



## OPEN ACCESS

EDITED BY  
Philip Greulich,  
University of Southampton,  
United Kingdom

REVIEWED BY  
Jeremy B. A. Green,  
King's College London, United Kingdom  
Adriana Sánchez Danés,  
Champalimaud Foundation, Portugal

\*CORRESPONDENCE  
Steffen Rulands,  
rulands@lmu.de

SPECIALTY SECTION  
This article was submitted to Stem Cell  
Research, a section of the journal  
Frontiers in Cell and Developmental  
Biology

RECEIVED 26 September 2022  
ACCEPTED 14 November 2022  
PUBLISHED 14 December 2022

CITATION  
Dang Y and Rulands S (2022), Making  
sense of fragmentation and merging in  
lineage tracing experiments.  
*Front. Cell Dev. Biol.* 10:1054476.  
doi: 10.3389/fcell.2022.1054476

COPYRIGHT  
© 2022 Dang and Rulands. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Making sense of fragmentation and merging in lineage tracing experiments

Yiteng Dang<sup>1,2,3</sup> and Steffen Rulands<sup>1,2,4\*</sup>

<sup>1</sup>Max-Planck-Institute for the Physics of Complex Systems, Dresden, Germany, <sup>2</sup>Center for Systems Biology Dresden, Dresden, Germany, <sup>3</sup>Max-Planck-Institute for Molecular Cell Biology and Genetics, Dresden, Germany, <sup>4</sup>Arnold-Sommerfeld-Center for Theoretical Physics, Ludwig-Maximilians-Universität München, München, Germany

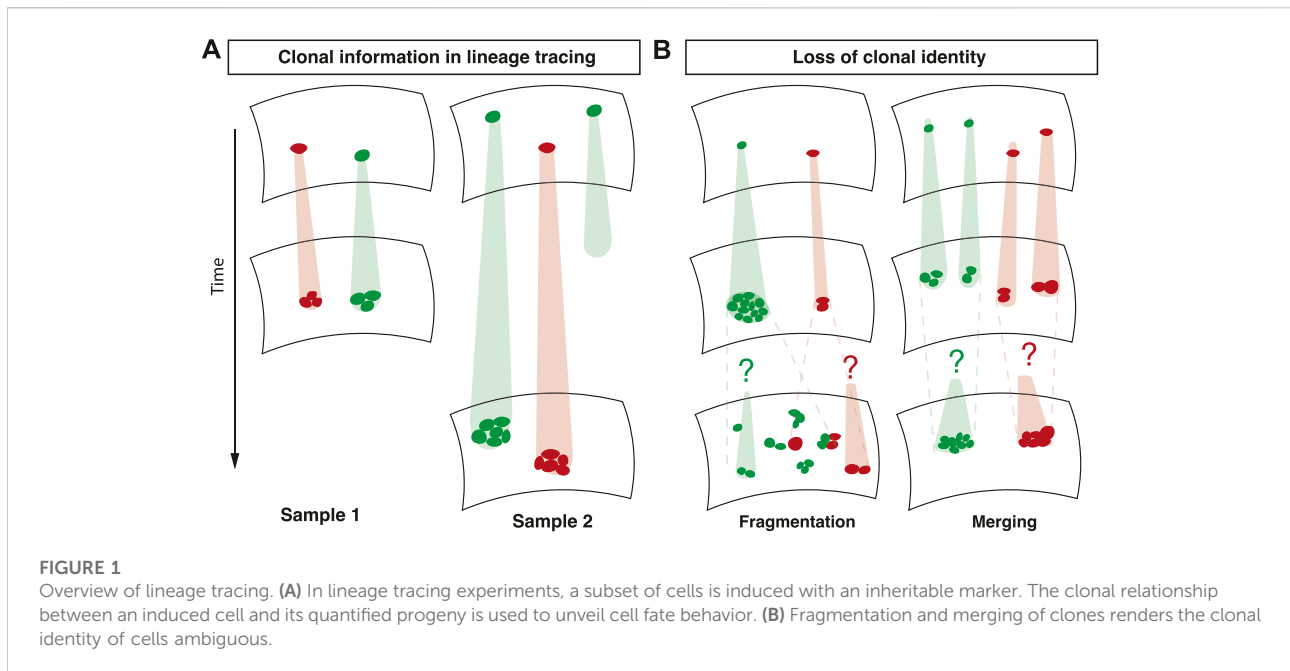
Lineage tracing experiments give dynamic information on the functional behaviour of dividing cells. These experiments therefore have become an important tool for studying stem and progenitor cell fate behavior *in vivo*. When cell proliferation is high or the frequency of induced clones cannot be precisely controlled, the merging and fragmentation of clones renders the retrospective interpretation of clonal fate data highly ambiguous, potentially leading to unguarded interpretations about lineage relationships and fate behaviour. Here, we discuss and generalize statistical strategies to detect, resolve and make use of clonal fragmentation and merging. We first explain how to detect the rates of clonal fragmentation and merging using simple statistical estimates. We then discuss ways to restore the clonal provenance of labelled cells algorithmically and statistically and elaborate on how the process of clonal fragmentation can indirectly inform about cell fate. We generalize and extend results from the context of their original publication.

## KEYWORDS

lineage tracing, stem cells, development, wound healing, cancer, stochastic modelling, statistics

## 1 Introduction

The development and maintenance of tissues relies on the tight regulation of cell migration, proliferation, differentiation and death. Cellular behaviour is typically linked to a progression through cell states in a lineage hierarchy, from multi-potent stem and progenitor cells to terminally differentiated cells. Research on the cellular programs that underlie these cell states has historically focused on molecular processes, such as the expression of genes in gene regulatory networks and epigenetic modifications of the DNA and chromatin. Recent advances in single-cell genomics have given rise to unprecedented possibilities in identifying these molecular states and associated molecular markers of cell states (Tam and Ho (2020); Tanay and Regev (2017); Stuart and Satija (2019)). Detailed molecular profiles do not, however, inform about the functional consequences of molecular cell states in terms of cell migration, proliferation, differentiation and death. By quantifying the



enrichment of marker genes obtained from single-cell sequencing compared to lists obtained from perturbation experiments cellular function can be inferred statistically in a correlative manner (Ashburner et al. (2000)).

A more direct study of the functional behaviour of cells requires time-resolved measurements on the functional level. Despite recent advances in microscopy, live imaging remains highly challenging in most tissues. Lineage tracing experiments, where a subset of cells is genetically labelled, give dynamic information on a functional level *in vivo* (Blanpain and Simons (2013); Simons and Clevers (2011)). In these experiments, transgenic mouse models are used to label a subset of cells with genetic markers. These markers are bequeathed to all progeny of a labelled cell, termed a clone. Lineage tracing comes in many flavours: single and multi-color constructs, genomic barcoding, combined with genetic perturbations (Rulands and Simons (2016); Yum et al. (2021)). All of these methods have in common that the number of cells in a given clone, the clone size, reflects a history of cell division, differentiation or cell death events between the time point of initial labelling (induction) and the time point of analysis (Ruske et al. (2020)). While a given clone size is compatible with multiple histories between labelling and analysis, statistical ensembles of clones allow to rigorously infer mechanisms of cell fate behaviour and lineage relationships from clonal fate data. As an example, the average size of clones is related to the rate and mode of cell proliferation. For example, in homeostasis, the balance between clonal loss and growth leads to a linear increase of the average clone size (Klein and Simons (2011)). If this

balance is tilted towards symmetric proliferation, such as is typical in early development or cancer, clone size distributions increase exponentially (Alcolea et al. (2014)). The probability distribution of clone sizes in addition reflects how cell fate is regulated, such as *via* extrinsic feedback or intrinsically. Methods from statistical physics and statistics have been successfully employed to unveil mechanisms of cell fate decisions in numerous tissues and contexts (Snippert et al. (2010); Doupé et al. (2010); Driessens et al. (2012); Shakiba et al. (2019); Scheele et al. (2017); Robertson et al. (2022)). Theoretical work has also highlighted generic features of clonal fate data (Klein et al. (2007); Yamaguchi et al. (2017); Rulands et al. (2018); Corominas-Murtra et al. (2020); Greulich et al. (2021)). Both in homeostasis and during tissue expansion clone size distributions converge over time to shapes that only partially depend on the details of cell fate regulation.

