeroen had toen net een schattige puppy aangeschaft

Eigenlijk mag dit niet van ICT, maar onze groep slaat de timelapse-filmpjes op in Dropbox. Onbeperkt ruimte, en veel goedkoper!

Corentin

Hij is zó niet woke!

Dhawal, nadat Sinterklaas de mensen op AMOLF heeft begroet met "Goedemorgen, jongens en meisjes!"

Bacteriën zijn eigenlijk heel makkelijk. Eenvoudig te kweken, eenvoudig te manipuleren en eenvoudig te analyseren. Echt, het onderzoek gaat zo makkelijk!

Fotis op DutchBiophysics 2019, nog aan het begin van zijn promotie

Het is raar, het lijkt er zoveel op maar is tegelijk zo anders.

Guizela, nadat anderen haar artikel hebben aangepast

Voor dit soort problemen mag je me wel vaker roepen!

Hinco, nadat hij binnen 30 seconden een gasfles open weet te krijgen, na 30 minuten vruchteloos proberen door Max en mezelf

We zijn wat later begonnen, maar beneden is er nu prosecco om de ERC te vieren!

Jack, nadat bekend is geworden dat Sander een beurs van de Europese Onderzoeksraad heeft ontvangen

Hoi allen, heeft iemand een gigantische stapel enkelzijdige organoïdenpapers geprint bij de "col-vide1" printer? Als niemand zich meldt dan ga ik ze als kladpapier gebruiken!

Jeroen

Het is niet helemaal duidelijk hoe de COR nou eigenlijk werkt.

Jesse, probeert iets op de agenda van de ondernemingsraad van NWO te krijgen

Die organoïden zijn een soort monsters. Ze eten al het andere onderzoek van de groep op!

Joleen

Nee, voor deze chemicaliën hoef je eigenlijk het licht niet uit te doen. Het is meer een ritueel, een ritueel in het donker. Net als het Christendom!

Kasper

Ik wil niet opscheppen, maar ik heb hier wel de allernieuwste iPhone. Jawel, net gekocht. Kunnen we hier niet mee filmen?

Kathi bij het maken van filmpje van het lab

Ik zag de pandemie gewoon aankomen! Ik had moeten investeren in mondkapjes!

Luca na de eerste lockdown

Prins Bernhard enteignen und BlackRock Raus!

Max, over de huizenmarkt

Ik zie jou niet vaak koffie drinken, maar als je koffie drinkt, dan is het gelijk met dit gigantische volume.

Pascal

Ok, er kwam iets tussendoor. Nu een reparatie-mannetje aan de deur. Zou jij misschien bullet-point-gewijs willen opschrijven wat er in die mail moet? Dan zal ik hem vanavond netjes formuleren, vertalen en verzenden.

Ramon, voorzitter ondernemingsraad

Ons labuitje moet beter zijn dan die van Jeroens groep!

Sander

We kregen het scherm niet aan de praat, dus toen hebben we Paula om hulp gevraagd.

Simone en Leonie schakelden de instituutsdirecteur in

Gebruik indien mogelijk alleen Biowizard 2 vandaag (en de volgende dagen als het nog warm is) om de warmteproductie in het lab te verminderen. Het is nu 34 graden in het lab, laten we dat niet nog meer maken :P

Yvonne

Publications

Related to this thesis

Betjes, M.A., Zheng, X., **Kok, R.N.U.**, Van Zon, J.S., & Tans, S.J. (2021). Cell tracking for organoids: lessons from developmental biology. *Frontiers in Cell and Developmental Biology*, 1347. (**Chapter 1**)

Kok, R.N.U., Hebert, L., Huelsz-Prince, G., Goos, Y.J., Zheng, X., Bozek, K., ... & Van Zon, J.S. (2020). OrganoidTracker: Efficient cell tracking using machine learning and manual error correction. *PLOS One*, 15(10), e0240802. (**Chapter 2**)

Kok, R.N.U., Tans, S.J., & Van Zon, J.S. (2022). Minimizing cell number fluctuations in self-renewing tissues with a stem cell niche. *Physical Review E*, 108(6), 064403. (**Chapter 3**)

Huelsz-Prince, G., **Kok, R.N.U.**, Goos, Y.J., Bruens, L., Zheng, X., Ellenbroek, S.I., ... & Van Zon, J.S. (2022). Mother cells control daughter cell proliferation in intestinal organoids to minimize proliferation fluctuations. *eLife*, 11, e80682. (**Chapter 4**)

Kok, R.N.U., Zheng, X., Clement, T.V.M., Spoelstra, W.K., Tans, S.J., & Van Zon, J.S. (2024). Loss of Wnt signaling starts a differentiation timer in the intestinal crypt. *Manuscript in preparation*. (**Chapter 5**)

Kok, R.N.U., Spoelstra, W.K., Betjes, M.A., Van Zon, J.S., & Tans, S.J. (2024). Label-free cell tracking in intestinal organoids. *Manuscript in preparation*. (**Chapter 6**)

Other publications

De Boer, F.Y., **Kok, R.N.U.**, Imhof, A., & Velikov, K.P. (2018). White zein colloidal particles: synthesis and characterization of their optical properties on the single particle level and in concentrated suspensions. *Soft Matter*, 14(15), 2870-2878.

