eLife

Reviewed Preprint
v1 • March 17, 2025
Not revised

For correspondence:
mattia.miotto@roma1.infn.it

Competing interests: The authors declare no competing interests

Reviewing editor: Jan Brugués, Technische Universität Dresden, Germany

Robust assessment of asymmetric division in colon cancer cells

Domenico Caudo[1,2], Chiara Giannattasio[3,1], Simone Scalise[1,2], Valeria de Turris[1], Fabio Giavazzi[3], Giancarlo Ruocco[1,2], Giorgio Gosti[4], Giovanna Peruzzi[1], Mattia Miotto[1]

[1]Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Rome, Italy • [2]Department of Physics, Sapienza University, Rome, Italy • [3]Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Italy • [4]Istituto di Scienze del Patrimonio Culturale, Consiglio Nazionale delle Ricerche, Rome, Italy

eLife Assessment

This study presents a **useful** method based on flow cytometry to study partitioning noise during cell division. The evidence supporting the claims of the authors is **incomplete**, as the method neglects other sources of noise present in cells. With the theoretical part extended, this paper would be of interest to cell biologists and biophysicists working on asymmetric partitioning during cell division.

https://doi.org/10.7554/eLife.104528.1.sa3

Abstract

Asymmetric partition of fate determinants during cell division is a hallmark of cell differentiation. Recent works suggested that such a mechanism is hijacked by cancer cells to increase both their phenotypic heterogeneity and plasticity and in turn their fitness. To quantify fluctuations in the partitioning of cellular elements, imaging-based approaches are used, whose accuracy is limited by the difficulty of detecting cell divisions. Our work addresses this gap proposing a general method based on high-throughput flow cytometry measurements coupled with a theoretical frameworks. We applied our method to a panel of both normal and cancerous human colon cells, showing that different kinds of colon adenocarcinoma cells display very distinct extents of fluctuations in their cytoplasm partition, explained by an asymmetric division of their size. To test the accuracy of our population-level protocol, we directly measure the inherited fractions of cellular elements from extensive time-lapses of live-cell laser scanning microscopy, finding excellent agreement across the cell types. Ultimately, our flow cytometry-based method promise to be accurate and easily applicable to a wide range of biological systems where the quantification of partition fluctuations would help accounting for the observed phenotypic heterogeneity and plasticity.

I. Introduction

Asymmetric cell division is referred to as the mechanism by which a mother cell splits into two daughter cells with distinct cellular fates. As first shown by Edwin Conklin [1], this process is usually achieved by asymmetric inheritance of intrinsic cell fate determinants, like specific proteins or RNA, and plays a crucial role during the organisms' development, favoring cell differentiation and self-renewal [2, 3].

Increasing evidence is vouching for a more ubiquitous presence of fluctuations in the partitioning of cellular elements. Indeed, asymmetric segregation has been observed in non-differentiating cell populations as different as bacteria [4], yeast [5], and tumoral cells [6, 7]. In bacterial populations, differential segregation of cellular components has been linked to antibiotic resistance [8]. Similarly, Katajisto *et al.* [9], studying the partition of mitochondria in mammalian epithelial stem-like, found that asymmetric segregation of older mitochondria enables cells to protect themselves

from aging by preventing the accumulation of misfolded proteins. Asymmetric partition of centrosomes has been observed in human neuroblastoma and colorectal cancer stem cells, controlled by polarity factors and influenced by the segregation of subcellular vesicles during cell division [10].

To quantify the strength of fluctuations at division, the common route goes through fluorescence microscopy measurements of either fixed or live cells, where fluorescent dyes are used as a proxy for the cellular element of interest. The partition statistics are then estimated by looking at the fraction of fluorescence intensity inherited by the two daughter cells [11, 12]. This approach is, in principle, much informative; however, it needs the (often manual) identification of hundreds of division events to be reliable: a time-consuming procedure that dampens a wide characterization of the partition statistics of different cell types.

To cope with these limitations, we propose using high-throughput flow cytometry measurements to quantify the fluctuations in the partitioning of cellular components. In particular, flow cytometry has already been applied to study asymmetric division in specific cellular systems: Yang *et al.* [5], sorted budding yeast cells to analyze the differential segregation of proteins between daughters cells, while Peruzzi and coworkers first applied multi-color flow cytometry to show that leukemia cells divide mitochondria and membrane proteins asymmetrically [13]. Here, we generalized the protocol to adherent cell types and probed its accuracy by comparing the measured fluctuations to those observed with extensive time-lapse microscopy experiments on a panel of normal and cancerous epithelial cells.

Our work shows that (i) there is an exact, analytical expression linking the inherited cellular components distribution dynamic to the specifics of the partition process. Based on this result, we (ii) proposed an experimental protocol to tag with fluorescent dyes cellular components and measure the dynamics of fluorescent distribution, whose analysis allows for the accurate estimates the degree of asymmetry in the partition process. Applying our method to a panel of both normal and cancerous human colon cells, we found that (iii) different lines of colon adenocarcinoma display very distinct extents of fluctuations in the cytoplasm partition, that reflect in (iv) different degrees of asymmetric division of their size.

II. Results

A. Modeling cells population dynamic in presence of partitioning noise

We look for a theoretical framework able to describe the time evolution of a cell population and its dependence on the partitioning process. In particular, we consider a population of cells characterized by a certain initial distribution, $P(m)_0$, of element m. Monitoring the time evolution of the abundance of specific cellular elements over multiple divisions in single-cell experiments, it is observed [14] that a component's counts presents dynamics qualitatively similar to the one reported in Figure 1(b). Indeed, variations in the components number are determined by: (i) the production and degradation processes along the duration of the cell growth phase; and (ii) non identical apportioning of the components between the two daughter cells. Here, we will assume that the components number varies only due to the partition events and look for a quantitative description of the evolution of the element distribution. Such an assumption allows us to neglect the time between divisions and reduces the dynamics to a progressive dilution of the components numbers in subsequent generations. Note that it is possible to model the complete growth and division process (see for instance refs. [15–17]), but it is more difficult to separate the different sources of noise.

Each cell i of the population divides into two daughter cells that we can label unambiguously as $2i$ and $2i + 1$, which inherit respectively m_{2i} and m_{2i+1} components. The process, exactly at division, must conserve the total number of elements, i.e. $m_i = m_{2i} + m_{2i+1}$ and we call partitioning fraction the quantity $f = \frac{m_{2i}}{m_i}$. To model noise in partitioning, we assume that at each division a random

value for f is extracted from a probability distribution function, $\Pi(f)$. Given these assumptions, we can write the probability distribution of the components number of a daughter cell m after g divisions from the mother M as:

$$P_g(m) = \left(\frac{1}{2^{g-1}}\right) \sum_{k=0}^{2^{g-1}} \int dM P(m|M) P_k(M) \tag{1}$$

where $P_k(M)$ is the components distribution at generation k and it is given by the sum of the distributions of all siblings cells of the k-th generation. After g generations, the cell lineage generated from cell i forms a lineage tree composed of $n_g = 2^g$ cells. Without losing generality, the division probability can be expressed, as:

$$P(m|M) = \int \delta(m - fM) \prod(f)\, df. \tag{2}$$

Therefore, we can write the probability distribution, $P(m)$, of the daughter cells inheriting a fraction, f, of the mother elements:

$$P(m) = \int df\, dM \delta(m - fM) \prod(f) P(M) \tag{3}$$

As is often the case, it is convenient to characterize the distributions in terms of their moments [16]. We identify the cell m_{2i} (m_{2i+1}) as the one inheriting a fraction f ($1 - f$). In particular, the mean of the daughter cell $2i$ components distribution can be expressed as:

$$\begin{aligned} \mu_{2i} &= E[m_{2i}] = \int P(m_{2i})\, m_{2i}\, dm_{2i} \\ &= \int df\, dm_i\, dm_{2i}\, m_{2i} \delta(m_{2i} - f m_i) \prod(f) P(m_i) \\ &= E[f] E[m_i] = \mu_f \mu_i \end{aligned} \tag{4}$$

Where $E[\cdot]$ stands for the average. Note that every computation can be symmetrically done for the cell $2i + 1$. Thus, we have that in a single partition, the mean number of inherited elements of the daughter cells depends on the asymmetry of the $\Pi(f)$ and on the mean of the $P(m_i)$.

In Figure 1(d), we show the evolution of the components number distribution at different generations. The $\Pi(f)$ chosen for the simulation is a Cauchy distribution with location $x_0 = \frac{1}{2}$ and scale $\gamma = 0.7$.

The total distribution is the sum of 2^g sub-populations which derives from the division process. For example, if we assume $\prod(f) = \frac{1}{2}(\delta(f) + \delta(1 - f))$, after one division we would have a population distributed around $f\mu_i$ and one at $(1-f)\mu_i$. After two division the populations would grow to four sub-populations centered in $f^2\mu_i$, $f(1-f)\mu_i$, $(1-f)f\mu_i$ and $(1-f)^2\mu_i$ and so on.

To compute the mean relative to an entire generation, we need to sum over all the different sub-populations:

$$\mu_g = \frac{1}{2^g} \sum_{k=1}^{2^g} \mu_g^k \tag{5}$$

where μ_g^k is the mean of k-th sub-population at generation g and 2^g is the total number of sub-populations for that generation. This equation [A5] can be simplified into the form (see Appendix for the full computation):

$$\mu_g = \left(\frac{1}{2}\right)^g \mu_0 \tag{6}$$

where μ_0 is the mean of the initial population. It is relevant to notice, that independently of the characteristic of Π(*f*), the mean always halves itself at each generation. This result is simply understood considering that for each cell that inherits a fraction f' of the compound, its sisters inherit a fraction $1 - f'$. Hence, at every division, the compound counts is diluted by an average factor of 2 in the population, and partition distributions with different properties lead to the same behavior of μ_g (see Figure 1(f)). Therefore, no relevant insights on the Π(*f*) can be obtained from the first moment. We thus moved to consider the second moment, i.e. the variance. Similarly to what is done for the mean, the variance of the inherited fraction of elements can be expressed as:

$$\begin{aligned}\sigma_{2i}^2 &= E\left[m_{2i}^2\right] - E[m_{2i}]^2 \\ &= \int df\, dm_i\, dm_{2i}\, m_{2i}^2 \delta\left(m_{2i} - f m_i\right) \prod(f)\, P\left(m_i\right) - \mu_f^2 \mu_i^2 \\ &= E\left[f^2\right] E\left[m_i^2\right] - \mu_f^2 \mu_i^2 = E\left[f^2\right] \sigma_i^2 + \mu_i^2 \sigma_f^2\end{aligned} \tag{7}$$

and in turn, the variance of a generation, starting from the consideration of the variance of a mixture of distributions can be written as (see Supplementary materials for details):

$$\sigma_g^2 = \frac{1}{2^g} \sum_{k=1}^{2^g} \left(\sigma_{g,k}^2 + \mu_g^2\right) - \mu_g^2 = A_g - \mu_g^2 \tag{8}$$

with

$$A_g = \frac{1}{2^g} \sum_{k=1}^{2^g} \left(\sigma_{g,k}^2 + \mu_g^2\right) \tag{9}$$

It is possible to obtain a recursive equation for A_g (see Appendices), $A_g = A_{g\text{-}1} E[f^2] = A_0 E[f^2]^g$ which leads to a compact form for equation [A8]:

$$\sigma_g^2 = \mu_0^2 \left(E\left[f^2\right]^g - (1/2)^{2g}\right) + \sigma_0^2 E\left[f^2\right]^g \tag{10}$$

where $E\left[f^2\right] = \sigma_f^2 + \mu_f^2 = \sigma_f^2 + 1/4$.. Finally, equation 10, links the variance of the components distribution for the entire cells population at generation g, to the properties of the Π(*f*). It is interesting to highlight that there's no dependence on the symmetric properties of the process. Up to this order, the only relevant feature that shapes the dynamic is the extent of the fluctuations. Indeed, as shown in Figure 1(e), the same Π(*f*), but with a different variance shows a different behavior for σ_g^2. Our analytical results show that the dynamical behavior of the population variance depends non-trivially on the second moment of the underlining partitioning distribution. The next section will be devoted to demonstrating, how this condition can be exploited in experiments to measure the partition noise component.