The dynamic information encoded in clonal fate data relies on the strict clonal relationship between a group of cells at the time of analysis and a single induced cell at an earlier time point (Figure 1A). The loss of this association significantly complicates the interpretation of lineage tracing experiments. This association can be lost in two ways: First, at the time point of analysis, a group of cells might be associated with multiple induced cells (merging) (Figure 1B). In tissues, where clonal identity is usually defined by spatial proximity of groups of labelled cells, clonal merging is a consequence of the joining of multiple clones labelled with the same fluorescent marker. Second, multiple groups of cells at the time of analysis might be associated with a single

induced cell (fragmentation) (Figure 1B). Clonal fragmentation is a consequence of large-scale tissue rearrangements, cell migration, or stochastic forces by cell divisions and cell loss (Ranft et al. (2010)). Cell intercalations, also termed T1 transitions, where cells rearrange and change neighbours, are also among the drivers of clonal fragmentation (Bocanegra-Moreno et al. (2022)). Clonal fragmentation therefore is particularly prominent in tissues with a high rate of cell proliferation, such as in development, regeneration or in tumours.

What can be learnt from lineage tracing data in these scenarios? A combination of theoretical and experimental work has shown that the ensuing size distributions tend to take a universal shape that is independent of the biological processes governing cell fate (Rulands et al. (2018)). This is because these processes lead to variability in clone sizes that dominates variability caused by other, cell-fate associated processes. Therefore, the merging and fragmentation of clones leads to the erasure of biological information contained in clone size distributions. In order to safeguard against drawing unfounded conclusions from clonal fate data it is therefore necessary to faithfully detect clonal fragmentation and merging in lineage tracing experiments and to restore clonality using statistical inference methods.

Here, we discuss statistical strategies that have recently been successfully employed to estimate rates of clonal fragmentation and merging and to restore the clonal provenance of fragmented cell clusters in a broad range of tissues. Where applicable we generalise these results beyond their original biological contexts. We argue that an estimation of the rates of clonal fragmentation and merging is an important step in the analysis of all lineage tracing experiments. In order for these methods to be applied by biologists who conduct these experiments we here focus on simplicity, intuition and applicability rather than quantitative exactness. We also discuss how the contiguity and the shape of clones provide information about morphogenetic behaviours such as oriented cell division and cell intercalation (Kauka et al. (2016)).

## 2 Sensing clonal fragmentation and merging

A qualitative assessment of the prevalence of clonal fragmentation and merging can be obtained by inspecting the shape of the probability distribution of the sizes of labelled cell clusters. Merging and fragmentation of clones lead to stereotypical shapes - in particular the log-normal shape - that can be indicative of these processes (Rulands et al. (2018)). In the following, we will discuss strategies of how the rates of merging and fragmentation in a given experiment can be estimated quantitatively.

### 2.1 Estimating the induction frequency

A given number of labelled cell clusters can originate from a combination of multiple induction, merging and fragmentation events. Therefore, an independent estimate of the induction frequency is an important step in the analysis of lineage tracing experiments with an uncertain degree of clonality. In Lescroart et al. (2014) the authors studied the temporal commitment of Mesp1 expressing progenitors to heart morphogenesis. They labelled cells at different time points during early gastrulation (E6.5-E7.5), where most of the labelled cells divide symmetrically and the processes leading to the formation of different heart compartments involve large-scale cell rearrangements. These processes lead to the fragmentation of labelled clones between the time of labelling and the time of analysis at E12.5. To deal with fragmentation, a possible strategy could be to aim to label 1 cell per heart on average, which would allow calculating the fragmentation rate by simply calculating the average number of clusters of labelled cells per sample. However, since induction is generally stochastic, this would have implied a significant fraction of unlabeled hearts of roughly 30%. Therefore, Lescroart et al. aimed for an average number of induced cells that is larger than one. Since in this case a given number of labelled cell clusters can originate from any combination of multiple induction and fragmentation events, the retrospective clonal analysis required an independent estimate of the induction frequency.

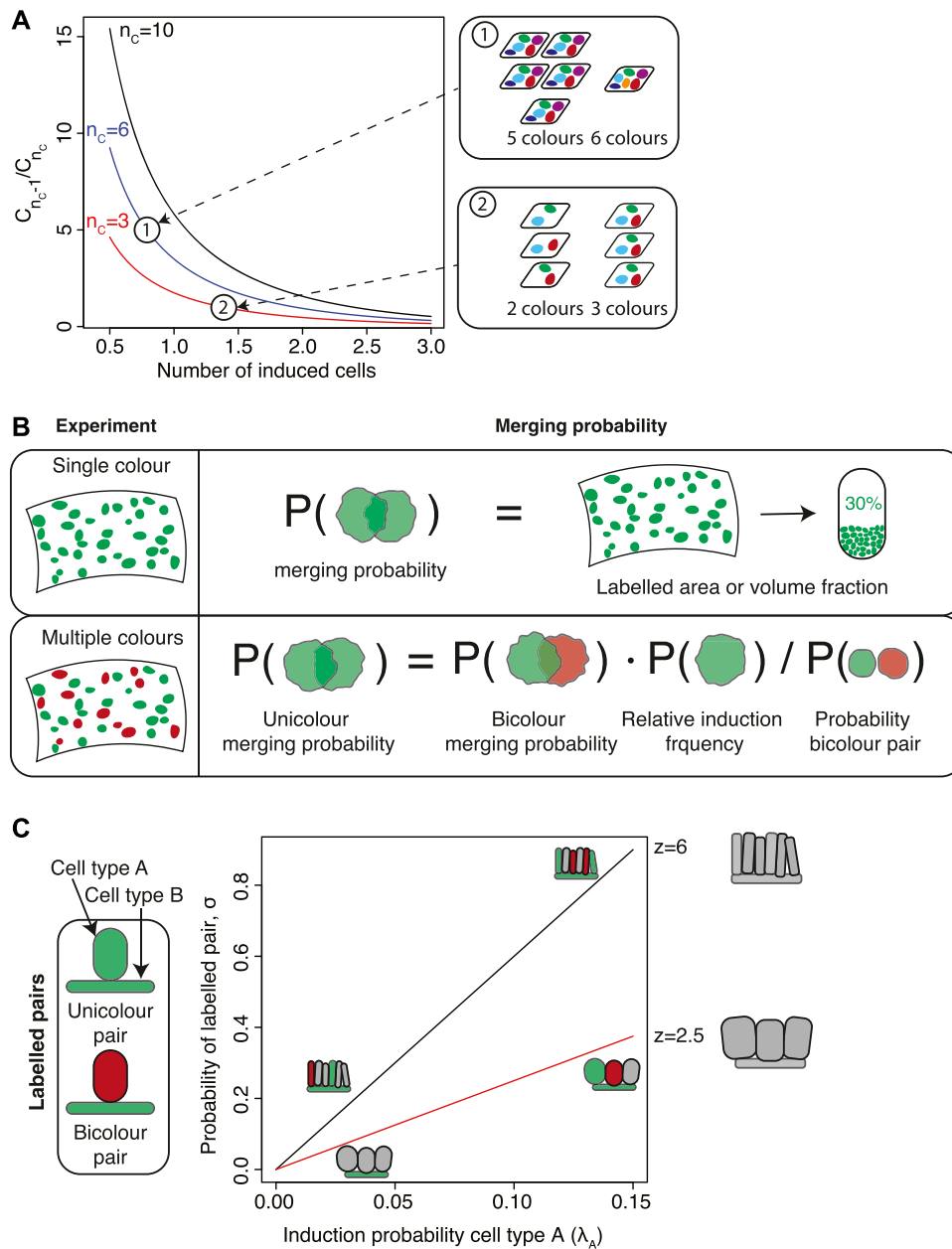
In order to obtain such an estimate, the authors made use of the multi-color labelling strategy based on the *Confetti* construct used in this study. If induction events in different colors are statistically independent, then the number of different colors in a sample follows a binomial distribution. For a labelling strategy with a total of  $n_C$  different colors the fraction of organs that is labelled in  $n$  different colors,  $C_n$  then is