Artegiani, B., Hendriks, D., Beumer, J., **Kok, R.N.U.**, Zheng, X., Joore, I., ... & Clevers, H.C. (2020). Fast and efficient generation of knock-in human organoids using homology-independent CRISPR-Cas9 precision genome editing. *Nature Cell Biology*, 22(3), 321-331.

Curriculum Vitae

Rutger N.U. Kok

1995	Geboren te Naarden
2007-2013	Gymnasium, Goois Lyceum, Bussum (cum laude)
2013-2016	B.Sc. scheikunde, Universiteit Utrecht
2016-2018	M.Sc. chemische wetenschappen, Universiteit Utrecht (cum laude)
2018-2022	Promotieonderzoek, AMOLF, Amsterdam De resultaten van het hier uitgevoerde onderzoek zijn in dit proefschrift beschreven.

Bibliography

1. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-Free Prediction of Three-Dimensional Fluorescence Images from Transmitted-Light Microscopy. *Nature Methods* **15,** 917–920. doi:10.1038/s41592-018-0111-2 (2018).
2. Christiansen, E. M. *et al.* In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* **173,** 792–803.e19. doi:10.1016/j.cell.2018.03.040 (2018).
3. Baron, C. S. & van Oudenaarden, A. Unravelling Cellular Relationships during Development and Regeneration Using Genetic Lineage Tracing. *Nature Reviews Molecular Cell Biology* **20,** 753–765. doi:10.1038/s41580-019-0186-3 (2019).
4. Wagner, D. E. & Klein, A. M. Lineage Tracing Meets Single-Cell Omics: Opportunities and Challenges. *Nature Reviews Genetics.* doi:10.1038/s41576-020-0223-2 (Mar. 2020).
5. McKenna, A. & Gagnon, J. A. Recording Development with Single Cell Dynamic Lineage Tracing. *Development* **146,** 1–10. doi:10.1242/dev.169730 (June 2019).
6. Snippert, H. J. *et al.* Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. *Cell* **143,** 134–144. doi:10.1016/j.cell.2010.09.016 (2010).
7. Pei, W. *et al.* Polylox Barcoding Reveals Haematopoietic Stem Cell Fates Realized in Vivo. *Nature* **548,** 456–460. doi:10.1038/nature23653 (2017).
8. Mohme, M. *et al.* Optical Barcoding for Single-Clone Tracking to Study Tumor Heterogeneity. *Molecular Therapy* **25,** 621–633. doi:10.1016/j.ymthe.2016.12.014 (2017).
9. Weber, K. *et al.* RGB Marking Facilitates Multicolor Clonal Cell Tracking. *Nature Medicine* **17,** 504–509. doi:10.1038/nm.2338 (2011).
10. Weinreb, C. & , Alejo Rodriguez-Fraticelli, Fernando D. Camargo, A. M. K. State To Fate During Differentiation. *Science* **367,** 755. doi:10.1126/science.aaw3381.Weinreb (2020).
11. Mayr, U. *Transient and Heterogeneous YAP1 Activity Drives Self-Organization in Intestinal Organoid Development* PhD thesis (Universität Basel, 2019).
12. Lindenburg, L. & Merkx, M. Engineering Genetically Encoded FRET Sensors. *Sensors (Switzerland)* **14,** 11691–11713. doi:10.3390/s140711691 (2014).
13. van de Moosdijk, A. A. A., van de Grift, Y. B. C., de Man, S. M. A., Zeeman, A. L. & van Amerongen, R. A Novel Axin2 Knock-in Mouse Model for Visualization and Lineage Tracing of WNT/CTNNB1 Responsive Cells. *Genesis* **58,** 1–14. doi:10.1002/dvg.23387 (2020).

14. Artegiani, B. *et al.* Fast and Efficient Generation of Knock-in Human Organoids Using Homology-Independent CRISPR-Cas9 Precision Genome Editing. *Nature Cell Biology* **22,** 321–331. doi:10.1038/s41556-020-0472-5 (2020).

15. Pennisi, E. Development Cell by Cell. *Science* **362,** 1344–1345. doi:10.1126/science.362.6421.1344 (Dec. 2018).

16. Ulman, V. *et al.* An Objective Comparison of Cell-Tracking Algorithms. *Nature Methods* **14,** 1141–1152. doi:10.1038/nmeth.4473 (2017).

17. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The Embryonic Cell Lineage of the Nematode Caenorhabditis Elegans. *Developmental Biology* **100,** 64–119. doi:10.1016/0012-1606(83)90201-4 (1983).