B. Measuring fluctuations via flow cytometry

As shown by Peruzzi *et al.* [13] for leukemia cells growing in suspension, it is possible to use flow cytometry to follow the evolution of properly marked cellular elements in time. Here, we generalized the protocol considering a panel of colon cell lines, *i.e.* Caco2, a human epithelial colorectal adenocarcinoma cell line; HCT116, a human colorectal carcinoma cell line; and CCD-18Co, a normal human colon fibroblast cell line. To measure the degree of asymmetric division, we marked cells' cytoplasm via a fluorescent dye and followed its dilution through successive generations. As explained in detail in the Materials and Methods section, the experimental procedure we used is based on the following main steps: staining of the selected compound,

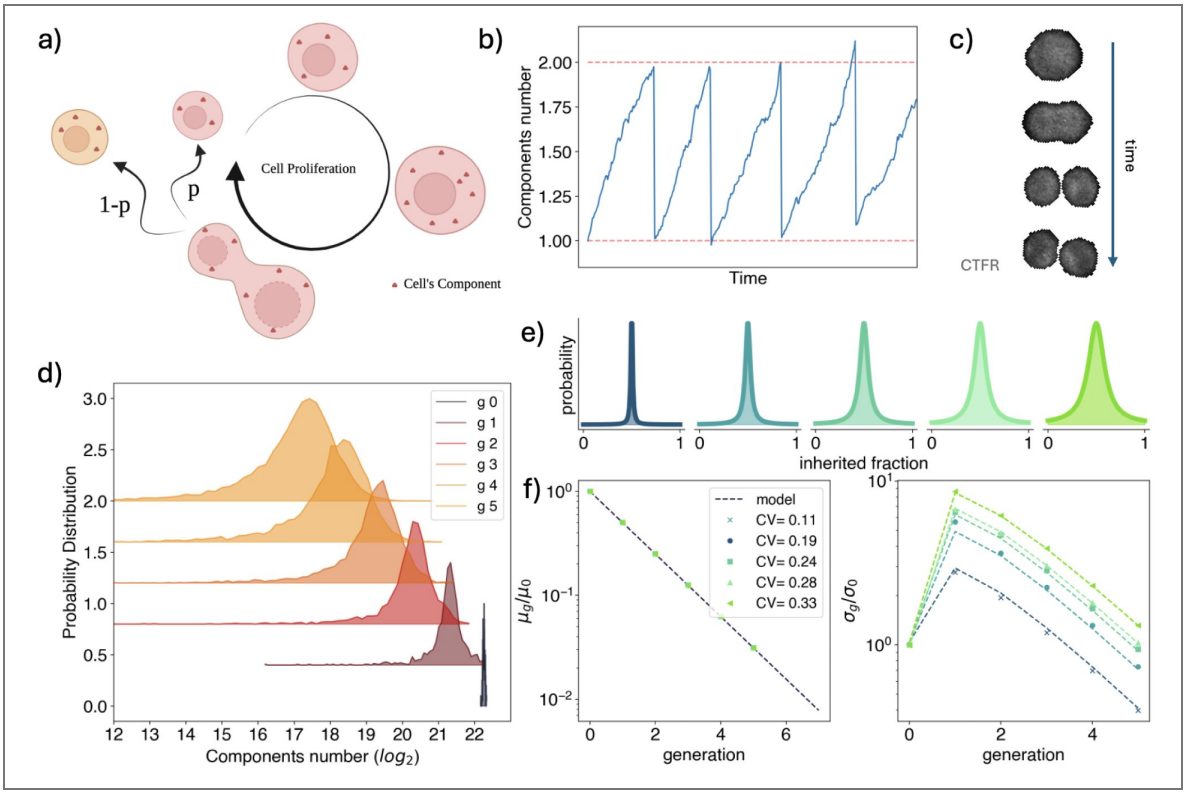


Fig. 1. Modeling partitioning noise at cell division.

a) Schematic representation of the growth and division process of a mother cell. The cell undergoes an initial phase of duplication, where its internal elements are multiplied, followed by a division phase, where these elements are partitioned between the two daughter cells. One daughter inherits a fraction p of the mother elements, while the other receives the remaining fraction $q = 1 - p$. **b)** Idealized behavior of the components number of a cell element over time, considering three different noise terms: fluctuations in compound counts during growth, uncertainty in the timing of division, and noise in the compound's partition fraction. **c)** Microscopy images of a single cell division event for an HCT116 cell, whose cytoplasm is stained with Celltrace Yellow. **d)** Time evolution of the distribution of the components number of a cellular element for a population of cells subjected only to partitioning noise. The distribution used for the simulation is a Cauchy distribution with location $x_0 = \frac{1}{2}$ and scale $\gamma = 0.07$. **e)** Examples of partition distributions Π, with increasing coefficient of variation. **f)** Mean (μ_g) and variance (σ_g) of the components number as a function of the generations, g, for the proliferation of a population subject to different partition noise distributions. Different colors correspond to distributions with varying coefficients of variation, as represented in the top panels. The dashed line represents the theoretical behavior obtained from the model. Dots are colored according to the distributions shown in panel e).

sorting of the initial population, plating, and acquisition of the distinct populations in the following days. The sorting allows us to have an initial population with a narrow peak at a specific fluorescence intensity.

To achieve this, we used fluorescence-activated cell sorting (FACS) to isolate a specific cell population based on cell morphology and fluorescence intensity. The sorted population is divided into distinct wells, one per acquisition, in equal numbers and kept under identical growth conditions to mitigate inoculum-dependent variability among samples [18]. Figure 2(a) shows an example of the time course dynamics obtained with HCT116 cells (see Methods for details). Each distribution represents a different acquisition of a whole well at increasing times from plating (0 to 84 hours from the bottom to the top). It is possible to neatly observe the succession of peaks associated with different generations that succeed one another and the average decrease of the fluorescence intensity. Note that different generations can coexist at the same time point. One can spot the flow of cells towards higher generations and the behavior of the peak heights, which grow over time, reach dominance and then decay. The shift on the y-axis allows to follow the evolutionary dynamics while being able to see the overlapping of the same generation at different time points.

As in each measurement, the distribution of fluorescent intensity is a mixture of distributions corresponding to different generations:

$$p\left(f\right)=\sum_{g}\pi_{g}P\left(f_{g}\right) \tag{11}$$

To obtain the necessary information on the properties of each generations we performed a fitting of the data via a Gaussian Mixture Model (GMM) combined with an Expectation Maximization algorithm (details are explained in “Materials and Methods”). Indeed, assuming that each generation is log-normally distributed with mean μ_g and variance σ_g^2 of the corresponding Gaussian distribution, equation 11 can be rewritten as :

$$p\left(log_2\left(f\right)\right)=\sum_{g}\pi_{g}N\left(log_2\left(f_{g}\right)|\mu_{g},\sigma_{g}\right).$$

An example of the outcome of the GMM algorithm is shown in Figure 2(b) for two acquisitions (48h and 60 h from plating). Blue-shaded distributions represent experimental data, while the best solution of the Gaussian mixture is shown with a solid line. Each generation’s fit is identified with a different color. Upon an overnight, we can identify generations 0, 1, and 2 in both the acquisitions. One can see that their relative fractions changed: generation 3 became dominant overnight and generation 4 appeared.

Figure 2(c) shows the behaviors of μ_g and σ_g as functions of the generation number g for three independent replicas of the time course experiments with CaCo2 cells. Dashed lines represent the best fit of the data against Equation 10, which yields the variance σ_f of the partition distribution probability, Π(f). The outcomes of the fitting procedure on the time course dynamic for all the replicas of three studied cell lines are reported in Figure 2(d). In particular, Caco2 and CCD18Co cells show a higher degree of asymmetry with respect to HCT116 as measured by the division asymmetry of the corresponding partition distribution.

C. Validation of the method via live fluorescence microscopy

The flow cytometry-based protocol combined with the theoretical model provides insights into the properties of the cell partition function (see Figure 2). To validate our approach, we performed extensive live time-lapse microscopy experiments to measure the partition fraction of a labelled compound during proliferation, providing a direct measure of the Π(f). Specifically, we aimed to follow a growing cell colony over time, detect divisions, evaluate the fluorescence intensity of mother and daughter cells, and calculate the inherited fraction. The detailed experimental procedure is described in the Materials and Methods. Briefly, cells are stained for the cytoplasm

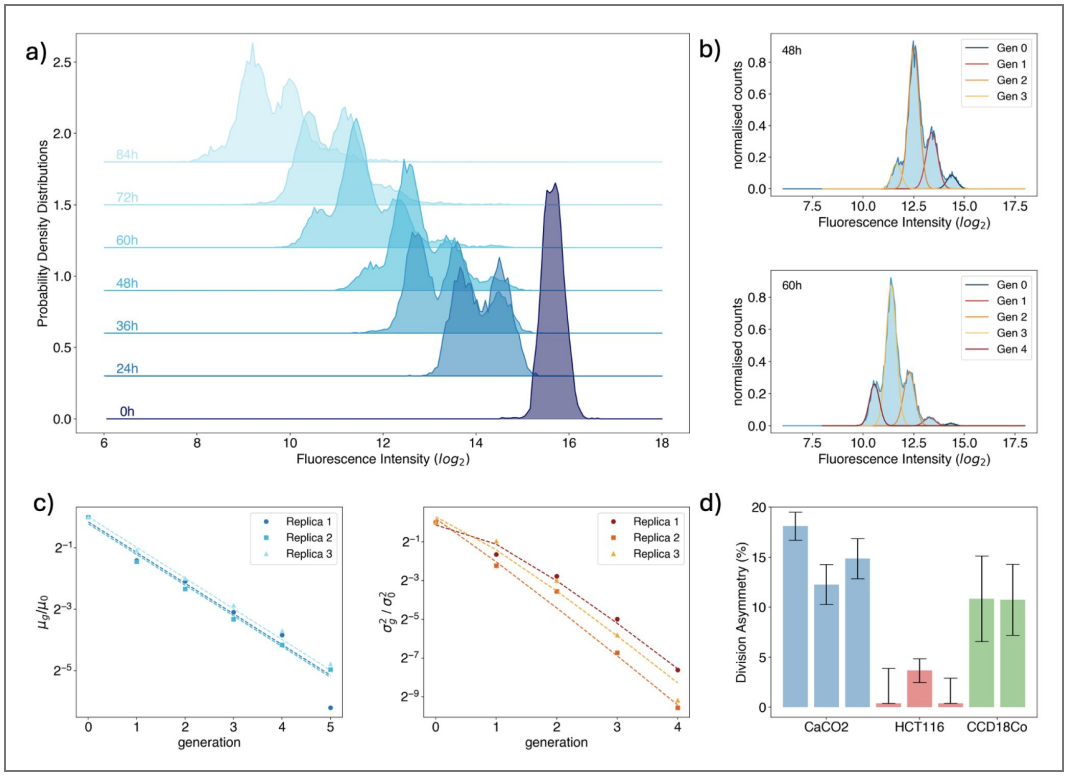


Fig. 2. Quantification of partitioning noise via population-level measurements.

a) Time evolution of CellTrace-Violet^{TM} fluorescence intensity distribution measured in a flow cytometry time course experiment for a population of HCT116 cells. Time progresses from the darker shade of blue to the lightest, spanning from [0, 84] hours (bottom line to top). **b)** Snapshots of the evolution of the distribution of CellTrace-Violet fluorescence intensity measured in a flow cytometry time course experiment for a population of HCT116 cells. Experimental data are represented by the light blue histogram, while the best-fit Gaussian Mixture Model is displayed as lines, with different colors representing different generations. **c)** Mean (left) and variance (right) of the intensity of the fluorescent markers as a function of generations, normalized to the initial population values. Each replica of the experiment is identified by a different color and point marker's shape. The experiment showed are conducted on Caco2 cells and the points correspond to the mean values for each generation. Dashed lines represent the best fit according to Equations 10 and A7, respectively. **d)** Division asymmetry of the $\Pi(f)$ obtained by fitting Equation 10 to data for all experiments and cell lines. The division asymmetry is measured via the percentage of the coefficient of variation.