$$C_n = \binom{n_C}{n} (1 - J)^n J^{n_C - n}, \quad (1)$$

where  $J$  is the probability that the tissue is unlabelled in a single color. If  $m_0$  denotes the average number of induced cells in a given tissue region then  $J = e^{-m_0}$  (Table 1). If only labelled organs are recorded, then Eq. 1 needs to be divided by the probability that an organ is labelled at all,  $1 - J^{n_C}$ . While Eq. 1 cannot in general be solved for  $m_0$ , we can solve for  $m_0$  in terms of the ratio  $C_n/C_{n+1}$ . Therefore, the ratio of the number of samples labelled in  $n$  colors,  $C_n$  and in  $n + 1$  colors  $C_{n+1}$  can be used to infer the average number of labelled cells per sample,  $m_0$  (Figure 2A),

$$m_0 = \log \left[ 1 + \frac{n+1}{n_C - n} \frac{C_{n+1}}{C_n} \right]. \quad (2)$$

In Lescroart et al. (2014) the authors applied this strategy by calculating the number of induced colors in all embryonic



**FIGURE 2**

Estimating the induction frequency and merging probability. **(A)** In experiments involving multiple fluorescent markers, the induction frequency can be inferred by comparing the numbers of samples with different numbers of colors, denoted  $C_n$  for  $n$  colors. The ratio  $C_{n_c-1}/C_{n_c}$  is most sensitive to the induction frequency at low values of the induction frequencies and in the case of many colors. **(B)** Calculating clonal merging rates from experiments. In unicolor experiments, the merging rate can be approximated as the area or volume fraction of the tissue that is labelled. In multicolor experiments, the unicolor merging probability is proportional to the bicolor merging probability times additional factors depending on the labelling frequencies of each individual color. **(C)** The probability that an induced cell of type B is in contact with an induced cell of type A depends the geometry of the tissue (coordination number) and the induction specificity. The insets show typical examples depicting different values of these parameters.

hearts analysed. In order to obtain a robust estimate of  $m_0$  this requires, however, a large number of animals. In larger tissues, for example in later stages of development or adulthood, an alternative strategy would be to define equally sized regions that are sufficiently separated to be statistically independent. Then one can quantify  $C_n$  and  $C_{n+1}$  across all of these regions. The average number of induced cells per region can then be calculated as above under the assumption that induction in all regions is statistically independent and with equal frequency.

## 2.2 Estimating the merging probability

The potential merging of induced clones disproportionately affects average clone sizes and the clone size distribution. For example, an estimate made from Monte Carlo simulations yields that a merging probability of 10% leads to a 20% increase in the average clone size. This is because larger clones have a higher probability of merging. Therefore, since the number and positions of induced clones are generally stochastic even for seemingly clonal experiments the merging probability should be controlled statistically. With an estimate of the induction frequency at hand the probability that a given cluster of labelled cells is polyclonal (merging probability) can, in principle, be estimated by comparing the number of induced cells with the number of observed cell clusters. If the number of observed cell clusters is significantly lower than the estimated number of induced cells then clonal merging is prevalent. Since this approach is based on a comparison between the time point of induction and the time point of analysis it implicitly relies on assumptions about the kinetic processes influencing the number of labelled cell clusters. Specifically, it neglects the possibility of clonal loss and fragmentation. In the following, we will discuss how the merging probability can be estimated independently of such assumptions.

### 2.2.1 Multi-color experiments

For multi-color experiments the probability of clonal merging can be obtained elegantly by comparing mergers between cell clusters of different colors. As a first rough estimate, one could make use of two different colors,  $c_1$  and  $c_2$ , that are expected to be induced with similar frequencies. Then, the merging probability of two clones with color  $c_1$  (unicolor mergers) is roughly equal to the probability that a cell cluster labelled in color  $c_1$  is in spatial contact with a cell cluster of color  $c_2$  (bicolor mergers). The latter is equal to the fraction of cell clusters of color  $c_1$  that are bicolor mergers, which can be easily be quantified from microscopy images.

If colors are not induced at similar frequencies, a more detailed analysis is necessary. Aragona et al. (2017) performed lineage tracing to study fate decisions of epidermal cells in the

mouse tail during wound healing. They labelled interfollicular epidermis (IFE) stem cells using a *Confetti* labelling strategy, and observed that the induced IFE cells formed elongated stripes that extended towards the center of the wound. In order to distinguish monoclonal stripes from stripes originating from the merger of multiple stripes, they devised a statistical framework to calculate merging rates in multicolor clonal data. To this end, Aragona et al. calculated the correction that is to be made to the heuristic argument presented in the previous paragraph due to the fact that the fraction of bicolor mergers is in itself influenced by the presence of mergers. If  $k_{bicolor}$  is the number of observed clusters that are bicolor mergers between a color  $c$  and any other color and  $k$  is the total number of clusters, then the fraction of unicolor mergers is

$$G_{unicolor}^c \approx \frac{k_{bicolor}}{k} \frac{r_c}{1 - \sum_c r_c^2}. \quad (3)$$

The nominator of the second term,  $r_c$ , is the relative frequency of clones of color  $c$  which rescales the fraction  $k_{bicolor}/k$  by the relative contribution of color  $c$ . The denominator corrects for the possibility of mergers when counting  $k$  (Figure 2B). This correction is particularly relevant if colors are not equally frequent. For example, in Aragona et al. (2017) the authors find prevalence of CFP, RFP, YFP clones at roughly 30% each and GFP at roughly 10%, which corresponds to a correction factor to the merging rate of roughly 0.72. For the labelling strategy used in Baggiolini et al. (2015); Kaucka et al. (2016), migratory neural crest cells induced at low recombination density express RFP at roughly 73%, YFP at 17%, CFP at 9% and GFP at less than 1%. This gives a correction factor of 0.43.