18. Ritsma, L. *et al.* Intestinal Crypt Homeostasis Revealed at Single-Stem-Cell Level by in Vivo Live Imaging. *Nature* **507,** 362–365. doi:10.1038/nature12972 (2014).

19. Sidhaye, J. & Knoblich, J. A. Brain Organoids: An Ensemble of Bioassays to Investigate Human Neurodevelopment and Disease. *Cell Death & Differentiation* **28,** 52–67. doi:10.1038/s41418-020-0566-4 (Jan. 2021).

20. Clevers, H. C. Modeling Development and Disease with Organoids. *Cell* **165,** 1586–1597. doi:10.1016/j.cell.2016.05.082 (June 2016).

21. Rios, A. C. & Clevers, H. Imaging Organoids: A Bright Future Ahead. *Nature Methods* **15,** 24–26. doi:10.1038/nmeth.4537 (2018).

22. Svensson, C. M., Medyukhina, A., Belyaev, I., Al-Zaben, N. & Figge, M. T. Untangling Cell Tracks: Quantifying Cell Migration by Time Lapse Image Data Analysis. *Cytometry Part A* **93,** 357–370. doi:10.1002/cyto.a.23249 (2018).

23. Amat, F. *et al.* Fast, Accurate Reconstruction of Cell Lineages from Large-Scale Fluorescence Microscopy Data. *Nature Methods* **11,** 951–958. doi:10.1038/nmeth.3036 (2014).

24. Bao, Z. *et al.* Automated Cell Lineage Tracing in Caenorhabditis Elegans. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 2707–12. doi:10.1073/pnas.0511111103 (Feb. 2006).

25. Schiegg, M. *et al.* Graphical Model for Joint Segmentation and Tracking of Multiple Dividing Cells. *Bioinformatics* **31,** 948–956. doi:10.1093/bioinformatics/btu764 (2015).

26. McDole, K. *et al.* In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell* **175,** 859–876.e33. doi:10.1016/j.cell.2018.09.031 (2018).

27. McKinley, K. L. *et al.* Cellular Aspect Ratio and Cell Division Mechanics Underlie the Patterning of Cell Progeny in Diverse Mammalian Epithelia. *eLife* **7,** e36739. doi:10.7554/eLife.36739 (June 2018).

28. Serra, D. *et al.* Self-Organization and Symmetry Breaking in Intestinal Organoid Development. *Nature.* doi:10.1038/s41586-019-1146-y (2019).

29. Held, M., Santeramo, I., Wilm, B., Murray, P. & Lévy, R. Ex Vivo Live Cell Tracking in Kidney Organoids Using Light Sheet Fluorescence Microscopy. *PLoS ONE* **13.** doi:10.1371/journal.pone.0199918 (July 2018).

30. Xing, F., Xie, Y., Su, H., Liu, F. & Yang, L. Deep Learning in Microscopy Image Analysis: A Survey. *IEEE Transactions on Neural Networks and Learning Systems* **29,** 4550–4568. doi:10.1109/TNNLS.2017.2766168 (Oct. 2018).

31. Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A. & Olivo-Marin, J. C. Imagining the Future of Bioimage Analysis. *Nature Biotechnology* **34,** 1250–1255. doi:10.1038/nbt.3722 (2016).

32. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: A Generalist Algorithm for Cellular Segmentation. *Nature Methods* **18,** 100–106. doi:10.1038/s41592-020-01018-x (2021).

33. Wolff, C. *et al.* Multi-View Light-Sheet Imaging and Tracking with the MaMuT Software Reveals the Cell Lineage of a Direct Developing Arthropod Limb. *eLife* **7,** 1–31. doi:10.7554/eLife.34410 (2018).

34. He, Z. *et al.* Lineage Recording in Human Cerebral Organoids. *Nature Methods* **19,** 90–99. doi:10.1038/s41592-021-01344-8 (2022).

35. Kiviet, D. J. *et al.* Stochasticity of Metabolism and Growth at the Single-Cell Level. *Nature* **514,** 376–379. doi:10.1038/nature13582 (2014).

36. Rodríguez-Colman, M. J. *et al.* Interplay between Metabolic Identities in the Intestinal Crypt Supports Stem Cell Function. *Nature* **543,** 424–427. doi:10.1038/nature21673 (2017).

37. Prewitt, J. M. S. & Mendelsohn, M. L. The Analysis of Cell Images. *Annals of the New York Academy of Sciences* **128,** 1035–1053. doi:10.1111/j.1749-6632.1965.tb11715.x (1966).

38. Zhu, R., Sui, D., Qin, H. & Hao, A. *An Extended Type Cell Detection and Counting Method Based on FCN* in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)* (IEEE, Washington, DC, Oct. 2017), 51–56. doi:10.1109/BIBE.2017.00-79.

39. Xie, W., Noble, J. A. & Zisserman, A. Microscopy Cell Counting and Detection with Fully Convolutional Regression Networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* **6,** 283–292. doi:10.1080/21681163.2016.1149104 (2018).