and are plated at low density (approximately 10^3 cells/well) on IBIDI cell imaging chambers (μ-Slide 4 and 8 wells). After 24 hours (sufficient time for complete cell adhesion to the slide) the wells are washed, the media is renewed and live imaging acquisition time-lapse is started. Multiple fields are followed in each well, and for each field, we acquire a z-stack bright-field and fluorescence image (a single plane is shown in Figure 3a) at constant intervals (20 min) for 3 days. The fields are manually selected at the beginning of the acquisition based on cell density and fluorescence intensity. The recorded time lapses have been then manually analyzed to identify divisions. A sample outcome of the analysis procedure is shown in Figure 3(b). The total fluorescence of the cells is displayed as a function of time, before and after the division. The mother cell fluorescence decays exponentially over time and at division it halves due to components partitioning between the two daughters. In this specific case, we can observe that the fraction of cytoplasm inherited by the two cells is not equal, indicating an asymmetrical division. To compute the partition fraction *f*, we fit the fluorescence intensity of the daughters' cells vs time with:

$$log\left(I_{2i}\left(t\right)\right) = mt + q_{2i} \tag{12}$$

$$log\left(I_{21+1}\left(t\right)\right) = mt + q_{21+1} \tag{13}$$

where $I(t)$ is the fluorescence intensity at time t, m accounts for the decaying process and $2i$ and $2i + 1$ respectively refer to the two daughter cells. We constrain the slope to be the same between the two cells. The fraction of inherited fluorescence, f, is obtained as :

$$f_{2i} = \frac{e^{q_{2i}}}{e^{q_{2i}} + e^{q_{2i+1}}} \text{ and } f_{2i+1} = 1 - f_{2i} \tag{14}$$

To ensure the highest possible accuracy, we (i) measured the total fluorescence of the mother cell and those of the two daughter cells for at least 2 hours before and after the division event to increase the determination of the fluorescence splitting for each detected mitosis. Moreover, (ii) we removed all dynamics that showed an absolute Pearson correlation value between luminosity and pixel size higher than 0.9 as a function of time (see Supplementary Information). In fact, single cell's pixel size, which corresponds to the cell size projection on the focal plane, is found to shrink before mitosis and enlarges after division. Fluorescent intensity is instead proportional to the number of stained cytoplasmic proteins, which are expected to remain constant (except for a constant decay of the fluorescence). Therefore, a high correlation between pixel size and fluorescence intensity indicates a high noise-to-signal ratio. Figure 3(c) (bottom) displays the strip plot of the obtained partition fraction, f, for Caco2, HCT116, and CCD18Co cells, randomly spread on the y-axis, and their corresponding fitted $\Pi(f)$ distributions (top). Note that via microscopy imaging, it is possible to measure the whole partition distribution and not just its moments; however, hundreds of events are required to get a reliable estimate. Here, we characterize it by fitting the data with a double Gaussian distribution of the form:

$$N\left(1/2, \sigma\right) = \frac{N\left(f, \sigma'\right) + N\left(1 - f, \sigma'\right)}{2}$$

that allows for the measure of both $< f >$ and $\sigma_f^2 = \sigma'^2 + f^2 - f - 1/4$.

The fitted distributions clearly show how HCT116 cells are the most symmetric with only one central and narrow peak, while for CCD18Co and CaCO2 cells two asymmetric peaks are visible, associated to higher fluctuations. In Figure 3(d), we compare the outcomes of the flow cytometry vs microscopy measurements. For all three cell lines, the found degrees of asymmetry are statistically consistent between the two approaches.

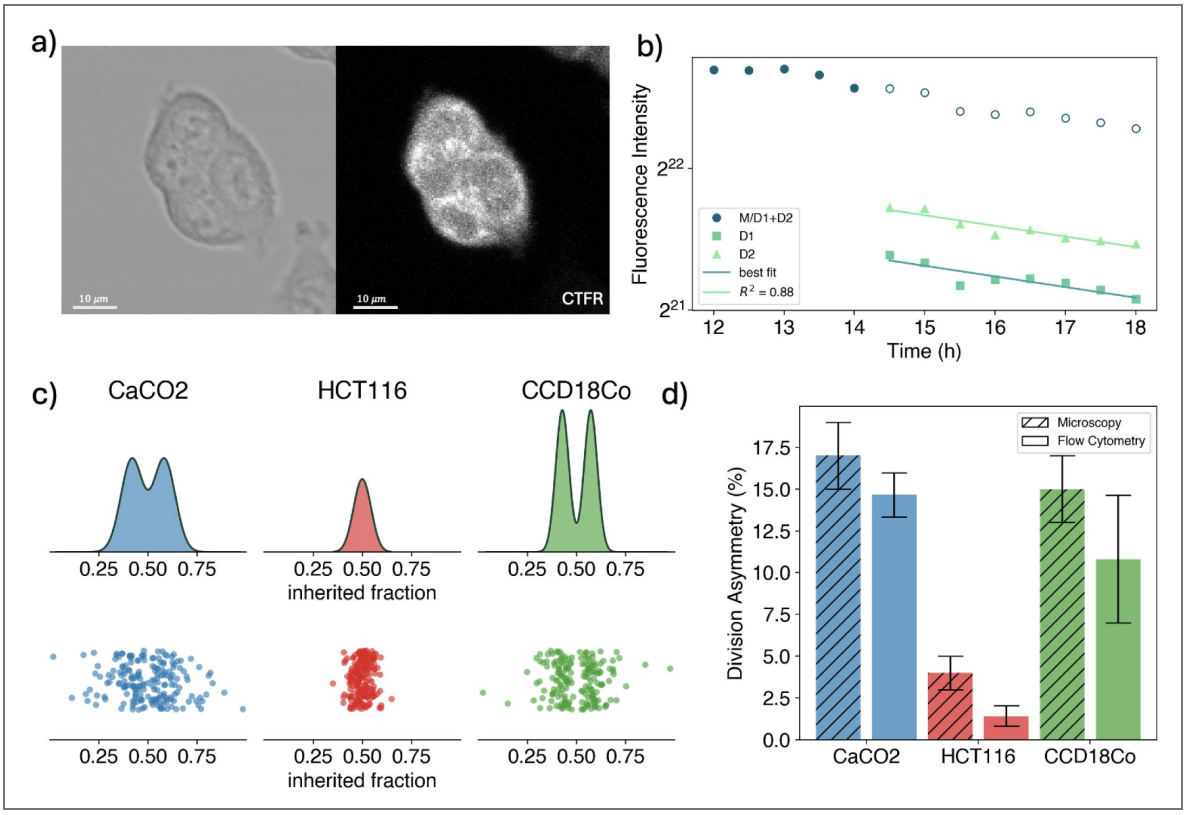


Fig. 3. Quantification of partitioning noise via single-cell measurements.

a) Example of a recorded cell colony of HCT116 cells in brightfield (left) and on CTFR fluorescence (right). **b)** Cell cytoplasm fluorescence intensity as a function of time for a cell before and after division. Dark green circles correspond to the fluorescence intensity of the mother cell up to the division frame, and then to the sum of the daughters' fluorescence. Lighter green triangles and squares represent the fluorescence intensity of the daughter cells. Solid lines are the linear fit of the points. The intercepts of the linear fit are used to compute the fraction of tagged components inherited by the daughter cells. The time is counted from the start of the experiment. **c)** (bottom) Strip plot of the distribution of the inherited fraction of cytoplasm for the different cell lines. The points are randomly spread on the y-axis to avoid overlay. (top) Fit of the inherited fraction distribution with the sum of two Gaussians with mean symmetric to 1/2. **d)** Comparison of division asymmetry obtained with time-lapse fluorescent microscopy measures (striped bars) with the ones obtained from flow cytometry experiments (plain bars). The flow cytometry bars are obtained as the mean over the multiple conducted experiments.

D. Size division bias accounts for cytoplasmic fluctuations

To investigate the observed fluctuations in the partition of the cell cytoplasm, we start by recalling that the used dyes bind a-specifically to cytoplasmic amines. Thus, (i) it is expected to be uniformly distributed in the cellular cytoplasmic space and (ii) the number of labeled cytoplasmic components can be considered large. In this framework, the least complex model one can assume is a binomial one with the parameter p, measuring the bias in the process. Via equation 10 we got a direct link between the measurable variance of the population and the second moment of the underlining partition probability distribution, $\Pi(f)$. Henceforth we sought for a relationship between the second moment of $\Pi(f)$ and the parameters of the binomial distribution.

To begin with, we can write the partition distribution for the fraction of inherited component, assuming a level of asymmetry p ($q = 1 - p$), as:

$$\prod(f) = \frac{1}{2}\left(\prod(f)^{(p)} + \prod(f)^{(q)}\right). \tag{15}$$

As we have already observed in the general model, due to symmetry, $< f >= 1/2$ for any form of the $\Pi(f)$. No information on the system can be obtained by looking at the first moment.

For the second moment, we note that the variance σ_f^{2p} of a single branch of the $\Pi(f)$ is:

$$\begin{aligned}\sigma^{2(p)} &=< f^2 >^{(p)} - < f >^{2(p)} \\ &= \int dN_i \sigma^2_{f|N}{}^{(p)} P(N_i)\end{aligned} \tag{16}$$

where $P(N_i)$ is the probability distribution of the mother's components number at the moment of division, while $\sigma^2_{f|N_i}{}^{(p)}$, is the variance of the $\Pi(f)^{\mathrm{p}}$ given the number of internal component N_i in the mother cell.

Recalling that m_{2i} is the number of inherited intracellular components of one of the two daughter cells, the partitioning fraction f can be expressed as $f = m_{2i}/N_i$ which leads to:

$$\begin{aligned}\sigma^2_{f|N_i}{}^{(p)} &=< f^2 >^{(p)} - < f >^{2(p)} \\ &= \tfrac{1}{N_i^2}\sigma^2_m{}^{(p)} = \tfrac{1}{N_i^2} N_i pq = \tfrac{pq}{N_i}.\end{aligned} \tag{17}$$

Therefore, by substituting it into equation B2 we obtain:

$$\sigma^{2(p)} = pq \int dN_i \frac{1}{N_i} P(N_i), \tag{18}$$

Which explicitly depends on $P(N_i)$ and coherently returns $\sigma^{2(p)} = \frac{pq}{N}$ for $P(N_i) = \delta(N_i - N)$.

An equivalent computation can be done for the symmetric branch ($\sigma^{2(p)}$). The variance of the full distribution is given by:

$$\begin{aligned}\sigma_f^2 &=< f^2 > - < f >^2 \\ &= \tfrac{1}{2}\left(< f^2 >^{(p)} - < f^2 >^{(q)}\right) - \tfrac{1}{4} \\ &= pq \int dN_i \frac{1}{N_i} P(N_i) + \tfrac{p^2+q^2}{2} - \tfrac{1}{4}.\end{aligned} \tag{19}$$

Note that this relationship depends on the integral:

$$\sum(N) = \int dN_i \frac{1}{N_i} P(N_i)$$

which requires the knowledge of the mother element distribution $P(N_i)$ to be computed. To get an idea of its behavior, we observe that assuming a population of cells with a fixed number of elements before division, i.e. $P(N_i) = \delta(N - N_i)$, the integral simply goes as $1/N$; in a more realistic scenario, the elements can be considered log-normally distributed, $P(N_i) = \frac{1}{N_i\sqrt{2\pi\sigma^2_{N_i}}}\exp\left(\frac{-(\ln(N_i)-\mu)^2)}{2\sigma^2_{N_i}}\right)$. as that the log-normal distribution is the limit distribution for the product of random variables, similar to what happen to number of tagged components in a cell as product of the inherited fraction of all the previous divisions. In this case, one gets the following expression for σ_f^2:

$$\sigma_f^2 = \frac{1}{2}e^{\sigma^2_{N_i}/2-\mu}\left(1 - er\,f\left(\frac{\sigma^2_{N_i}-\mu}{\sqrt{(2)}\sigma_{N_i}}\right)\right) + \frac{p^2+q^2}{2} - \frac{1}{4} \tag{20}$$

where $\mu =< log(N_i) >$.

The quantity σ_{N_i} is obtainable from flow cytometry/microscopy measurements. If the total cell fluorescence is I and F is the fluorescence intensity of each component, then $I = N_i F$ and $\sigma_I^2 = \sigma^2_{N_i}$. The values of μ are instead not directly measurable. However, form Figure 4(a), showing the value of $\Sigma(N)$ as a function of N for fixed values of σ_{N_i}, one can see that already for $N = 100$, the role of $\Sigma(N)$ is negligible. In general, for the partitioning of components characterized by a great components number this factor can be neglected. In this regime, the expression for σ_f^2 becomes:

$$\sigma_f^2 = \frac{p^2+q^2}{2} - \frac{1}{4}. \tag{21}$$

By inverting this relationship, we obtained a general mapping between the variance and the level of asymmetry of the binomial distribution.

$$p = \frac{1}{2} - \sigma_f \tag{22}$$

In Figure 4(c), we compared the analytical expression to the values obtained via microscopy measurements, showing that also for the binomial approximation the two methods are compatible. CaCO2 cells confirm to be as the most asymmetric among the cell lines and since they are known for their heterogeneity in cell morphology [19], we explored the hypothesis of shape heterogeneity and cytoplasmic asymmetry to be linked. Figure 4(d) displays three sample cases of division examined through time-lapse fluorescent microscopy following cytoplasm.