### 2.2.2 Single-color experiments

Calculating the rate of clonal mergers in experiments with a single fluorescent color is much more challenging. In this case, the merging rate must be inferred from statistical models that are based on simplifying assumptions about clonal induction events and the spatial location of clones. Fortunately, in many cases the assumptions that clonal induction events are statistically independent and spatially uniform and that clonal expansion is spatially isotropic are approximately valid. They therefore serve as a starting point for statistical frameworks of clonal merging.

An intuitive argument already gives an estimate of the merging probability. For this argument, let us consider a tissue with a number of labelled clones. Let us pick out a random clone and put it in a random location in the tissue. What is the probability that this clone is put onto a location that is covered by another clone? If we neglect the spatial extension of the randomly picked clone then this probability is equal to the fraction of the tissue occupied by labelled cells (Figure 2B). Therefore, to quantify the degree of clonal merging an

estimate can be obtained from a simple quantification of the area or volume fraction of labelled cells.

A similar argument was employed in Frede et al. (2016) who studied oesophageal tumour growth and observed a small number of clones that were much larger than the typical clone size. They reasoned that these clones might be a result of merging events and estimated the merging probability. Since the circumference of the tumour lesions they studied is effectively one-dimensional, the merging probability for a given clone size  $s$  is equal to the probability that the distance to the neighbouring clone is smaller than  $s$ . If  $p$  is the probability that a cell is induced this is equal to  $1 - (1 - p)^s$ . Taking into account that clone sizes are variable and are described by a distribution  $P(s)$ , one then sums over contributions from different clone sizes,

$$p_{merge} = 1 - \sum_{s=1}^{\infty} (1 - p)^s P(s). \tag{4}$$

This argument, which was formulated for one spatial dimension (the circumference of the lesion) can be generalised to planar tissues (two spatial dimensions) and volumar tissues (three spatial dimensions). Under the conditions of statistical independence and spatial uniformity of induction events and isotropy of clone expansion, the probability of clonal merging events can be estimated by noting that the strongest contribution to the merging probability comes from merging events with nearest neighbouring clones, while merging events between three or more clones are much less probable. Then, the probability of merging is equal to the probability that the distance between the centers of nearest neighbours is smaller than the sum of their radii. From this we obtain as a generalisation of Eq. 4,

$$p_{merge} = \int_0^{\infty} (1 - e^{-\rho V_d(r)}) C(r) dr. \tag{5}$$

where  $\rho = m_0/V$  is the density of clones in the tissue,  $V_d(r) = \pi^{\frac{d}{2}} r^d \Gamma(\frac{d}{2} + 1)$  is the volume of a sphere of radius  $r$  in  $d$  spatial dimensions, where  $\Gamma$  is the Gamma function and  $C(r)$  is the probability that the sum of two radii is equal to  $r$  which is derived from the clone size distribution. This means that, similar to Frede et al. (2016), the merging probability can in general be obtained from an estimate of the induction frequency,  $m_0$ , and the clone size distribution.

### 2.3 Application to multipotency

An important application of lineage tracing experiments is to establish lineage relationships between different cell types. The idea is that if an induced cell type gives rise to clones containing another type of cells then the latter is lineage related to the former. However, if induction is not entirely specific such that both types of cells are induced, then the chance of merging of

clones might lead to unguarded conclusions about multipotency. Therefore, in order to establish the hypothesis of multipotency one needs to devise a statistical test quantifying the significance of an enrichment of clones containing both cell types compared to clonal merging by chance.

Wuidart et al. (2016) devised a statistical framework to test for multipotency which they applied to mammary gland and prostate tissues, two ductal epithelial tissues where basal cells and luminal cells are in direct contact with one another (Figure 2C). If induction events are statistically independent the probability that a given pair of basal and luminal cells are chance mergers is binomial. This constitutes the null hypothesis. Therefore, a binomial test can be used to establish the statistical significance of the alternative hypothesis of multipotency. If  $k^*$  is the number of observed pairs of basal and luminal cells of the same color, the  $p$ -value is

$$p = \sum_{k=k^*}^N \binom{N}{k} \nu^k (1 - \nu)^{N-k}. \tag{6}$$

The value of the sample size  $N$  and the parameter  $\nu$  depend on the statistical model used to describe the null hypothesis.

Based on this idea, Wuidart et al. derived two orthogonal statistical tests. In the first test, they considered only pairs of labelled basal and luminal cells. The null hypothesis then is that among these pairs the ones that are of the same color arise by chance labelling events. In this case,  $\nu$  is the probability that a given pair of labelled basal and luminal cells is of the same color. This probability is equal to

$$\nu = \sum_C r_C^B r_C^L, \tag{7}$$

where  $r_C^A$  and  $r_C^B$  the relative probabilities of labelling luminal and basal cells in color  $c$  respectively. As an example, in an experiment using the *Confetti* construct on two cell types A and B, the majority colors RFP, YFP and CFP are approximately induced at the same probability  $r_C^A = 1/3$  and  $r_C^B = 1/3$  for all colors and consequently  $\nu = 1/3$ . Then, if 50 out of 100 pairs are unicolor,  $k = 50$  and  $N = 100$  such that the  $p$ -value, Eq. 6, is  $p = 0.0004$ .

This example demonstrates that although this test is insensitive to assumptions about the tissue architecture and overall labelling frequencies, it requires a large number of unicolor pairs in order to establish statistical significance. If induction frequencies are low these pairs are rare such that a large number of tissues potentially needs to be analysed.

For a broader statistical test, Wuidart et al. used the fraction of basal cells in contact with a luminal cell of any color. Under the null hypothesis of chance merging, this probability depends on the geometry of the tissue, which is encoded in the coordination number  $z$  giving the number of luminal cells in contact with one basal cell. As the coordination number may differ between



individual basal cells, we define the distribution of coordination numbers,  $\mathcal{F}(z)$ , which can be directly obtained experimentally from microscopy images. The probability that an induced basal cell is paired to a labelled luminal cell then takes the rather lengthy form

$$\nu = \left( \sum_C \lambda_C \right) \sum_{z=1}^{\infty} \left[ z \left( 1 - \sum_{C'} \lambda_{C'} \right)^{z-1} \mathcal{F}(z) \right], \quad (8)$$

which now depends on the luminal induction probability,  $\lambda_C$ , which should be determined independently. If variations in coordination number are small we can approximate the pairing probability as

$$\nu \approx \bar{z} \sum_C \lambda_C, \quad (9)$$

where  $\bar{z}$  denotes the average coordination number. Therefore, in this case the pairing probability depends only on the geometry of the tissue and the specificity of the genetic labelling construct. Substituting this into Eq. 6 gives another binomial test for multipotency where  $k^*$  now is the number of labelled basal cells in contact with a labelled luminal cell and  $N$  is the total number of labelled basal cells. A similar approach was also followed in Lilja et al. (2018) in order to quantify the degree of bipotency in the mammary gland during different time points in development.

These tests are not restricted to tissues containing basal and luminal cells, but can be applied and generalised to any tissue where cell layers are in direct contact with each other. Since both tests rely on complementary aspects of the clonal data it is advisable to strengthen one's conclusion about multipotency by performing both of them. Should these tests indicate unipotency, then a direct comparison between the predicted values of Eqs 7–9 with experimental data could further strengthen this conclusion. In the case that the analysis supports multipotency, these comparisons can allow inferring the rate of differentiation between both cell types.