40. Xie, Y. *et al.* Efficient and Robust Cell Detection: A Structured Regression Approach. *Medical Image Analysis* **44,** 245–254. doi:10.1016/j.media.2017.07.003 (Feb. 2018).

41. Pan, X. *et al.* Cell Detection in Pathology and Microscopy Images with Multi-Scale Fully Convolutional Neural Networks. *World Wide Web* **21,** 1721–1743. doi:10.1007/s11280-017-0520-7 (2018).

42. Caicedo, J. C. *et al.* Nucleus Segmentation across Imaging Experiments: The 2018 Data Science Bowl. *Nature Methods* **16,** 1247–1253. doi:10.1038/s41592-019-0612-7 (2019).

43. Hay, E. A. & Parthasarathy, R. Performance of Convolutional Neural Networks for Identification of Bacteria in 3D Microscopy Datasets. *PLOS Computational Biology* **14,** 1–17. doi:10.1371/journal.pcbi.1006628 (2018).

44. Dunn, K. W. *et al.* DeepSynth: Three-dimensional Nuclear Segmentation of Biological Images Using Neural Networks Trained with Synthetic Data. *Scientific Reports* **9,** 18295. doi:10.1038/s41598-019-54244-5 (Dec. 2019).

45. Ho, D. J. *et al. Center-Extraction-Based Three Dimensional Nuclei Instance Segmentation of Fluorescence Microscopy Images* in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (IEEE, May 2019), 1–4. doi:10.1109/BHI.2019.8834516. arXiv: 1909.05992.

46. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science* **9351,** 234–241. doi:10.1007/978-3-319-24574-4_28 (2015).

47. Khan, F. A., Voß, U., Pound, M. P. & French, A. P. Volumetric Segmentation of Cell Cycle Markers in Confocal Images. *bioRxiv,* 1–10. doi:10.1101/707257 (2019).

48. Liu, R. *et al. An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution* in *Advances in Neural Information Processing Systems* (eds Bengio, S. *et al.*) (Curran Associates, Inc., 2018), 9605–9616.

49. Haubold, C., Aleš, J., Wolf, S. & Hamprecht, F. A. *A Generalized Successive Shortest Paths Solver for Tracking Dividing Targets* in *European Conference on Computer Vision* (Springer, Amsterdam, The Netherlands, 2016), 566–582. doi:10.1007/978-3-319-46478-7_35.

50. Winter, M. *et al.* Vertebrate Neural Stem Cell Segmentation, Tracking and Lineaging with Validation and Editing. *Nature Protocols* **6,** 1942–1952. doi:10.1038/nprot.2011.422 (Dec. 2011).

51. Tinevez, J. Y. *et al.* TrackMate: An Open and Extensible Platform for Single-Particle Tracking. *Methods* **115,** 80–90. doi:10.1016/j.ymeth.2016.09.016 (2017).

52. Han, H., Wu, G., Li, Y. & Zi, Z. eDetect: A Fast Error Detection and Correction Tool for Live Cell Imaging Data Analysis. *iScience* **13,** 1–8. doi:10.1016/j.isci.2019.02.004 (Mar. 2019).

53. Wallach, T. E. & Bayrer, J. R. Intestinal Organoids: New Frontiers in the Study of Intestinal Disease and Physiology. *Journal of Pediatric Gastroenterology and Nutrition* **64,** 180–185. doi:10.1097/MPG.0000000000001411 (2017).

54. van der Walt, S. *et al.* Scikit-Image: Image Processing in Python. *PeerJ* **2,** e453. doi:10.7717/peerj.453 (June 2014).

55. Matula, P. *et al.* Cell Tracking Accuracy Measurement Based on Comparison of Acyclic Oriented Graphs. *PLOS ONE* **10.** doi:10.1371/journal.pone.0144959 (2015).

56. Toyoshima, Y. *et al.* Accurate Automatic Detection of Densely Distributed Cell Nuclei in 3D Space. *PLoS Computational Biology* **12,** 1–20. doi:10.1371/journal.pcbi.1004970 (2016).

57. Gehart, H. & Clevers, H. Tales from the Crypt: New Insights into Intestinal Stem Cells. *Nature Reviews Gastroenterology & Hepatology* **16,** 19–34. doi:10.1038/s41575-018-0081-y (Jan. 2019).

58. Clevers, H. C. & Bevins, C. L. Paneth Cells: Maestros of the Small Intestinal Crypts. *Annual Review of Physiology* **75,** 289–311. doi:10.1146/annurev-physiol-030212-183744 (2013).

59. Coelho, L. P. Mahotas: Open Source Software for Scriptable Computer Vision. *Journal of Open Research Software* **1,** e3. doi:10.5334/jors.ac (July 2013).

60. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9,** 90–95. doi:10.1109/MCSE.2007.55 (2007).

61. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering* **13,** 22–30. doi:10.1109/MCSE.2011.37 (2011).