Different time frames are overlaid to depict the entire dynamic at once, with time progressing from the bottom left to the top right of each image.

We computed the pixel size ratio between the daughters cells in the last studied frame and compared it with the fraction of inherited cytoplasm from the mother cell. Although pixel size is a rough proxy for cell volume, we observe that the cell inheriting a larger fraction also displays a greater size. This finding leads us to conclude that asymmetry in cytoplasm partition is linked to size fluctuations, where size dictates the observed bias in the segregation processes.

III. Discussions

Cell division is orchestrated by hundreds of molecular interactions, which are intrinsically stochastic processes [20]. The presence of such stochasticity results in the insurgence of variability among the cells phenotypes sharing the same genome, as it is the case for cancer cells. In this respect, gene expression noise has been extensively characterized [21] as well as the different strategies that cells evolved to regulate the extent of the noise in these channels [22, 23]. Beside such noise sources, proliferating cells are subject to the fluctuations originated by the partition of cellular elements at division. Analysis of this partition noise highlighted that under some

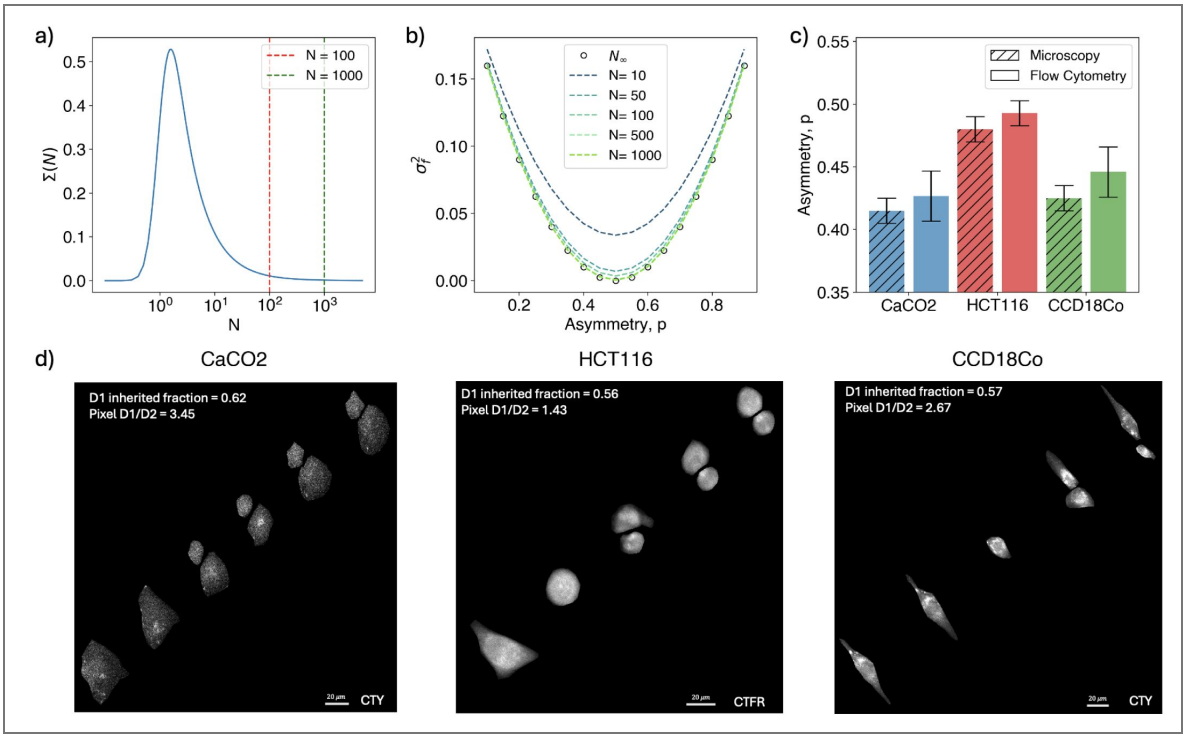


Fig. 4. Cytoplasm partition fluctuations vs cells size.

a) Behavior of the integral term Σ(*N*) (Eq.B10) as a function of the number of dividing elements, *N*, at fixed σ_{N_i} = 0.8. Vertical lines mark typical values of cellular elements, like mitochondria. **b)** Theoretical value of the variance of Π(*f*) in the binomial assumption for different levels of asymmetries and increasing values of *N*. **c)** Asymmetry of the partitioning distribution Π(*f*) in the binomial limit as measured by the binomial bias, p. **d)** Sample cases of volume asymmetric division for different cell lines. Images show the overlay of consecutive times during the division dynamic. Time flows from the bottom left to the top right.

circumstances, its contribution exceeds the others in creating heterogeneity [14, 24]. In fact, asymmetrically dividing cells can produce daughters that differ in size, cellular components, and, in turn, fate [25]. In this framework, the origin of such asymmetry may be ascribed to the basal stochasticity of molecular processes taking place at different levels during the division and/or to specific, evolutionary-conserved mechanisms used by cells to control cell fate and generate diversity [3, 26]. In fact, asymmetric partitioning of components alters the initial counts of molecules in the following cell cycle, which can lead the system to a different phenotypic state [27, 28]

Here, we show that an accurate determination of partition fluctuations can be obtained via standard flow cytometry measurements of properly marked cell populations. Our approach is easier and faster with respect to the use of microscopy imaging, which requires the acquisition and analysis of extensive recording of the cells, ultimately limiting the reachable statistics in contrast to the high-throughput readouts of flow cytometry. Moreover, most of the microscopy-based approaches use fixed cells and measure the inherited fraction of fluorescence from a single snapshot of dividing cells. Notably, comparing the degree of asymmetry in the division of the cytoplasm in the three different analysed cell types, we found that microscopy-derived fluctuations are systematically higher than those measured by flow cytometry. While this could be linked to the different used protocols, it seems more probable that the difference is due to a systematic overestimation of the fluctuations measured by microscopy due to its lower statistics.

From a biological point of view, the segregation statistics among the analysed colon cell types are neatly distinct. CaCo2 cells exhibits the highest degree of cytoplasmic partition fluctuations, while HCT116 division is the one with lowest asymmetry. It is interesting to note that normal colon fibroblasts sit in between, suggesting that the cancer-associated dis-regulation of the normal cell activities leads to the reactivation of asymmetric cell division in a cancer-specific manner. Indeed, CaCo2 cells are known to display a markedly heterogeneous distribution of phenotypes [19].

To get insights on the possible mechanisms behind the observed asymmetries, we looked for specific shape of the partition distribution. Following a maximum entropy approach [20, 29, 30], we started by considering the distribution that employs the fewest assumptions about the system, i.e. we consider independent segregation, which directly maps into binomial partitioning. Indeed, this is a common assumption in segregation models that traces back to the pioneering works of Berk [31] and Rigney [32]. Notably, our results indicated biased segregations, with colon cancer cells displaying different biases.

To explain the origin of those biases, we looked at relative sizes between the daughter cells right after division. The found correlation between daughter sizes and the fractions of inherited fluorescence vouches size as the origin of the observed biased segregation.

In conclusion, by combining experimental data with statistical modelling, we show how it is possible to use flow cytometry data to extract reliable estimates of the strength of fluctuations during cell division. Our approach has the potential to be applied to different cell types where a quantification of the level of division asymmetries daughter cells can experience may provide new insights into the mechanisms of asymmetric cell division and its role in cancer heterogeneity and plasticity.

IV. Materials and Methods

A. Cell Culture

Caco-2, colorectal adenocarcinoma cell line, was purchased from ATCC (Manassas, VA, USA, HBT-37) and maintained in complete culture media DMEM (D6046) containing 20% FBS, penicillin/streptomycin plus glutamine, nonessential amino acids (NEAA) and sodium pyruvate, all 1/100 dilution. HCT116 VIM RFP, colorectal carcinoma cell line, was purchased from ATCC (Manassas, VA, USA, CCL-247EMT) and maintained in complete culture media McCoy's 5A (M9309) containing 10% FBS, penicillin/streptomycin plus glutamin.

CCD-18Co, colon fibroblast cell line, were purchased from ATCC (Manassas, VA, USA, CRL-1459) and maintained in complete culture media DMEM containing 10% FBS, penicillin/streptomycin plus glutamin, nonessential amino acids (NEAA) and sodium pyruvate, all 1/100 dilution.

All cells were kept in culture at 37°*C* in 5% CO2 and passaged according to the experimental protocol. For all experiments cells were harvested, counted, and washed twice in serum-free solutions and resuspended in room temperature (RT) PBS w/o salts for further staining.

In detail, to detach cells from culture flasks we used Trypsin EDTA 0.25%. Culture media was removed and the flasks were washed once with PBS, then Trypsin EDTA was added and allowed to work for 1 min in an incubator at 37°*C*. Once the cells appeared detached the flasks were mechanically agitated to facilitate cells' loss of adhesion. Trypsin was then inactivated with one volumes of complete media and cells were collected, pelleted, and washed a second time in PBS. To determine cell viability, prior to dye staining, the collected cells were counted with the hemocytometer using Trypan Blue, an impermeable dye not taken up by viable cells.

B. Flow Cytometry and Cell Sorting

To track cell proliferation by dye dilution for establishing the progeny of a mother cell, cells were stained with CellTrace Violet^{TM} (CTV). The CTV dye staining (C34557, Life Technologies, Paisley, UK), used to monitor multiple cell generations, was performed according to the manufacturer's instruction, diluting the CTV 1/1000 in 0.5-1 ml of PBS for 20 min in a water bath at 37°C, mixing every 10 min. Afterward, 5x complete media was added to the cell suspension for an additional 5 min incubation before the final washing in PBS.

Labeled cells were sorted using a FACSAriaIII (Becton Dickinson, BD Biosciences, USA) equipped with Near UV 375nm, 488nm, 561nm, and 633nm lasers and FACSDiva software (BD Biosciences version 6.1.3). Data were analyzed using FlowJo software (Tree Star, version 10.7.1). Briefly, cells were first gated on single cells, by doublets exclusion with morphology parameters area versus width (A versus W), both side and forward scatter. The unstained sample was used to set the background fluorescence. The sorting gate was set around the max peak of fluorescence of the dye distribution. In this way, the collected cells were enriched for the highest fluorescence intensity. Following isolation, an aliquot of the sorted cells was analyzed with the same instrument to determine the post-sorting purity and population width, resulting in an enrichment > 97 % for each sample.

To monitor multiple cell division, the sorted cell population was seeded in 12 wells plates (Corning, Kennebunk, ME, USA) at cell density bewteen [30–70]·10^3 cells/well, according to the experiment, and kept in culture for up to 84 h. Each well corresponds to a time point of the acquisition and cells in culture were analyzed every 24, 36, 48, 60, 72, 84 h by the LSRFortessa flow cytometer. In order to set the time zero of the kinetic, prior culturing, a tiny aliquot of the collected cells was analyzed immediately after sorting at the flow cytometer. The unstained sample was used to set the background fluorescence as described above.

C. Time-lapse Microscopy

To better investigate the cell proliferation dynamics, we performed time-lapse experiments, for up to 3 days. In our previous paper [13] we verified that cell growth is not affected by different dyes combination. Therefore, in the present work, we used both CellTrace Yellow^{TM} (C34573 A, Life Technologies, Paisley, UK) and CellTrace Far Red^{TM} (C34572, Life Technologies, Paisley, UK), to stain the cytoplasm for microscopy analysis. Low passage cells at around 60 % confluency were counted and stained with CellTrace dyes 1/500 and plated on IBIDI cell imaging chambers (μ-Slide 4 and 8 wells) at low cell density of 10^3 cells/well. After overnight incubation, the chamber is transferred to the inverted microscope adapted with an incubator to keep cells in appropriate growing conditions. Brightfield and fluorescent confocal image stacks were acquired with a 20x air objective (Olympus, Shinjuku, Japan) and Zen Microscopy Software (Zeiss, Oberkochen, Germany), every 20 min. Time-lapse images were analyzed using ImageJ and inhouse Python programs.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

All codes used to produce the findings of this study are available from the corresponding author upon request. The code for the Gaussian Mixture algorithm is available at *https://github.com/ggosti/fcGMM*.