### 2.3.1 Clonal merging in barcoding experiments

In recent years, fluorescent-marker-based assays have been supplemented with labelling strategies using nucleic-acid-based barcoding (Kebschull and Zador (2018); Wagner and Klein (2020)). Due to the large number of induced cells, these approaches require the creation of a barcode library with sufficient diversity to induce individual cells with unique barcodes, as well as a sequencing-based method to read out the barcodes. *In vivo* barcoding methods generally fall into one of two classes: 1) methods based on recombinase enzymes, which flip or excise DNA sequences to generate sequence diversity and 2) CRISPR-based methods where an endonuclease enzyme such as Cas9 introduces insertions or deletions into a genomic region of the DNA. Readout of the barcodes can be performed by PCR amplification followed by bulk sequencing, through single-cell

RNA sequencing, through fluorescence *in situ* hybridization (FISH) methods or even through spatial transcriptomics (Rao et al. (2021)).

While typically a large number of different barcodes is generated, not all of them are induced with the same probability. This gives rise to the chance induction of several cells with the same barcode and hence clonal merging. The non-spatial methods discussed in Sections 2.1–2.3 can be applied to barcoding experiments, where different barcodes take the role of fluorescent colors. Alternative approaches rely on modelling the molecular processes leading to the generation of barcodes in order to estimate the induction frequency of individual barcodes (Marcou et al. (2018)).

## 2.4 Estimating the fragmentation rate

Estimating the rate of clonal fragmentation is not only important to understand the degree of clonality of a given experiment, but can also be informative for understanding the underlying processes that cause fragmentation (*cf.* Section 3.3). In order to estimate the rate of fragmentation, Lescroart et al. (2014) derived the probability distribution  $F(k)$  of the number of fragments  $k$  in a given sample. Since this distribution depends both on the average number of fragmentation events between labelling and analysis,  $f$ , and the number of induced cells,  $m$ , it in principle allows inferring both simultaneously from clonal fate data. To this end, they defined the joint probability of inducing  $m$  cells and observing  $k$  fragments,  $J(k, m)$ . This joint probability is equal to the conditional probability of observing  $k$  fragments given  $m$  induced cells multiplied with the marginal probability of inducing  $m$  cells. Since individual induction and fragmentation events are considered to be statistically independent, both follow Poisson distributions, such that

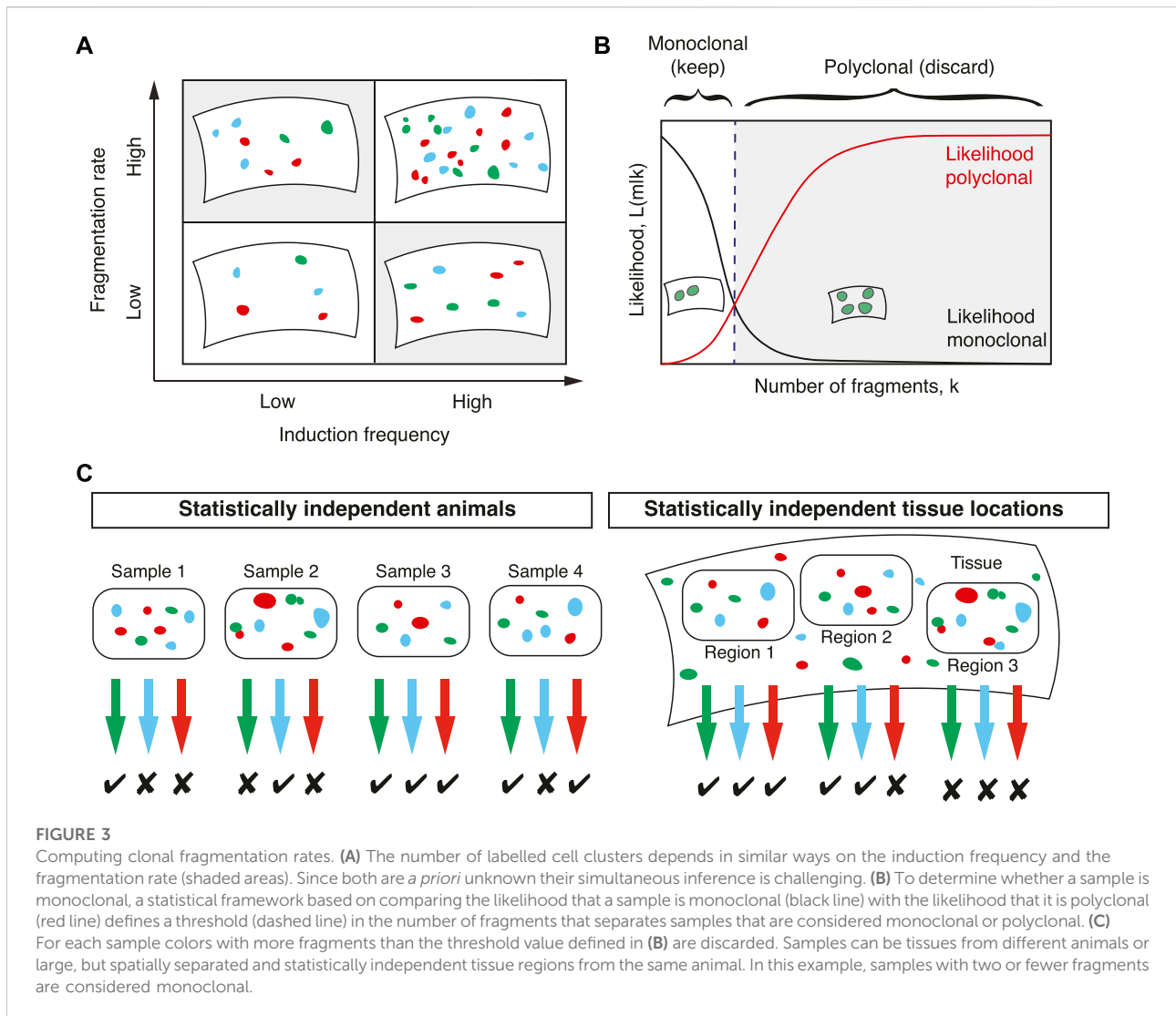
$$J(k, m) = \frac{(mf)^{n-m} m_0^m}{(k-m)! m!} e^{-mf-m_0}. \quad (10)$$

Because completely unlabelled organs are often not quantified, this needs to be divided by the probability that a tissue is unlabelled,  $J(0,0)$ . Taken together, the number of fragments then follows

$$F(k) = \frac{\sum_{m=1}^k J(m, k)}{J(0,0)}. \quad (11)$$

With this, the fragmentation rate  $f$  and the induction frequency  $m_0$  can be estimated from the data, for example by maximum likelihood estimation.

Because two parameters with similar effect on the distribution  $F(k)$  need to be estimated simultaneously (Figure 3A), it is expected that with the sizes of typical lineage tracing experiments the uncertainties associated with these estimates are large. Therefore, in the case of multi-color



labelling, the induction frequency should be estimated independently as in Section 2.1. Then, in a second step, the fragmentation rate  $f$  can be inferred more robustly.

### 3 Interpreting nonclonal experiments

Given that merging or fragmentation of clones confound a lineage tracing experiment, how can information about cell fate behaviour still be inferred? In the following, we will discuss three approaches that solve this problem in different ways: by discarding nonclonal experimental samples, by restoring the clonal origin of labelled cells and by taking clonal fragmentation as an opportunity to learn about the mechanical effects of cell fate decisions.