62. Bradski, G. & Kaehler, A. *Learning OpenCV: Computer Vision with the OpenCV Library* (O'Reilly Media, Inc., 2008).

63. Riverbank Computing Limited. *What Is PyQt?* https://www.riverbankcomputing.com/software/pyqt/intro.

64. Martín Abadi *et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* 2015.

65. Gohlke, C. *Tifffile* https://pypi.org/project/tifffile/.

66. Rybarski, J. & Verweij, R. *Nd2reader* https://rbnvrw.github.io/nd2reader/.

67. Ayachit, U. *The ParaView Guide: A Parallel Visualization Application* (Kitware, 2015).

68. McCarthy, N., Kraiczy, J. & Shivdasani, R. A. Cellular and Molecular Architecture of the Intestinal Stem Cell Niche. *Nature Cell Biology* **22,** 1033–1041. doi:10.1038/s41556-020-0567-z (2020).

69. Weterings, S. D. C., van Oostrom, M. J. & Sonnen, K. F. Building Bridges between Fields: Bringing Together Development and Homeostasis. *Development* **148,** dev193268. doi:10.1242/dev.193268 (July 2021).

70. Simons, B. D. & Clevers, H. Strategies for Homeostatic Stem Cell Self-Renewal in Adult Tissues. *Cell* **145,** 851–862. doi:10.1016/j.cell.2011.05.033 (2011).

71. Jilkine, A. Mathematical Models of Stem Cell Differentiation and Dedifferentiation. *Current Stem Cell Reports* **5,** 66–72. doi:10.1007/s40778-019-00156-z (2019).

72. Chatzeli, L. & Simons, B. D. Tracing the Dynamics of Stem Cell Fate. *Cold Spring Harbor Perspectives in Biology* **12,** a036202. doi:10.1101/cshperspect.a036202 (June 2020).

73. Klein, A. M. & Simons, B. D. Universal Patterns of Stem Cell Fate in Cycling Adult Tissues. *Development* **138,** 3103–3111. doi:10.1242/dev.060103 (2011).

74. Kitadate, Y. *et al.* Competition for Mitogens Regulates Spermatogenic Stem Cell Homeostasis in an Open Niche. *Cell Stem Cell* **24,** 79–92.e6. doi:10.1016/j.stem.2018.11.013 (2019).

75. Clayton, E. *et al.* A Single Type of Progenitor Cell Maintains Normal Epidermis. *Nature* **446,** 185–189. doi:10.1038/nature05574 (2007).

76. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. *Science* **330,** 822–825. doi:10.1126/science.1196236 (Nov. 2010).

77. Klein, A. M., Doupé, D. P., Jones, P. H. & Simons, B. D. Kinetics of Cell Division in Epidermal Maintenance. *Physical Review E* **76,** 021910. doi:10.1103/PhysRevE.76.021910 (2007).

78. Sun, Z. & Komarova, N. L. Stochastic Modeling of Stem-Cell Dynamics with Control. *Mathematical Biosciences* **240,** 231–240. doi:10.1016/j.mbs.2012.08.004 (2012).

79. Corominas-Murtra, B. *et al.* Stem Cell Lineage Survival as a Noisy Competition for Niche Access. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 16969–16975. doi:10.1073/pnas.1921205117 (July 2020).

80. Cannataro, V. L., Mckinley, S. A. & St. Mary, C. M. The Implications of Small Stem Cell Niche Sizes and the Distribution of Fitness Effects of New Mutations in Aging and Tumorigenesis. *Evolutionary Applications* **9,** 565–582. doi:10.1111/eva.12361 (2016).

81. Cannataro, V. L., McKinley, S. A. & St. Mary, C. M. The Evolutionary Trade-off between Stem Cell Niche Size, Aging, and Tumorigenesis. *Evolutionary Applications* **10,** 590–602. doi:10.1111/eva.12476 (2017).

82. Shahriyari, L. & Komarova, N. L. The Role of the Bi-Compartmental Stem Cell Niche in Delaying Cancer. *Physical Biology* **12,** 055001. doi:10.1088/1478-3975/12/5/055001 (July 2015).

83. Sei, Y., Feng, J., Chow, C. C. & Wank, S. A. Asymmetric Cell Division-Dominant Neutral Drift Model for Normal Intestinal Stem Cell Homeostasis. *American Journal of Physiology Gastrointestinal Liver Physiology* **316,** G64–G74. doi:10.1152/ajpgi.00242.2018 (2019).

84. Yang, J., Plikus, M. V. & Komarova, N. L. The Role of Symmetric Stem Cell Divisions in Tissue Homeostasis. *PLoS Computational Biology* **11,** 1–30. doi:10.1371/journal.pcbi.1004629 (2015).

85. Azzalini, A. & Capitanio, A. Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **61,** 579–602. doi:10.1111/1467-9868.00194. arXiv: 0911.2093 (1999).

86. Huelsz-Prince, G. *et al.* Mother Cells Control Daughter Cell Proliferation in Intestinal Organoids to Minimize Proliferation Fluctuations. *eLife* **11,** e80682. doi:10.7554/eLife.80682 (Nov. 2022).