Acknowledgements

This research was partially funded by grants from ERC-2019-Synergy Grant (ASTRA, n. 855923); EIC-2022-PathfinderOpen (ivBM-4PAP, n. 101098989); Project 'National Center for Gene Therapy and Drugs based on RNA Technology' (CN00000041) financed by NextGeneration EU PNRR MUR—M4C2—Action 1.4—Call 'Potenziamento strutture di ricerca e creazione di campioni nazionali di R&S' (CUP J33C22001130001).

Additional information

Author contributions statement

M.M. and G.G. conceived research; D.C., G.P., and S.S. performed cell biology experiment and flow cytometry measurements. D.C., C.G., V.d.T., and M.M. performed timelapse microscopy experiments; F.G. and G.R. contributed additional ideas and helped interpret results; D.C. analyzed data; M.M. performed analytical calculations; D.C. and G.G. performed numerical simulations and statistical analysis; all authors analyzed results, and wrote and revised the paper.

Appendix A: General model for the partitioning noise

We assume that at time zero each cell has m_i marked compounds, with a certain distribution in the whole cell population of mean μ and variance σ^2. Each cell in the population, if sufficiently healthy, will divide in two cells and we would like to follow the evolution of the marker intensity distribution. The splitting ratio is one of the parameter we want to characterize.

Therefore, by calling mi the number of compounds of the mother cell, m_{2i} and m_{2i+1} the compound in the two daughter cell. The probability of finding a cell with m_{2i} marked compounds at generation g is:

$$\begin{aligned} P\left(m_{2i}|m_i\right) &= \int \delta\left(m_{2i} - f m_i\right) \prod\left(f\right) df \\ P\left(m_{2i+1}|m_i\right) &= \int \delta\left(m_{2i} - \left(1-f\right) m_i\right) \prod\left(f\right) df \end{aligned} \tag{A1}$$

where Π(f) is the distribution probability of the fraction of compound that goes into a daughter cell.

Since we do not study single cells, but populations, we must deal with distributions. We will omit the computation for m_{2i+1} since it is equal to m_{2i} if f is substituted with 1 – f.

$$P\left(m_{2i}\right) = \int P\left(m_{2i}|m_i\right) P\left(m_i\right) dm_i$$

Which according to equation A1 becomes:

$$P(m_{2i}) = \int df\, dm_i \delta(m_{2i} - f m_i) \prod(f) P(m_i) \tag{A2}$$

Hence, the expectation value for m_{2i} is :

$$\mu_{2i} = E[m_{2i}] = \int df\, dm_i P(m_{2i})\, m_{2i} dm_{2i} = \int df\, dm_i dm_{2i} m_{2i} \delta(m_{2i} - f m_i) \prod(f) P(m_i) = E[f] E[m_i] = \mu_f \mu_i \tag{A3}$$

while for m_{2i+1}:

$$\begin{aligned} \mu_{2i+1} &= E[m_{2i+1}] = \int df\, dm_i P(m_{2i+1})\, m_{2i+1} dm_{2i1+} = \\ &= \int df\, dm_i dm_{2i+1} m_{2i+1} \delta(m_{2i+1} - (1-f) m_i) \prod(f) P(m_i) \\ &= E[1-f] E[m_i] = \mu_{1-f} \mu_i \end{aligned}$$

For the variance we can do a similar computation:

$$\begin{aligned} \sigma_{2i} &= E[m_{2i}^2] - E[m_{2i}]^2 \\ &= \int df\, dm_i dm_{2i} m_{2i}^2 \delta(m_{2i} - f m_i) \prod(f) P(m_i) - \mu_f^2 \mu_i^2 = \\ &\qquad E[f^2] E[m_i^2] - \mu_f^2 \mu_i^2 \\ &= E[f^2] \sigma_i^2 + \mu_i^2 \sigma_f^2 \end{aligned} \tag{A4}$$

Therefore, both variance and mean of the daughters subpopulations are linked to the variance and mean of the population of mother cells and to the distribution probability of *f*.

To be able to compare the model with experimental data we need to compute the mean and variance of the whole population, which is the sum of coexisting subsequent generations.

$$\mu_g = \frac{1}{2^g} \sum_{k=1}^{2^g} \mu_g^k \tag{A5}$$

Where we are assuming that a mother cell always give rise to only two daughters. We can split equation A5 into two groups, the one that always inherits a fraction $E[f]$ of the mother's compounds and the symmetrical one which inherits a fraction $E[1-f]$.

$$\mu_g = \frac{1}{2^g} \left[\sum_{k=1}^{2^{g-1}} E[f] \mu_{g-1}^k + \sum_{k=1}^{2^{g-1}} E[1-f] \mu_{g-1}^k \right]$$

Which can be rewritten as:

$$\mu_g = \frac{1}{2^g} \sum_{k=1}^{2^{g-1}} \mu_{g-1} = \frac{1}{2} \mu_{g-1} \tag{A6}$$

Therefore, knowing the initial mean value of the population, μ_0 we have:

$$\mu_g = \left(\frac{1}{2}\right)^g \mu_0 \tag{A7}$$

Equation A7 shows that the mean value of the number of compounds for each generation is independent of the underlying process and does not depend on the distribution Π(*f*). Interesting, although expected, but useless for comparing expectations to experiment.

To obtain a dependence on the partitioning process we need to compute the variance. We want the variance of a distribution which is given by the mixture of the distribution of all the sub-populations of daughter cells. It can be demonstrated that the total variance is:

$$\sigma_g^2 = \frac{1}{2^g}\sum_{k=1}^{2^g}\left(\sigma_{g,k}^2 + \mu_g^2\right) - \mu_g^2 \tag{A8}$$

As we did in A6 we can rewrite equation A8 by taking into account the two daughters sub-populations:

$$A_g = \frac{1}{2^g}\sum_{k=1}^{2^g}\left(\sigma_{g,k}^2 + \mu_g^2\right) = \frac{1}{2^g}\left[\sum_{k=1}^{2^{g-1}}\sigma_{g,k}^2 + \mu_{g,k}^2 + \sum_{k=1}^{2^{g-1}}\sigma_{g,k}^2 + \mu_{g,k}^2\right]$$

We can rewrite this formula by recalling the relationships A4 and A3.

$$\begin{aligned} A_g &= \tfrac{1}{2^g}\sum_{k=1}^{2^g}\left(\sigma_{g,k}^2 + \mu_g^2\right) \\ &= \tfrac{1}{2^g}\left[\sum_{k=1}^{2^{g-1}} E\left[f^2\right]\sigma_{g-1,k}^2 + \sigma_f^2\mu_{g-1,k}^2 + \mu_f^2\mu_{g-1,k}^2\right. \\ &\left. + \sum_{k=1}^{2^{g-1}} E\left[(1-f)^2\right]\sigma_{g-1,k}^2 + \sigma_{1-f}^2\mu_{g-1,k}^2 + \mu_{1-f}^2\mu_{g-1,k}^2\right] \end{aligned} \tag{A9}$$

We can notice that the last two elements of every sum can be rewritten as:

$$\begin{aligned} \sigma_f^2\mu_{g-1,k}^2 + \mu_f^2\mu_{g-1,k}^2 &= E\left[f^2\right]\mu_{g-1,k}^2 \\ \sigma_{1-f}^2\mu_{g-1,k}^2 + \mu_{1-f}^2\mu_{g-1,k}^2 &= E\left[(1-f)^2\right]\mu_{g-1,k}^2 \end{aligned}$$

Which leads to:

$$\begin{aligned} A_g &= \frac{1}{2^g}\left(\sum_{g-1,k}^{2}\sigma_{g-1,k}^2 + \mu_{g-1,k}^2\right)\left(E\left[f^2\right] - E\left[(1-f)^2\right]\right) \\ &= \frac{1}{2}A_{g-1}\left(1 - 2\mu_f + 2E\left[f^2\right]\right) \end{aligned}$$

and considering that $\mu_f = 1/2$.

$$A_g = A_{g-1}E\left[f^2\right] \tag{A10}$$

Therefore, the final form of equation A6 is :

$$\begin{aligned} &\sigma_g^2 = A_g - \mu_g^2 = A_{g-1}E\left[f^2\right] - \mu_f^2\mu_{g-1}^2 = A_0E[f^2]^g - \mu_f^{2g}\mu_0^2 = \\ &\mu_0^2\left(E[f^2]^g - \mu_f^{2g}\right) + \sigma_0^2E[f^2]^g = \mu_0^2\left(E[f^2]^g - (1/2)^{2g}\right) + \sigma_0^2E[f^2]^g \end{aligned} \tag{A11}$$

Appendix B: Modeling partition as a binomial process

The model we have developed and tested by comparing two sets of independent experiments allows for a robust characterisation of the partitioning noise. The measures are done via flowcytometry which allows to obtain a great statistics. The key point of what we have proved up to now, is that the variance of the partitioning distribution is the main actor in shaping the population dynamic. In this section we want to propose an interesting case of study, were we make an assumption on the shape of the partitioning distribution. The reasonable assumption is to consider the binomial distribution for the partitioning of components. In particular, for the cytoplasm the high number of tagged components.

Therefore, we want to link the measure of σ of the partitionig distribution done with the general method, to the value of p of the binomial distribution which would lead to the same observed value.

The binomial model consists in assuming that each cell's component has a certain probability p of ending up in one of the two daughter cell, and $1-p$ to end up in the other one.

Therefore, the total partition distribution can be written as:

$$P(f) = \frac{1}{2}\left(P(f)^{(p)} + P(f)^{(q)}\right). \tag{B1}$$

As we have already observed, due to symmetry, the $<f>= 1/2$. But, we are interested in how the general noise, map into the binomial one.

The variance of a single branch $\sigma_f^{2(p)}$, can be computed in the following way:

$$\begin{aligned}
\sigma^{2(p)} &= < f^2 >^{(p)} - < f >^{2(p)} \\
&= \int df \int dN_i f^2 P(f|N_i) P(N_i) - < f >^{2(p)} \\
&= \int dN_i \int dN_i f^2 P(f|N_i) P(N_i) - < f >_p^2 \\
&= \int dN_i \left(\sigma_{f|N_i}^{2\ (p)} + < f >_{p|N_i}^2\right) P(N_i) - < f >_p^2 \\
&= \int dN_i \left(\sigma_{f|N_i}^{2\ (p)}\right) P(N_i)
\end{aligned}$$

Consider initially one of the two daughter cell distributions for example, $P(f)^{(p)}$. Its average value will be given by:

$$< f >^{(p)} = \int df \int dN_i\, f P(f|N_i) P(N_i) = \int dN_i p P(N_i) = p,$$

and similarly $<f>^{(q)}= q$. For the variance instead, $\sigma_f^{2(p)}$:

$$\begin{aligned}
\sigma^{2(p)} &= < f^2 >^{(p)} - < f >^{2(p)} \\
&= \int df \int dN_i f^2 P(f|N_i) P(N_i) - < f >^{2(p)} \\
&= \int dN_i \int dN_i f^2 P(f|N_i) P(N_i) - < f >_p^2 \\
&= \int dN_i \left(\sigma_{f|N_i}^{2\ (p)} + < f >_{p|N_i}^2\right) P(N_i) - < f >_p^2 \\
&= \int dN_i \left(\sigma_{f|N_i}^{2\ (p)} P(N_i)\right)
\end{aligned} \tag{B2}$$

where the notation $\sigma^{2}_{f|N_i}{}^{(p)}$ indicates the variance of the distribution p as a function of f given a certain N_i. Recalling that f, the fraction of intracellular components and/or fluorescence, is defined as $f = \frac{x_{2i}}{N_i}$, it follows that:

$$\sigma^{2}_{f|N_i}{}^{(p)} = < f^2 >^{(p)} - < f >^{2(p)} = \frac{1}{N_i^2}\sigma_x^{2(p)} = \frac{1}{N_i^2} N_i pq = \frac{pq}{N_i}. \tag{B3}$$

Plugging Eq. B3 in B2 we get:

$$\sigma^{2(p)} = pq \int \frac{1}{N_i} P(N_i), \tag{B4}$$

which, consistently, in the case of a delta distribution (i.e., for a single value of N_i), results in $\sigma^{2(p)} = \frac{pq}{N_i}$.