### 3.1 Filtering for monoclonal samples

In the first approach, samples with fragments deriving from multiple induction events are filtered out and the downstream analysis is restricted to monoclonal samples (Figure 3B). This approach was followed in Lescroart et al. who devised a statistical framework that allows, with known uncertainty, to filter for monoclonal samples. The decision of whether a sample is monoclonal is quantified by the ratio of two probabilities: the probability that a given number of fragments are polyclonal and the probability that a given number of fragments are monoclonal. Following Bayes' theorem, this ratio is equal to the ratio of two likelihood functions: if  $L(k|m)$  should be  $L(k|m)$  is the probability that an observed number of  $k$  fragments arises from  $m$  labelled cells then this ratio is



$$\Lambda = \frac{\sum_{m=2}^{\infty} L(m|k)}{L(m=1|k)}. \quad (12)$$

By defining a threshold value of  $\Lambda$  this defines a number of fragments in a given color below which a given color in a given sample is considered monoclonal (Figure 3B). If  $\Lambda \geq 1$  a given sample is more likely to be polyclonal than monoclonal and these samples were discarded in Lescroart et al. (2014). To achieve a given higher significance level (a stricter criterion for monoclonality) a smaller number than one can be chosen as a threshold at the cost of a larger number of discarded data points. With this threshold defined, the experimental data can then be filtered for samples containing fewer fragments than this threshold (Figure 3C). All downstream analysis then only considers samples that are monoclonal with defined uncertainty. Lescroart et al. (2014) used hearts from different animals and different fluorescent colours to define such samples. An alternative approach is, as mentioned in Section 2.1, to define regions in the tissue that are much larger than the typical expected size of dispersed clones and sufficiently distant to be considered statistically independent (Figure 3C).

### 3.2 Restoring clonality

If the rates of merging and fragmentation are small, clonality can potentially be restored statistically by making use of the fact that labelled cells belonging to the same clone have the tendency to be located closer to each other compared to cells originating from different induction events. As a first approach one might therefore employ standard clustering techniques, such as hierarchical clustering or k-means clustering on the spatial coordinates of labelled cells. These approaches, however, have shortcomings in the context of lineage tracing experiments: First, they often lack unambiguous ways of determining the total number of induced clones, which then becomes an independent parameter that needs to be determined using different tools (cf. Section 2.1). Secondly, they lack ways for quantifying levels of uncertainty associated with the assignment of cells to clones. Importantly, these uncertainties need to be propagated to downstream analyses, such as the calculation of average clone sizes. Transparency about uncertainties also allows one to be tolerant with respect to mis-assignments in clonal identity. Thirdly, standard clustering algorithms usually do not allow to take into account additional information experimentalists might have on a specific biological context, such as on the rates of cell migration.

Than-Trong et al. (2020) studied cell fate behaviour of neural stem cells in the Zebrafish brain. Because labelled clones dispersed over time, they devised a statistical framework that overcomes these limitations by combining a biophysical model of clone dispersion with Bayesian inference. The key idea is to calculate the probability that a given partition of labelled cells into clones is the true clonal partition (posterior). According to Bayes' theorem, this probability is

proportional to the probability that a given clonal partition arises from statistically independent induction events (termed likelihood) multiplied with the *a priori* probability that a given partition arises in the first place (termed prior). The likelihood was modelled as a stochastic process of clonal induction and dispersion while the prior contains information about clonal induction frequencies or clone dispersion estimated independently from the data. Then, the partition of cells into clones that maximized the posterior, and its uncertainty, was approximately obtained by successively coarse graining cells in larger and larger clones.

In a more simplistic way, experimentalists quantifying clonal fate data could also decide on the clonality of a group of fragments based on the distance between these fragments. In order to specify a distance  $w^*$  below which a pair of fragments can be considered clonal we consider a pair of fragments and assume that these fragments are nearest neighbors. Then, if we aim for a given fraction  $\alpha$  of polyclonal cell clusters at a distance shorter than  $w^*$  (significance level), we get as a condition

$$\int_0^{w^*} P(m=2|w)P(w)dw = \alpha, \quad (13)$$

where  $P(m=2|w)$  is the probability that two fragments at a given distance  $w$  arise from two separate induction events and  $P(w)$  is the probability to find any two fragments at distance  $w$ . According to Bayes' theorem this is equal to  $\int_0^{w^*} P(w|m=2)P(m=2)dw$ , where  $P(w|m=2)$  is the probability that a pair of monoclonal fragments have a distance  $w$ . If clone induction events are statistically independent then  $P(w|m=2) = e^{-\rho\pi^{d/2}w^d/\Gamma(d/2+1)}$  (Diggle (2013)), where  $\rho = m_0/V$  is the density of clones,  $V$  is the area or volumes analysed, and  $\Gamma$  is the Gamma function.  $P(m=2)$  is the *a priori* probability that two clusters of labelled cells at any distance arise from two separate induction events. If the fragmentation rate is known this is equal to the probability that an induced clone has fragmented,  $e^f$ . Taken together, we can solve for  $w^*$  and obtain

$$w^* = \left[ \frac{\Gamma(d/2+1)(f-1)V}{\pi^{d/2}k \ln(1-\alpha e^{-f})} \right]^{1/d}. \quad (14)$$

As expected, the distance threshold  $w^*$  decreases with the number of fragmentation events between induction and analysis,  $f$ , and with the overall number of cell clusters,  $k$ .  $w^*$  is the distance below which two labelled cell clusters are clonal with significance level  $\alpha$ . A reasonable choice could be  $\alpha = 0.5$  such that at distances shorter than  $w^*$  two fragments are more likely clonal than polyclonal.

### 3.3 Learning from clonal fragmentation about cell fate

Instead of treating clonal fragmentation as a liability for the analysis of clonal fate data, in recent years several groups have

TABLE 1 Notations used throughout this review.

$m_0$	Parameter describing the average number of induced cells in a given experiment
$m$	Random variable describing the actual number of induced cells in a given sample
$k$	Random number describing the number of observed cell clusters in a given sample
$r_c$	Probability that a given labelled cell has color $c$
$z$	Coordination number giving the number of neighbours a cell has
$n_c$	Total number of colors expressed in a lineage tracing experiment
$f$	Average number of fragmentation events per clone between induction and analysis
$p_{merge}$	Probability that a given cluster of labelled cells is polyclonal

taken an orthogonal perspective. They noted that the dispersion of clones after labelling reflects a history of mechanical forces acting on the cells in the clone and that an analysis of clonal fragmentation can give insights into these processes. [Watson et al. \(2020\)](#) used a zebrafish line that fluorescently labels osteoblasts and then used CRISPR to induce mutations of *Plod2* and *Bmp1a* in a subset of cells. The ensuing clone size distribution takes the universal shape as predicted to arise from fragmentation and merger events ([Rulands et al. \(2018\)](#)). Therefore, clone size distributions were not informative about the phenotype. However, by using statistical measures for the spatial distribution of labelled cells they could identify systematic differences between wildtype and mutant samples. One of these measures is termed Moran's I which is a statistical measure of spatial correlations. The authors further used multivariate statistical tests to compare the spatial fluorescence signal between conditions and performed Monte Carlo simulations. This work exemplifies that while clone size distributions may converge to universal shapes that are independent of the biological condition, quantities capturing the spatial statistics of labelled cells may carry cell fate specific information.