87. Inaba, M. & Yamashita, Y. M. Asymmetric Stem Cell Division: Precision for Robustness. *Cell Stem Cell* **11,** 461–469. doi:10.1016/j.stem.2012.09.003 (Oct. 2012).

88. Venkei, Z. G. & Yamashita, Y. M. Emerging Mechanisms of Asymmetric Stem Cell Division. *Journal of Cell Biology* **217,** 3785–3795. doi:10.1083/jcb.201807037 (Nov. 2018).

89. Li, L. & Xie, T. Stem Cell Niche: Structure and Function. *Annual Review of Cell and Developmental Biology* **21,** 605–631. doi:10.1146/annurev.cellbio.21.012704.131525 (Nov. 2005).

90. Lander, A. D., Gokoffski, K. K., Wan, F. Y. M., Nie, Q. & Calof, A. L. Cell Lineages and the Logic of Proliferative Control. *PLoS Biology* **7** (ed Stevens, C. F.) e1000015. doi:10.1371/journal.pbio.1000015 (Jan. 2009).

91. Rué, P. & Martinez Arias, A. Cell Dynamics and Gene Expression Control in Tissue Homeostasis and Development. *Molecular Systems Biology* **11,** 792. doi:10.15252/msb.20145549 (Feb. 2015).

92. Feller, W. Die Grundlagen Der Volterraschen Theorie Des Kampfes Ums Dasein in Wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheoretica* **5,** 11–40. doi:10.1007/BF01602932 (May 1939).

93. Farin, H. F., Van Es, J. H. & Clevers, H. Redundant Sources of Wnt Regulate Intestinal Stem Cells and Promote Formation of Paneth Cells. *Gastroenterology* **143,** 1518–1529.e7. doi:10.1053/j.gastro.2012.08.031 (2012).

94. Sato, T. *et al.* Paneth Cells Constitute the Niche for Lgr5 Stem Cells in Intestinal Crypts. *Nature* **469,** 415–418. doi:10.1038/nature09637 (2011).

95. Shoshkes-Carmel, M. *et al.* Subepithelial Telocytes Are an Important Source of Wnts That Supports Intestinal Crypts. *Nature* **557,** 242–246. doi:10.1038/s41586-018-0084-4 (May 2018).

96. Scoville, D. H., Sato, T., He, X. C. & Li, L. Current View: Intestinal Stem Cells and Signaling. *Gastroenterology* **134,** 849–864. doi:10.1053/j.gastro.2008.01.079 (Mar. 2008).

97. Winton, D. J., Ponder, B. A. J. & Clarke, B. C. Stem-Cell Organization in Mouse Small Intestine. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **241,** 13–18. doi:10.1098/rspb.1990.0059 (July 1990).

98. Sato, T. *et al.* Single Lgr5 Stem Cells Build Crypt-Villus Structures in Vitro without a Mesenchymal Niche. *Nature* **459,** 262–265. doi:10.1038/nature07935 (2009).

99. Doupé, D. P. *et al.* A Single Progenitor Population Switches Behavior to Maintain and Repair Esophageal Epithelium. *Science* **337,** 1091–1093. doi:10.1126/science.1218835 (Aug. 2012).

100. Klein, A. M., Nakagawa, T., Ichikawa, R., Yoshida, S. & Simons, B. D. Mouse Germ Line Stem Cells Undergo Rapid and Stochastic Turnover. *Cell Stem Cell* **7,** 214–224. doi:10.1016/j.stem.2010.05.017 (Aug. 2010).

101. Teixeira, V. H. *et al.* Stochastic Homeostasis in Human Airway Epithelium Is Achieved by Neutral Competition of Basal Cell Progenitors. *eLife* **2,** e00966. doi:10.7554/eLife.00966 (Oct. 2013).

102. Ragkousi, K. & Gibson, M. C. Cell Division and the Maintenance of Epithelial Order. *Journal of Cell Biology* **207,** 181–188. doi:10.1083/jcb.201408044 (Oct. 2014).

103. Swain, P. S. *Lecture Notes on Stochastic Models in Systems Biology* July 2016. doi:10.48550/arXiv.1607.07806. arXiv: 1607.07806 [q-bio].

104. Farin, H. F. *et al.* Visualization of a Short-Range Wnt Gradient in the Intestinal Stem-Cell Niche. *Nature* **530,** 340–343. doi:10.1038/nature16937 (Feb. 2016).

105. Langlands, A. J. *et al.* Paneth Cell-Rich Regions Separated by a Cluster of Lgr5+ Cells Initiate Crypt Fission in the Intestinal Stem Cell Niche. *PLOS Biology* **14** (ed Hynes, N.) e1002491. doi:10.1371/journal.pbio.1002491 (June 2016).