Once the main quantities of the individual distribution have been determined, we consider the sum distribution (eq. B1).

The mean value of this distribution is:

$$< f >= \int f\, P(f)\, df = \frac{1}{2}\left(< f >^{(p)} + < f >^{(q)} \right) = \frac{1}{2}(p+q) = \frac{1}{2}. \tag{B5}$$

Therefore, to determine p, it is necessary to estimate the variance σ_f^2. By definition, it is:

$$\sigma_f^2 = < f^2 > - < f >^2 . \tag{B6}$$

Substituting Eq. B5 one gets:

$$\sigma_f^2 = \frac{1}{2}\left(< f^2 >^{(p)} - < f^2 >^{(q)} \right) - \frac{1}{4}, \tag{B7}$$

from which, adding and subtracting $< f >^{2(p)}$, $< f >^{2(q)}$, one gets:

$$\sigma_f^2 = \frac{1}{2}\left(< f^2 >^{(p)} - < f^2 >^{(p)} + < f >^{2(p)} + < f^2 >^{(q)} - < f >^{2(q)} + < f >^{2(q)} \right) = \tag{B8}$$

$$= \frac{1}{2}\left(\}sigma a_f^{2(p)} + p^2 + \sigma_f^{2(q)} + q^2 \right) - \frac{1}{4} \tag{B9}$$

and via Eq. B4

$$\sigma_f^2 = pq \int dN_i \frac{1}{N_i} P(N_i) + \frac{p^2+q^2}{2} - \frac{1}{4}. \tag{B10}$$

a. Delta Distribution of mother's components

In principle, $P(N_i)$ can have any functional form, even a very complex one, which is not known. Therefore, let's start with the simplest case: all mother cells possess the same number of components N_0, so the distribution $P(N_i)$ is a Dirac delta.

In this case, the theoretical variance (eq. B12) will be:

$$\sigma_f^2 = pq \int dN_i \frac{1}{N_i} \delta\left(N_i - N_0\right) + \frac{p^2+q^2}{2} - \frac{1}{4} \tag{B11}$$

$$= \frac{pq}{N_0} + \frac{p^2+q^2}{2} - \frac{1}{4}. \tag{B12}$$

Since the values of the initial distribution have been considered randomly, each of them will have the same probability of being selected: $P(N_i) = \frac{1}{N_{max}-N_{min}}$. Consequently, the integral term in the variance expression (eq. B12) can be simplified as:

$$\int dN_i \frac{1}{N_i} P\left(N_i\right) = \frac{1}{N_{max}-N_{min}} \int dN_i \frac{1}{N_i} = \frac{ln\left(N_{max}\right) - ln\left(N_{min}\right)}{N_{max}-N_{min}}, \tag{B13}$$

and one gets to the final expression:

$$\sum\left(N\right) = \sigma_f^2 = pq \frac{ln\left(N_{max}\right) - ln\left(N_{min}\right)}{N_{max}-N_{min}} + \frac{p^2+q^2}{2} - \frac{1}{4}. \tag{B14}$$

Neglecting the integral term, the equation reduces to :

$$\sigma_f^2 = \frac{p^2+q^2}{2} - \frac{1}{4}. \tag{B15}$$

b. Lognormal distribution of mothers' components

In an even more realistic case, the distribution of the number of internal components of the mother cells will be given by a log-normal distribution:

$$P\left(N_i\right) = \frac{e^{\frac{-\left(ln\left(N_i\right)-\mu\right)^2\right)}{2\sigma_{N_i}^2}}}{\sqrt{2\pi\sigma_{N_i}^2} N_i}, \tag{B16}$$

where μ is the average value: $\mu =< log(N) >$.

The integral term in this case is given by:

$$\sum\left(N\right) = \int dN_i \frac{1}{N_i} P\left(N_i\right) = \frac{1}{\sqrt{2\pi\sigma_{N_i}^2}} \int dN_i \frac{1}{N_i^2} e^{\frac{-\left(ln\left(N_i\right)-\mu\right)^2\right)}{2\sigma_{N_i}^2}}. \tag{B17}$$

In general, the distribution of internal components of a cell is not known. However, in microscopy experiments, it is possible to determine fluorescence distributions. Additionally, a direct proportionality between the two variables has been assumed: $I = N_i \cdot F$, where F is a constant value representing the fluorescence of an internal component.

Now, consider the variance of the distribution as a function of the number of internal components:

$$\begin{aligned} \sigma_{N_i}^2 = \sigma_{I/F}^2 &=< ln(I/F)^2 > - < ln\left(I/F\right) >^2 \\ &=< ln\left(I\right) - ln\left(F\right))^2 > - < ln\left(I\right) - ln\left(F\right) >^2 \\ &=< \left(ln^2\left(I\right) > - < ln\left(I\right) >^2 = \sigma_I^2 \end{aligned} \tag{B18}$$

The variance of the distribution as a function of N_i, therefore, coincides with that of the fluorescence, which can be determined from experimental measurements.

Consequently, for a log-normal distribution, the contribution of the integral term to the variance, integrated over the interval [1, ∞), will be:

$$\sum(N) = \int_1^{\infty} dN_i \frac{1}{\sqrt{2\pi\sigma^2_{N_i}}} \frac{1}{N_i^2} e^{\frac{-(ln(N_i)-\mu)^2)}{2\sigma^2_{N_i}}} = \frac{1}{2} e^{\sigma^2/2-\mu} \left(1 - erf\left(\frac{\sigma^2-\mu}{\sqrt{(2)}\sigma}\right)\right), \tag{B19}$$

where the variance coincides with that of the fluorescence distribution (eq. B18) and $\mu = \langle \log(N) \rangle$.

Substituting the integral term (eq. B19) into the complete expression for the variance (B10) for a log-normal distribution of the mother cells, we obtain:

$$\sigma_f^2 = pq \cdot \frac{1}{2} e^{\frac{\sigma^2}{2} - <log(N)>} \left(1 - erf\left(\frac{\sigma^2 - < log(N) >}{\sqrt{2}\sigma}\right)\right) + \frac{p^2+q^2}{2} - \frac{1}{4}. \tag{B20}$$

References

[1] **Conklin Edwin G.** (1905) Mosaic development in ascidian eggs. *Journal of Experimental Zoology* **2**:145-223

[2] **Knoblich Juergen A.** (2001) Asymmetric cell division during animal development. *Nature Reviews Molecular Cell Biology* **2**:11-20

[3] **Sunchu Bharath,** Cabernard Clemens (2020) Principles and mechanisms of asymmetric cell division. *Development* **147**

[4] **Mushnikov Nikolai V.,** Fomicheva Anastasia, Gomelsky Mark, Bowman Grant R. (2019) Inducible asymmetric cell division and cell differentiation in a bacterium. *Nature Chemical Biology* **15**:925-931

[5] **Yang Jing,** McCormick Mark A., Zheng Jiashun, Xie Zhengwei, Tsuchiya Mitsuhiro, Tsuchiyama Scott, El-Samad Hana, Ouyang Qi, Kaeberlein Matt, Kennedy Brian K., *et al.* (2015) Systematic analysis of asymmetric partitioning of yeast proteome between mother and daughter cells reveals "aging factors" and mechanism of lifespan asymmetry. *Proceedings of the National Academy of Sciences* **112**:11977-11982

[6] **Buss Julieti Huch,** Begnini Karine Rech, Lenz Guido (2024) The contribution of asymmetric cell division to phenotypic heterogeneity in cancer. *Journal of Cell Science* **137**

[7] **Chao Shanshan,** Yan Huiwen, Bu Pengcheng (2024) Asymmetric division of stem cells and its cancer relevance. *Cell Regeneration* **13**

[8] **Lindner Ariel B.,** Madden Richard, Demarez Alice, Stewart Eric J., Taddei Francois (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences* **105**:3076-3081

[9] **Katajisto Pekka,** Dohla Julia, Chaffer Christine L., Pentinmikko Nalle, Marjanovic Nemanja, Iqbal Sharif, Zoncu Roberto, Chen Walter, Weinberg Robert A., Sabatini David M. (2015) Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness. *Science* **348**:340-343

[10] **Izumi Hideki,** Kaneko Yasuhiko (2012) Evidence of asymmetric cell division and centrosome inheritance in human neuroblastoma cells. *Proceedings of the National Academy of Sciences* **109**:18048-18053

[11] **Dey-Guha Ipsita,** Wolfer Anita, Yeh Albert C., Albeck John G., Darp Revati, Leon Eduardo, Wulfkuhle Julia, Petricoin Emanuel F., Wittner Ben S., Ramaswamy Sridhar (2011) Asymmetric cancer cell division regulated by akt. *Proceedings of the National Academy of Sciences* **108**:12845-12850

[12] **Dey-Guha Ipsita,** Alves Cleidson P., Yeh Albert C., Salony, Sole Xavier, Darp Revati, Ramaswamy Sridhar (2015) A mechanism for asymmetric cell division resulting in proliferative asynchronicity. *Molecular Cancer Research* **13**:223-230

[13] **Peruzzi Giovanna,** Miotto Mattia, Maggio Roberta, Ruocco Giancarlo, Gosti Giorgio (2021) Asymmetric binomial statistics explains organelle partitioning variance in cancer cell proliferation. *Communications Physics* **4**:188

[14] **Soltani Mohammad,** Vargas-Garcia Cesar A., Antunes Duarte, Singh Abhyudai (2016) Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. *PLOS Computational Biology* **12**:e1004972

[15] **Martino Andrea De,** Gueudre Thomas, Miotto Mattia (2019) Exploration-exploitation tradeoffs dictate the optimal distributions of phenotypes for populations subject to fitness fluctuations. *Physical Review E* **99**:012417

[16] **Miotto Mattia,** Scalise Simone, Leonetti Marco, Ruocco Giancarlo, Peruzzi Giovanna, Gosti Giorgio (2024) A size-dependent division strategy accounts for leukemia cell size heterogeneity. *Communications Physics* **7**

[17] **Scalise Simone,** Gosti Giorgio, Ruocco Giancarlo, Peruzzi Giovanna, Miotto Mattia (2024) *Probing leukemia cells behavior under starvation*

[18] **Bena Chiara Enrico,** Giudice Marco Del, Grob Alice, Gueudre Thomas, Miotto Mattia, Gialama Dimitra, Osella Matteo, Turco Emilia, Ceroni Francesca, Martino Andrea De, *et al.* (2021) Initial cell density encodes proliferative potential in cancer cell populations. *Scientific Reports* **11**

[19] **Lea Tor** (2015) *Caco-2 Cell Line* Springer International Publishing. pp. 103-111

[20] **Bialek W.** (2012) *Biophysics: Searching for Principles* Princeton University Press.

[21] **Elowitz Michael B.,** Levine Arnold J., Siggia Eric D., Swain Peter S. (2002) Stochastic gene expression in a single cell. *Science* **297**:1183-1186

[22] **Osella Matteo,** Bosia Carla, Cora Davide, Caselle Michele (2011) The role of incoherent microrna-mediated feedforward loops in noise buffering. *PLoS Computational Biology* **7**:e1001101

[23] **Miotto Mattia,** Marinari Enzo, Martino Andrea De (2019) Competing endogenous rna crosstalk at system level. *PLOS Computational Biology* **15**:e1007474

[24] **Huh Dann,** Paulsson Johan (2011) Random partitioning of molecules at cell division. *Proceedings of the National Academy of Sciences* **108**:15004-15009

[25] **Chhabra Shaan N.,** Booth Brian W. (2021) Asymmetric cell division of mammary stem cells. *Cell Division* **16**

[26] **Miotto Mattia,** Monacelli Lorenzo (2020) Genome heterogeneity drives the evolution of species. *Phys. Rev. Res* **2**:043026

[27] **Moris Naomi,** Pina Cristina, Arias Alfonso Martinez (2016) Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics* **17**:693-703

[28] **Miotto Mattia,** Rosito Maria, Paoluzzi Matteo, Turris Valeria de, Folli Viola, Leonetti Marco, Ruocco Giancarlo, Rosa Alessandro, Gosti Giorgio (2023) Collective behavior and self-organization in neural rosette morphogenesis. *Frontiers in Cell and Developmental Biology* **11**

[29] **Miotto Mattia,** Monacelli Lorenzo (2018) Entropy evaluation sheds light on ecosystem complexity. *Physical Review E* **98**

[30] **Miotto Mattia,** Monacelli Lorenzo (2021) TOLOMEO, a novel machine learning algorithm to measure information and order in correlated networks and predict their state. *Entropy* **23**:1138

[31] **Berg Otto G** (1978) A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of theoretical biology* **71**:587-603

[32] **Rigney David R** (1979) Stochastic model of constitutive protein levels in growing and dividing bacterial cells. *Journal of Theoretical Biology* **76**:453-480

Peer reviews

Reviewer #1 (Public review):

Summary:

The aim of this paper is to develop a simple method to quantify fluctuations in the partitioning of cellular elements. In particular, they propose a flow-cytometry-based method coupled with a simple mathematical theory as an alternative to conventional imaging-based approaches.