Further works associated clonal fragmentation with mechanical processes in tissues. [Ramanathan et al. \(2019\)](#) used models from statistical mechanics to show that clone dispersion is inversely proportional to cell size, such that smaller cells show higher clonal dispersion. In these models, termed vertex models, cells are considered as polygons with vertices and edges representing the boundaries between cells [Farhadifar et al. \(2007\)](#). The location of vertices and edges then follows equations of motion determined by different forces acting on vertices. Numerical results obtained from these models were also compared to experimental findings from the *Drosophila* wing disk of mutant embryos with smaller cells.

[Bocanegra-Moreno et al. \(2022\)](#) took a more direct approach of inferring mechanical and material properties from clonal fragmentation. They performed lineage tracing using the MADM system ([Zong et al. \(2005\)](#)) in the mouse neural tube. Based on these experiments they used

information on the fragmentation of clones to infer mechanical and material properties of tissues, and to learn about cellular processes that control fragmentation. Specifically, in order to compare patterns of clone fragmentation with theoretical predictions, they implemented a vertex model parametrised by normalised tension and normalised contractility, and applied noise to the normalised tension. In their simulations they labelled individual cells and tracked how the progeny of these cells dispersed in the tissue. While clonal fragmentation alone was not sufficient to determine the position of the tissue in parameter space, the combination of fragmentation, the rate of T1 transitions - where cells rearrange and change neighbours - and geometric parameters describing cell shapes allowed them to identify tissue and material properties of the neural tube. They also found *in vivo* and *in silico* that the fragmentation rate increases with the rate of cell divisions, while cell differentiation leads to a reduction in clone fragmentation.

[Kaucka et al. \(2016\)](#) combined lineage tracing, computer simulations, live imaging and mouse mutants to study clonal dynamics and cell movements of the cranial neural crest cells that give rise to the ectomesenchyme. They found that clones originating from a single neural crest cell form clonal envelopes, well-defined clusters of labelled cells that remain stable over time. However, these clusters overlap with one another, such that there is considerable mixing between cells from different clones. Computational modelling suggested that pushing of cells arising from cell division rather than cell migration is responsible for generating this clonal mixing. This is confirmed through live imaging, which revealed crowd motion of these clones with limited individual cell migration. Furthermore, they also find evidence of oriented cell divisions, which are consistent with a computational model that includes a signalling gradient, which in turn would explain the elongated shapes of clones they observed. Taken together, these works show that merging and fragmentation processes in themselves carry valuable information about the mechanical processes that underlie tissue morphogenesis.

## 4 Discussion

The interpretation of genetic tracing experiments relies on the integrity of the clonal relationship between induced cells and their progeny at the time point of analysis. Large-scale cell rearrangements or stochastic forces in the tissue lead to the fragmentation and merging of clones, which renders the retrospective analysis of such experiments highly ambiguous. This is on the one hand because the resulting size distributions obtain universal shapes that are largely independent of cell fate behaviour. On the other hand, since the induction frequency and the merging and fragmentation rates are *a priori* unknown, restoring the clonal origin of induced cells is statistically challenging.

If such non-clonality is undetected it might lead to unguarded claims about cell fate behavior. In this review, we discussed statistical strategies to detect non-clonality in genetic tracing experiments. Genetic constructs where clones can be induced in multiple colors provide simple and elegant routes to estimating the induction frequency by comparing the fractions of samples induced in different numbers of colours. They also allow for estimating the merging rate by counting the number of mergers between cell clusters of different colours. For experiments using only a single color, estimates of merging and fragmentation rates are feasible using stochastic models of clonal induction and *a priori* knowledge about the clone size distribution. Given that estimating the merging and fragmentation rates is feasible with no additional experimental effort, we suggest that such estimates should be conducted as a safety check in the analysis of clonal fate data. With the advent of single-cell genomics and deep sequencing, strategies that label clones with fluorescent colors have been complemented with nucleic-acid based barcoding. Since these barcodes are induced at different, not necessarily clonal frequencies, multiple cells can be labelled with the same barcode. From a statistical perspective this again leads to the effective merging of clones. By treating barcodes as different colours the statistical tools presented here to detect merging in multi-color experiments can be used to estimate the degree of clonality in barcoding experiments.

If clonal merging or fragmentation has been detected, several strategies exist for interpreting nonclonal data. A first strategy relies on filtering for monoclonal fragments. This approach is applicable even if the fragmentation rate is high, but it comes with the downside that a potentially large fraction of the data is not used for further analysis. If fragmentation and merging are infrequent, such that the number of fragmentation events is much smaller than the number of cell divisions in a given time interval, the clonal provenance of labelled cell clusters can be faithfully restored by statistical inference or through setting simple thresholds

in the distances between labelled cells. While these approaches cannot restore clonal information that is ultimately lost due to strong dispersion, they do come with well-defined statistical uncertainties in the assignment of cells to clones. As long as these uncertainties associated with the assignment of cells to clones are transparently propagated, occasional mis-assignments of the clonal origin of labelled cells do not compromise the interpretability of downstream results.

While clonal fragmentation and merging renders the *sizes* of clones ambiguous, valuable information about cell fate might be contained in other statistical quantities of labelled cell clusters. First, the rates of merging and fragmentation and their respective changes over time are associated with the rates of cell proliferation and cell rearrangements in the tissue. Further, the *shapes* of clones reflect anisotropic forces which are relevant for understanding morphogenesis. Such forces may arise through oriented cell divisions (Economou et al. (2013)), but also through collective motion leading to morphogenetic flows. In order to quantitatively interpret these features and to deduce mechanisms of tissue morphogenesis, models describing the collective effects of mechanical forces in tissues need to integrate clonal dynamics. Here, a promising line of research will be the integration of the shapes and boundary roughness of labelled cell clusters into quantitative frameworks of tissue mechanics.

## Author contributions

YD and SR wrote the manuscript, performed calculations and numerical simulations.

## Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 950349).