106. Itzkovitz, S., Blat, I. C., Jacks, T., Clevers, H. & Van Oudenaarden, A. Optimality in the Development of Intestinal Crypts. *Cell* **148,** 608–619. doi:10.1016/j.cell.2011.12.025 (2012).

107. Eisenhoffer, G. T. & Rosenblatt, J. Bringing Balance by Force: Live Cell Extrusion Controls Epithelial Cell Numbers. *Trends in Cell Biology* **23,** 185–192. doi:10.1016/j.tcb.2012.11.006 (Apr. 2013).

108. Messal, H. A. *et al.* Tissue Curvature and Apicobasal Mechanical Tension Imbalance Instruct Cancer Morphogenesis. *Nature* **566,** 126–130. doi:10.1038/s41586-019-0891-2 (2019).

109. Batlle, E. & Clevers, H. C. Cancer Stem Cells Revisited. *Nature Medicine* **23,** 1124–1134. doi:10.1038/nm.4409 (2017).

110. Wiese, K. E., Nusse, R. & van Amerongen, R. Wnt Signalling: Conquering Complexity. *Development* **145,** 1–9. doi:10.1242/dev.165902 (2018).

111. Rim, E. Y., Clevers, H. & Nusse, R. The Wnt Pathway: From Signaling Mechanisms to Synthetic Modulators. *Annual Review of Biochemistry* **91.** doi:10.1146/annurev-biochem-040320-103615 (June 2022).

112. Beumer, J. & Clevers, H. Cell Fate Specification and Differentiation in the Adult Mammalian Intestine. *Nature Reviews Molecular Cell Biology.* doi:10.1038/s41580-020-0278-0 (Sept. 2020).

113. Gregorieff, A. & Clevers, H. C. Wnt Signaling in the Intestinal Epithelium: From Endoderm to Cancer. *Genes and Development* **19,** 877–890. doi:10.1101/gad.1295405 (2005).

114. Spit, M., Koo, B.-K. & Maurice, M. M. Tales from the Crypt: Intestinal Niche Signals in Tissue Renewal, Plasticity and Cancer. *Open Biology* **8,** 180120. doi:10.1098/rsob.180120 (Sept. 2018).

115. Durand, A. *et al.* Functional Intestinal Stem Cells after Paneth Cell Ablation Induced by the Loss of Transcription Factor Math1 (Atoh1). *Proceedings of the National Academy of Sciences* **109,** 8965–8970. doi:10.1073/pnas.1201652109 (2012).

116. van Es, J. H. *et al.* Enteroendocrine and Tuft Cells Support Lgr5 Stem Cells on Paneth Cell Depletion. *Proceedings of the National Academy of Sciences,* 201801888. doi:10.1073/pnas.1801888117 (Dec. 2019).

117. Kim, T. H., Escudero, S. & Shivdasani, R. A. Intact Function of Lgr5 Receptor-Expressing Intestinal Stem Cells in the Absence of Paneth Cells. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 3932–3937. doi:10.1073/pnas.1113890109 (2012).

118. Böttcher, A. *et al.* Non-Canonical Wnt/PCP Signalling Regulates Intestinal Stem Cell Lineage Priming towards Enteroendocrine and Paneth Cell Fates. *Nature Cell Biology* **23,** 23–31. doi:10.1038/s41556-020-00617-2 (Jan. 2021).

119. Lustig, B. *et al.* Negative Feedback Loop of Wnt Signaling through Upregulation of Conductin/Axin2 in Colorectal and Liver Tumors. *Molecular and Cellular Biology* **22,** 1184–1193. doi:10.1128/MCB.22.4.1184-1193.2002 (Feb. 2002).

120. Wang, B. G., Riemann, I., Schubert, H., Halbhuber, K. J. & Koenig, K. In-Vivo Intratissue Ablation by Nanojoule near-Infrared Femtosecond Laser Pulses. *Cell and Tissue Research* **328,** 515–520. doi:10.1007/s00441-006-0367-1 (2007).

121. Seidlitz, T. *et al.* Human Gastric Cancer Modelling Using Organoids. *Gut* **68,** 207–217. doi:10.1136/gutjnl-2017-314549 (2019).

122. Schuijers, J. *et al.* Ascl2 Acts as an R-spondin/Wnt-responsive Switch to Control Stemness in Intestinal Crypts. *Cell Stem Cell* **16,** 158–170. doi:10.1016/j.stem.2014.12.006 (2015).

123. Zheng, X. *et al.* Following Cell Type Transitions in Space and Time by Combining Live Cell Tracking and Endpoint Cell Identity in Intestinal Organoids. *bioRxiv,* 15. doi:10.1101/2022.06.27.497728 (June 2022).

124. Haber, A. L. *et al.* A Single-Cell Survey of the Small Intestinal Epithelium. *Nature* **551,** 333–339. doi:10.1038/nature24489 (2017).

125. Gomez, H. F., Dumond, M. S., Hodel, L., Vetter, R. & Iber, D. 3D Cell Neighbour Dynamics in Growing Pseudostratified Epithelia. *eLife* **10,** e68135. doi:10.7554/eLife.68135 (Oct. 2021).