Strengths:

The approach they develop is simple to understand and its use with flow-cytometry measurements is clearly explained. Understanding how the fluctuations in the cytoplasm partition vary for different kinds of cells is particularly interesting.

Weaknesses:

The theory only considers fluctuations due to cellular division events. This seems a large weakness because it is well known that fluctuations in cellular components are largely affected by various intrinsic and extrinsic sources of noise and only under particular conditions does partitioning noise become the dominant source of noise.

https://doi.org/10.7554/eLife.104528.1.sa2

Reviewer #2 (Public review):

Summary:

The authors present a combined experimental and theoretical workflow to study partitioning noise arising during cell division. Such quantifications usually require time-lapse experiments, which are limited in throughput. To bypass these limitations, the authors propose to use flow-cytometry measurements instead and analyse them using a theoretical model of partitioning noise. The problem considered by the authors is relevant and the idea to use statistical models in combination with flow cytometry to boost statistical power is elegant. The authors demonstrate their approach using experimental flow cytometry measurements and validate their results using time-lapse microscopy. However, while I appreciate the overall goal and motivation of this work, I was not entirely convinced by the strength of this contribution. The approach focuses on a quite specific case, where the dynamics of the labelled component depend purely on partitioning. As such it seems incompatible with studying the partitioning noise of endogenous components that exhibit production/turnover. The description of the methods was partly hard to follow and should be improved. In addition, I have several technical comments, which I hope will be helpful to the authors.

Comments:

(1) In the theoretical model, copy numbers are considered to be conserved across generations. As a consequence, concentrations will decrease over generations due to dilution. While this consideration seems plausible for the considered experimental system, it seems incompatible with components that exhibit production and turnover dynamics. I am therefore wondering about the applicability/scope of the presented approach and to what extent it can be used to study partitioning noise for endogenous components. As presented, the approach seems to be limited to a fairly small class of experiments/situations.

(2) Similar to the previous comment, I am wondering what would happen in situations where the generations could not be as clearly identified as in the presented experimental system (e.g., due to variability in cell-cycle length/stage). In this case, it seems to be challenging to identify generations using a Gaussian Mixture Model. Can the authors comment on how to deal with such situations? In the abstract, the authors motivate their work by arguing that detecting cell divisions from microscopy is difficult, but doesn't their flow cytometry-based approach have a similar problem?

(3) I could not find any formal definition of division asymmetry. Since this is the most important quantity of this paper, it should be defined clearly.

(4) The description of the model is unclear/imprecise in several parts. For instance, it seems to me that the index "i" does not really refer to a cell in the population, but rather a subpopulation of cells that has undergone a certain number of divisions. Furthermore, why is the argument of Equation 11 suddenly the fraction f as opposed to the component number? I strongly recommend carefully rewriting and streamlining the model description and clearly defining all quantities and how they relate to each other.

(5) Similarly, I was not able to follow the logic of Section D. I recommend carefully rewriting this section to make the rationale, logic, and conclusions clear to the reader.

(6) Much theoretical work has been done recently to couple cell-cycle variability to intracellular dynamics. While the authors neglect the latter for simplicity, it would be important to further discuss these approaches and why their simplified model is suitable for their particular experiments.

(7) In the discussion the authors note that the microscopy-based estimates may lead to an overestimation of the fluctuations due to limited statistics. I could not follow that reasoning. Due to the gating in the flow cytometry measurements, I could imagine that the resulting populations are more stringently selected as compared to microscopy. Could that also be an explanation? More generally, it would be interesting to see how robust the results are in terms of different gating diameters.

(8) It would be helpful to show flow cytometry plots including the identified subpopulations for all cell lines, currently, they are shown only for HCT116 cells. More generally, very little raw data is shown.

(9) The title of the manuscript could be tailored more to the considered problem. At the moment it is very generic.

https://doi.org/10.7554/eLife.104528.1.sa1

Author response:

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The aim of this paper is to develop a simple method to quantify fluctuations in the partitioning of cellular elements. In particular, they propose a flow-cytometry-based method coupled with a simple mathematical theory as an alternative to conventional imaging-based approaches.*
>
> *Strengths:*

> *The approach they develop is simple to understand and its use with flow-cytometry measurements is clearly explained. Understanding how the fluctuations in the cytoplasm partition vary for different kinds of cells is particularly interesting.*
>
> *Weaknesses:*
>
> *The theory only considers fluctuations due to cellular division events. This seems a large weakness because it is well known that fluctuations in cellular components are largely affected by various intrinsic and extrinsic sources of noise and only under particular conditions does partitioning noise become the dominant source of noise.*

We thank the Reviewer for her/his evaluation of our manuscript. The point raised is indeed a crucial one. In a cell division cycle, there are at least three distinct sources of noise that affect component numbers [1] :

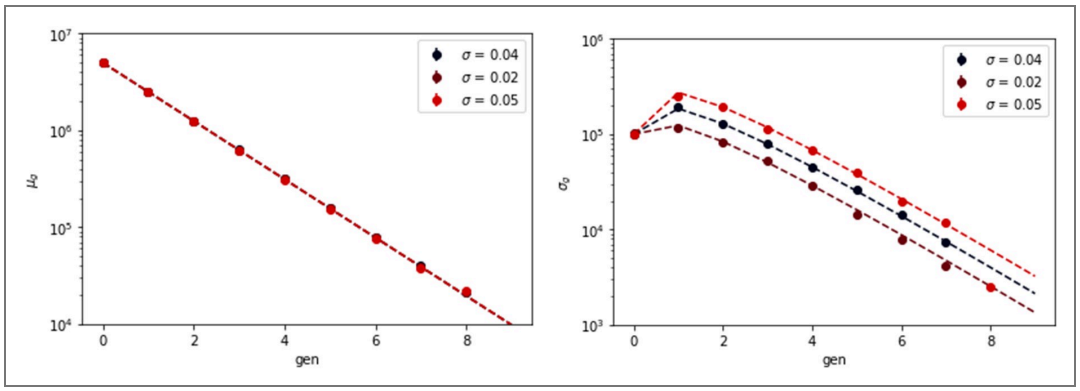
(1) Gene expression and degradation, which determine component numbers fluctuations during cell growth.

(2) Variability in cell division time, which depending on the underlying model may or may not be a function of protein level and gene expression.

(3) Noise in the partitioning/inheritance of components between mother and daughter cells.

Our approach specifically addresses the latter, with the goal of providing a quantitative measure of this noise source. For this reason, in the present work, we consider homogeneous cancer cell populations that could be considered to be stationary from a population point-of-view. By tracking the time evolution of the distribution of tagged components via live fluorescent markers, we aim at isolating partitioning noise effects. However, as noted by the Reviewer, other sources of noise are present, and depending on the considered system the relative contributions of the different sources may change. Thus, we agree that a quantification of the effect of the various noise sources on the accuracy of our measurements will improve the reliability of our method.

In this respect, assuming independence between noise sources, we reasoned that variability in cell cycle length would affect the timing of population emergence but not the intrinsic properties of those populations (e.g., Gaussian variance). To test this hypothesis, we conducted a preliminary set of simulations in which cell division times were drawn from an Erlang distribution (mean = 18 h, k=4k = 4k=4). The results, showing the behavior of the mean and variance of the component distributions across generations, are presented in Author response image 1. Under the assumption of independence between different noise sources, no significant effects were observed. Next, we plan to quantify the accuracy of our measurements in the presence of cross-talks between the various noise sources. As suggested, we will update the manuscript to include a more complete discussion on this topic and an evaluation of our model's stability.



Author response image 1. Variance and mean of the distribution of fluorescence intensity as a function of the generation for a time course dynamic with cell-cycle length variability. We repeated the same simulations as the one in figure 1 of the manuscript, but introducing a variable division time for each cell. The division time of each cell is extracted from an Erlang distribution (mean = 18 h and k = 4). As it is possible to observe in the plots, the results of our theoretical framework are not affected from the introduction of this variability. Hence, the Gaussian Mixture Model is still able to give the correct results even in a noisy environment.

(1) Soltani, Mohammad, et al. "Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes." *PLoS computational biology* 12.8 (2016): e1004972.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *The authors present a combined experimental and theoretical workflow to study partitioning noise arising during cell division. Such quantifications usually require time-lapse experiments, which are limited in throughput. To bypass these limitations, the authors propose to use flow-cytometry measurements instead and analyse them using a theoretical model of partitioning noise. The problem considered by the authors is relevant and the idea to use statistical models in combination with flow cytometry to boost statistical power is elegant. The authors demonstrate their approach using experimental flow cytometry measurements and validate their results using time-lapse microscopy. However, while I appreciate the overall goal and motivation of this work, I was not entirely convinced by the strength of this contribution. The approach focuses on a quite specific case, where the dynamics of the labelled component depend purely on partitioning. As such it seems incompatible with studying the partitioning noise of endogenous components that exhibit production/turnover. The description of the methods was partly hard to follow and should be improved. In addition, I have several technical comments, which I hope will be helpful to the authors.*

We are grateful to the Reviewer for her/his comments. Indeed, both partitioning and production turnover noise are in general fundamental processes. At present the only way to consider them together are time-consuming and costly transfection/microscopy/tracking experiments. In this work, we aimed at developing a method to effectively pinpoint the first component, i.e. partitioning noise thus we opted to separate the two different noise sources.

Below, we provide a point-by-point response that we hope will clarify all raised concerns.

> *Comments:*
>
> *(1) In the theoretical model, copy numbers are considered to be conserved across generations. As a consequence, concentrations will decrease over generations due to dilution. While this consideration seems plausible for the considered experimental system, it seems incompatible with components that exhibit production and turnover*

> *dynamics. I am therefore wondering about the applicability/scope of the presented approach and to what extent it can be used to study partitioning noise for endogenous components. As presented, the approach seems to be limited to a fairly small class of experiments/situations.*

We see the Reviewer's point. Indeed, we are proposing a high-throughput and robust procedure to measure the partitioning/inheritance noise of cell components through flow cytometry time courses. By using live-cell staining of cellular compounds, we can track the effect of partitioning noise on fluorescence intensity distribution across successive generations. This specific procedure is purposely optimized to isolate partitioning noise from other sources and, as it is, can not track endogenous components or dyes that require fixation. While this certainly poses limits to the proposed approach, there are numerous contexts in which our methodology could be used to explore the role of asymmetric inheritance. Among others, (i) investigating how specific organelles are differentially partitioned and how this influences cellular behavior could provide deeper insights into fundamental biological processes: asymmetric segregation of organelles is a key factor in cell differentiation, aging, and stress response. During cell division, organelles such as mitochondria, the endoplasmic reticulum, lysosomes, peroxisomes, and centrosomes can be unequally distributed between daughter cells, leading to functional differences that influence their fate. For instance, Kajaitso et al. [1] proposed that asymmetric division of mitochondria in stem cells is associated with the retention of stemness traits in one daughter cell and differentiation in the other. As organisms age, stem cells accumulate damage, and to prevent exhaustion and compromised tissue function, cells may use asymmetric inheritance to segregate older or damaged subcellular components into one daughter cell. (ii) Asymmetric division has also been linked to therapeutic resistance in Cancer Stem Cells [2]. Although the functional consequences are not yet fully determined, the asymmetric inheritance of mitochondria is recognized as playing a pivotal role [3]. Another potential application of our methodology may be (iii) the inheritance of lysosomes, which, together with mitochondria, appears to play a crucial role in determining the fate of human blood stem cells [4]. Furthermore, similar to studies conducted on liquid tumors [5][6], our approach could be extended to investigate cell growth dynamics and the origins of cell size homeostasis in adherent cells [7][8][9]. The aforementioned cases of study can be readily addressed using our approach that in general is applicable whenever live-cell dyes can be used. We will add a discussion of the strengths and limitations of the method in the Discussion section of the revised version of the manuscript.