## Acknowledgments

We thank P. Greulich and B.D. Simons for helpful discussions.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alcolea, M. P., Greulich, P., Wabik, A., Frede, J., Simons, B. D., and Jones, P. H. (2014). Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nat. Cell Biol.* 16, 615–622. doi:10.1038/ncb2963
- Aragona, M., Dekoninck, S., Rulands, S., Lenglez, S., Mascré, G., Simons, B. D., et al. (2017). Defining stem cell dynamics and migration during wound healing in mouse skin epidermis. *Nat. Commun.* 8, 14684. doi:10.1038/ncomms14684
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556
- Baggiolini, A., Varum, S., Mateos, J., Bettosini, D., John, N., Bonalli, M., et al. (2015). Premigratory and migratory neural crest cells are multipotent *in vivo*. *Cell Stem Cell* 16, 314–322. doi:10.1016/j.stem.2015.02.017
- Blanpain, C., and Simons, B. D. (2013). Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* 14, 489–502. doi:10.1038/nrm3625
- Bocanegra-Moreno, L., Singh, A., Hannezo, E., Zagorski, M., and Kicheva, A. (2022). Cell cycle dynamics controls fluidity of the developing mouse neuroepithelium. *bioRxiv* 2022, 477048. doi:10.1101/2022.01.20.477048
- Corominas-Murtra, B., Scheele, C. L. G. J., Kishi, K., Ellenbroek, S. I. J., Simons, B. D., van Rheeën, J., et al. (2020). Stem cell lineage survival as a noisy competition for niche access. *Proc. Natl. Acad. Sci. U. S. A.* 117, 16969–16975. doi:10.1073/pnas.1921205117
- Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. 3 edn. Boca Raton, Florida, USA: Chapman & Hall/CRC Monographs on Statistics & Applied Probability CRC Press.
- Doupé, D. P., Klein, A. M., Simons, B. D., and Jones, P. H. (2010). The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Dev. Cell* 18, 317–323. doi:10.1016/j.devcel.2009.12.016
- Driessens, G., Beck, B., Caauwe, A., Simons, B. D., and Blanpain, C. (2012). Defining the mode of tumour growth by clonal analysis. *Nature* 488, 527–530. doi:10.1038/nature11344
- Economou, A. D., Brock, L. J., Cobourne, M. T., and Green, J. B. A. (2013). Whole population cell analysis of a landmark-rich mammalian epithelium reveals multiple elongation mechanisms. *Development* 140, 4740–4750. doi:10.1242/dev.096545
- Farhadifar, R., Röper, J.-C., Aigouy, B., Eaton, S., and Jülicher, F. (2007). The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Curr. Biol.* 17, 2095–2104. doi:10.1016/j.cub.2007.11.049
- Frede, J., Greulich, P., Nagy, T., Simons, B. D., and Jones, P. H. (2016). A single dividing cell population with imbalanced fate drives oesophageal tumour growth. *Nat. Cell Biol.* 18, 967–978. doi:10.1038/ncb3400
- Greulich, P., MacArthur, B. D., Parigini, C., and Sánchez-García, R. J. (2021). Universal principles of lineage architecture and stem cell identity in renewing tissues. *Development* 148, dev194399. doi:10.1242/dev.194399
- Kauka, M., Ivashkin, E., Gyllborg, D., Zikmund, T., Tesarova, M., Kaiser, J., et al. (2016). Analysis of neural crest-derived clones reveals novel aspects of facial development. *Sci. Adv.* 2, e1600060. doi:10.1126/sciadv.1600060
- Kebeschull, J. M., and Zador, A. M. (2018). Cellular barcoding: Lineage tracing, screening and beyond. *Nat. Methods* 15, 871–879. doi:10.1038/s41592-018-0185-x
- Klein, A. M., Doupé, D. P., Jones, P. H., and Simons, B. D. (2007). Kinetics of cell division in epidermal maintenance. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76, 021910. doi:10.1103/PhysRevE.76.021910
- Klein, A. M., and Simons, B. D. (2011). Universal patterns of stem cell fate in cycling adult tissues. *Development* 138, 3103–3111. doi:10.1242/dev.060103
- Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolose, A., et al. (2014). Early lineage restriction in temporally distinct populations of mesp1 progenitors during mammalian heart development. *Nat. Cell Biol.* 16, 829–840. doi:10.1038/ncb3024
- Lilja, A. M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., et al. (2018). Clonal analysis of notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nat. Cell Biol.* 20, 677–687. doi:10.1038/s41556-018-0108-1
- Marcou, Q., Mora, T., and Walczak, A. M. (2018). High-throughput immune repertoire analysis with igor. *Nat. Commun.* 9, 561. doi:10.1038/s41467-018-02832-w
- Ramanathan, S. P., Krajnc, M., and Gibson, M. C. (2019). Cell-size pleomorphism drives aberrant clone dispersal in proliferating epithelia. *Dev. Cell* 51, 49–61. e4. doi:10.1016/j.devcel.2019.08.005
- Ranfjt, J., Basan, M., Elgeti, J., Joanny, J.-F., Prost, J., and Jülicher, F. (2010). Fluidization of tissues by cell division and apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* 107, 20863–20868. doi:10.1073/pnas.1011086107
- Rao, A., Barkley, D., França, G. S., and Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. *Nature* 596, 211–220. doi:10.1038/s41586-021-03634-9
- Robertson, N. A., Latorre-Crespo, E., Terradas-Terradas, M., Lemos-Portela, J., Purcell, A. C., Livesey, B. J., et al. (2022). Longitudinal dynamics of clonal hematopoiesis identifies gene-specific fitness effects. *Nat. Med.* 28, 1439–1446. doi:10.1038/s41591-022-01883-3
- Rulands, S., Lescroart, F., Chabab, S., Hindley, C. J., Prior, N., Sznurkowska, M. K., et al. (2018). Universality of clone dynamics during tissue development. *Nat. Phys.* 14, 469–474. doi:10.1038/s41567-018-0055-6
- Rulands, S., and Simons, B. D. (2016). Tracing cellular dynamics in tissue development, maintenance and disease. *Curr. Opin. Cell Biol.* 43, 38–45. doi:10.1016/j.ccb.2016.07.001
- Ruske, L. J., Kursawe, J., Tsakiridis, A., Wilson, V., Fletcher, A. G., Blythe, R. A., et al. (2020). Coupled differentiation and division of embryonic stem cells inferred from clonal snapshots. *Phys. Biol.* 17, 065009. doi:10.1088/1478-3975/aba041
- Scheele, C. L. G. J., Hannezo, E., Muraro, M. J., Zomer, A., Langedijk, N. S. M., van Oudenaarden, A., et al. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. *Nature* 542, 313–317. doi:10.1038/nature21046
- Shakiba, N., Fahmy, A., Jayakumar, G., McGibbon, S., David, L., Trcka, D., et al. (2019). Cell competition during reprogramming gives rise to dominant clones. *Science* 364, eaan0925. doi:10.1126/science.aan0925
- Simons, B. D., and Clevers, H. (2011). Stem cell self-renewal in intestinal crypt. *Exp. Cell Res.* 317, 2719–2724. doi:10.1016/j.yexcr.2011.07.010
- Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., et al. (2010). 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
- Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272. doi:10.1038/s41576-019-0093-7
- Tam, P. P. L., and Ho, J. W. K. (2020). Cellular diversity and lineage trajectory: Insights from mouse single cell transcriptomes. *Development* 147, dev179788. doi:10.1242/dev.179788
- Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. doi:10.1038/nature21350
- Than-Trong, E., Kiani, B., Dray, N., Ortica, S., Simons, B., Rulands, S., et al. (2020). Lineage hierarchies and stochasticity ensure the long-term maintenance of adult neural stem cells. *Sci. Adv.* 6, eaaz5424. doi:10.1126/sciadv.aaz5424
- Wagner, D. E., and Klein, A. M. (2020). Lineage tracing meets single-cell omics: Opportunities and challenges. *Nat. Rev. Genet.* 21, 410–427. doi:10.1038/s41576-020-0223-2
- Watson, C. J., Monstad-Rios, A. T., Bhimani, R. M., Gistelinc, C., Willaert, A., Coucke, P., et al. (2020). Phenomics-based quantification of crispr-induced mosaicism in zebrafish. *Cell Syst.* 10, 275–286. e5. doi:10.1016/j.cels.2020.02.007
- Wuidart, A., Ousset, M., Rulands, S., Simons, B. D., Van Keymeulen, A., and Blanpain, C. (2016). Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev.* 30, 1261–1277. doi:10.1101/gad.280057.116
- Yamaguchi, H., Kawaguchi, K., and Sagawa, T. (2017). Dynamical crossover in a stochastic model of cell fate decision. *Phys. Rev. E* 96, 012401. doi:10.1103/PhysRevE.96.012401
- Yum, M. K., Han, S., Fink, J., Wu, S.-H. S., Dabrowska, C., Trendafilova, T., et al. (2021). Tracing oncogene-driven remodelling of the intestinal stem cell niche. *Nature* 594, 442–447. doi:10.1038/s41586-021-03605-0
- Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D., and Luo, L. (2005). Mosaic analysis with double markers in mice. *Cell* 121, 479–492. doi:10.1016/j.cell.2005.02.012