126. Azkanaz, M. *et al.* Retrograde Movements Determine Effective Stem Cell Numbers in the Intestine. *Nature* **607,** 548–554. doi:10.1038/s41586-022-04962-0 (July 2022).

127. Sonnen, K. F. & Janda, C. Y. Signalling Dynamics in Embryonic Development. *Biochemical Journal* **478,** 4045–4070. doi:10.1042/BCJ20210043 (2021).

128. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169,** 985–999. doi:10.1016/j.cell.2017.05.016 (2017).

129. Jho, E.-h. *et al.* Wnt/β-Catenin/Tcf Signaling Induces the Transcription of Axin2, a Negative Regulator of the Signaling Pathway. *Molecular and Cellular Biology* **22,** 1172–1183. doi:10.1128/MCB.22.4.1172-1183.2002 (Feb. 2002).

130. Gehart, H. *et al.* Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell* **176,** 1158–1173.e16. doi:10.1016/j.cell.2018.12.029 (2019).

131. Barrett, T. *et al.* NCBI GEO: Archive for Functional Genomics Data Sets - Update. *Nucleic Acids Research* **41,** 991–995. doi:10.1093/nar/gks1193 (2013).

132. Edgar, R. Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Research* **30,** 207–210. doi:10.1093/nar/30.1.207 (Jan. 2002).

133. Hao, Y. *et al.* Integrated Analysis of Multimodal Single-Cell Data. *Cell* **184,** 3573–3587.e29. doi:10.1016/j.cell.2021.04.048 (June 2021).

134. Malin-Mayor, C. *et al.* Automated Reconstruction of Whole-Embryo Cell Lineages by Learning from Sparse Annotations. *Nature Biotechnology.* doi:10.1038/s41587-022-01427-7 (Sept. 2022).

135. Sugawara, K., Çevrim, Ç. & Averof, M. Tracking Cell Lineages in 3D by Incremental Deep Learning. *eLife* **11,** 1–19. doi:10.7554/elife.69380 (2022).

136. Ulicna, K., Vallardi, G., Charras, G. & Lowe, A. R. Automated Deep Lineage Tree Analysis Using a Bayesian Single Cell Tracking Approach. *Frontiers in Computer Science* **3,** 734559. doi:10.3389/fcomp.2021.734559 (2021).

137. Medeiros, G. D. *et al.* Multiscale Light-Sheet Organoid Imaging Framework. *Nature Communications* **13,** 4864. doi:10.1038/s41467-022-32465-z (Aug. 2022).

138. Cross-Zamirski, J. *et al.* Label-Free Prediction of Cell Painting from Brightfield Images. *Scientific Reports* **12,** 10001. doi:10.1038/s41598-022-12914-x (2022).

139. Matthews, J. M. *et al.* OrganoID: A Versatile Deep Learning Platform for Tracking and Analysis of Single-Organoid Dynamics. *PLOS Computational Biology* **18** (ed Schneidman-Duhovny, D.) e1010584. doi:10.1371/journal.pcbi.1010584 (Nov. 2022).

140. Xie, S. & Skotheim, J. M. A G1 Sizer Coordinates Growth and Division in the Mouse Epidermis. *Current Biology* **30,** 916–924.e2. doi:10.1016/j.cub.2019.12.062 (Mar. 2020).

141. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization* 2015. arXiv: 1412.6980 [cs].

142. Löffler, K., Scherr, T. & Mikut, R. A Graph-Based Cell Tracking Algorithm with Few Manually Tunable Parameters and Automated Segmentation Error Correction. *PLOS ONE* **16** (ed Metze, K.) e0249257. doi:10.1371/journal.pone.0249257 (Sept. 2021).

143. Gjorevski, N. *et al.* Tissue Geometry Drives Deterministic Organoid Patterning. *Science* **375,** eaaw9021. doi:10.1126/science.aaw9021 (Jan. 2022).

144. Tóth, B., Ben-Moshe, S., Gavish, A., Barkai, N. & Itzkovitz, S. Early Commitment and Robust Differentiation in Colonic Crypts. *Molecular Systems Biology* **13,** 902. doi:10.15252/msb.20167283 (2017).

145. Sjöqvist, M. & Andersson, E. R. Do as I Say, Not(Ch) as I Do: Lateral Control of Cell Fate. *Developmental Biology* **447,** 58–70. doi:10.1016/j.ydbio.2017.09.032 (Mar. 2019).

146. Hoffmann, E. K., Lambert, I. H. & Pedersen, S. F. Physiology of Cell Volume Regulation in Vertebrates. *Physiological Reviews* **89,** 193–277. doi:10.1152/physrev.00037.2007 (Jan. 2009).

147. Delpire, E. & Gagnon, K. B. in *Current Topics in Membranes* 3–52 (Elsevier, 2018). doi:10.1016/bs.ctm.2018.08.001.