(1) Katajisto, Pekka, et al. "Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness." Science 348.6232 (2015): 340-343.

(2) Hitomi, Masahiro, et al. "Asymmetric cell division promotes therapeutic resistance in glioblastoma stem cells." JCI insight 6.3 (2021): e130510.

(3) García-Heredia, José Manuel, and Amancio Carnero. "Role of mitochondria in cancer stem cell resistance." Cells 9.7 (2020): 1693.

(4) Loeffler, Dirk, et al. "Asymmetric organelle inheritance predicts human blood stem cell fate." Blood, The Journal of the American Society of Hematology 139.13 (2022): 2011-2023.

(5) Miotto, Mattia, et al. "Determining cancer cells division strategy." arXiv preprint arXiv:2306.10905 (2023).

(6) Miotto, Mattia, et al. "A size-dependent division strategy accounts for leukemia cell size heterogeneity." Communications Physics 7.1 (2024): 248.

(7) Kussell, Edo, and Stanislas Leibler. "Phenotypic diversity, population growth, and information in fluctuating environments." Science 309.5743 (2005): 2075-2078.

(8) McGranahan, Nicholas, and Charles Swanton. "Clonal heterogeneity and tumor evolution: past, present, and the future." Cell 168.4 (2017): 613-628.

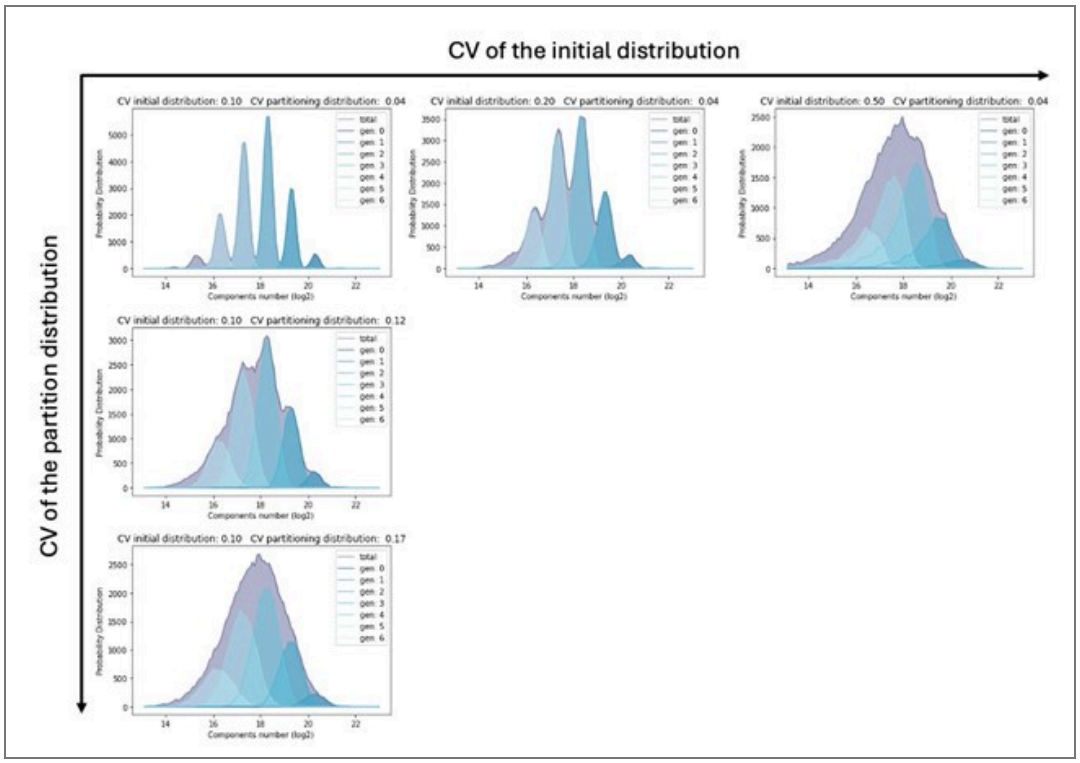
(9) De Martino, Andrea, Thomas Gueudré, and Mattia Miotto. "Exploration-exploitation tradeoffs dictate the optimal distributions of phenotypes for populations subject to fitness fluctuations." Physical Review E 99.1 (2019): 012417.

> *(2) Similar to the previous comment, I am wondering what would happen in situations where the generations could not be as clearly identified as in the presented experimental system (e.g., due to variability in cell-cycle length/stage). In this case, it seems to be challenging to identify generations using a Gaussian Mixture Model. Can the authors comment on how to deal with such situations? In the abstract, the authors motivate their work by arguing that detecting cell divisions from microscopy is difficult, but doesn't their flow cytometry-based approach have a similar problem?*

The point raised is an important one, as it highlights the fundamental role of the gating strategy. The ability to identify the distribution of different generations using the Gaussian Mixture Model (GMM) strongly depends on the degree of overlap between distributions. The more the distributions overlap, the less capable we are of accurately separating them.

The extent of overlap is influenced by the coefficients of variation (CV) of both the partitioning distribution function and the initial component distribution. Specifically, the component distribution at time t results from the convolution of the component distribution itself at time t−1 and the partitioning distribution function. Therefore, starting with a narrow initial component distribution allows for better separation of the generation peaks. The balance between partitioning asymmetry and the width of the initial component distribution is thus crucial.

As shown in Author response image 2, increasing the CV of either distribution reduces the ability to distinguish between different generations.



Author response image 2. Components distribution at varying CVs of initial components and partitioning distributions. Starting from a condition in which both division asymmetry and wideness of the initial components distribution are low and different generations are clearly separable, increasing either the CVs leads to distribution mixing and greater reconstruction difficulty.

However, the variance of the initial distribution cannot be reduced arbitrarily. While selecting a narrow distribution facilitates a better reconstruction of the distributions, it simultaneously limits the number of cells available for the experiment. Therefore, for components exhibiting a high level of asymmetry, further narrowing of the initial distribution becomes experimentally impractical.

In such cases, an approach previously tested on liquid tumors [1] involves applying the Gaussian Mixture Model (GMM) in two dimensions by co-staining another cellular component with lower division asymmetry.

Regarding time-lapse fluorescence microscopy, the main challenge lies not in disentangling the interplay of different noise sources, but rather in obtaining sufficient statistical power from experimental data. While microscopy provides detailed insights into the division process and component partitioning, its low throughput limits large-scale statistical analyses. Current segmentation algorithms still perform poorly in crowded environments and with complex cell shapes, requiring a substantial portion of the image analysis pipeline to be performed manually, a process that is time-consuming and difficult to scale. In contrast, our cytometry-based approach bypasses this analysis bottleneck, as it enables a direct population-wide measurement of the system's evolution. We will provide a detailed discussion on these aspects in the revised version of the manuscript.

(1) Peruzzi, Giovanna, et al. "Asymmetric binomial statistics explains organelle partitioning variance in cancer cell proliferation." Communications Physics 4.1 (2021): 188.

> *(3) I could not find any formal definition of division asymmetry. Since this is the most important quantity of this paper, it should be defined clearly.*

We thank the Reviewer for the note. With division asymmetry we refer to a quantity that reflects how similar two daughter cells are likely to be in terms of inherited components after a division process. We opted to measure it via the coefficient of variation (root squared variance divided by the mean) of the partitioning fraction distribution. We will amend this lack of definition in the reviewed version of the manuscript.

> *(4) The description of the model is unclear/imprecise in several parts. For instance, it seems to me that the index "i" does not really refer to a cell in the population, but rather a subpopulation of cells that has undergone a certain number of divisions. Furthermore, why is the argument of Equation 11 suddenly the fraction f as opposed to the component number? I strongly recommend carefully rewriting and streamlining the model description and clearly defining all quantities and how they relate to each other.*

We are amending the text carefully to avoid double naming of variables and clarifying each computation passage. In equation 11 the variable f refers to the fluorescent intensity, but the notation will be changed to increase clarity.

> *(5) Similarly, I was not able to follow the logic of Section D. I recommend carefully rewriting this section to make the rationale, logic, and conclusions clear to the reader.*

We will update the manuscript clarifying the scope of section D and its results. In brief, Section A presents a general model to derive the variance of the partitioning distribution from flow cytometry time-course data without making any assumptions about the shape of the distribution itself. In Section D, our goal is to interpret the origin of asymmetry and propose a possible form for the partitioning distribution. Since the dyes used bind non-specifically to cytoplasmic amines, the tagged proteins are expected to be uniformly distributed throughout the cytoplasm and present in large numbers. Given these assumptions the least complex model for division follows the binomial distribution, with a parameter that measures the bias in the process. Therefore, we performed a similar computation to that in Section A, which allows us to estimate not only the variance but also the degree of biased asymmetry. Finally, we fitted the data to this new model and proposed an experimental interpretation of the results.

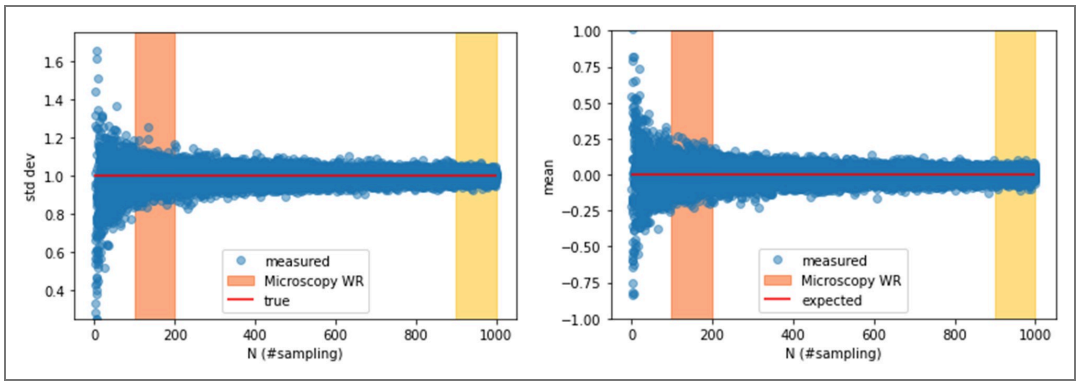
> *(6) Much theoretical work has been done recently to couple cell-cycle variability to intracellular dynamics. While the authors neglect the latter for simplicity, it would be important to further discuss these approaches and why their simplified model is suitable for their particular experiments.*

We agree with the Reviewer, we will discuss this aspect in the revised version of the manuscript.

> *(7) In the discussion the authors note that the microscopy-based estimates may lead to an overestimation of the fluctuations due to limited statistics. I could not follow that reasoning. Due to the gating in the flow cytometry measurements, I could imagine that the resulting populations are more stringently selected as compared to microscopy. Could that also be an explanation? More generally, it would be interesting to see how robust the results are in terms of different gating diameters.*

The Reviewer is right on the importance of the sorting procedure. As already discussed in a previous point, the gating strategy we employed plays a fundamental role: it reduces the overlap of fluorescence distributions as generations progress, enables the selection of an initial distribution distinct from the fluorescence background, allowing for longer tracking of proliferation, and synchronizes the initial population. The narrower the initial distribution, the more separated the peaks of different generations will be. However, this also results in a smaller number of cells available for the experiment, requiring a careful balance between precision and experimental feasibility. A similar procedure, although it would certainly limit

the estimation error, would be impracticable In the case of microscopy. Indeed, the primary limitation and source of error is the number of recorded events. Our pipeline allowed us to track on the order of hundreds of division dynamics, but the analysis time scales non-linearly with the number of events. Significantly increasing the dataset would have been extremely time-consuming. Reducing the analysis to cells with similar fluorescence, although theoretically true, would have reduced the statistics to a level where the sampling error would drastically dominate the measure. Moreover, different experiments would have been hardly comparable, since different fluorescences could map in equally sized cells. In light of these factors, we expect higher CV for the microscopy measure than for flow cytometry's ones. In the plots below, we show the behaviour of the mean and the standard deviation of N numbers sampled from a gaussian distribution N(0,1) as a function of the sampling number N. The higher is N the closer the sampled distribution will be to the true one. The region in the hundreds of samples is still very noisy, but to do much better we would have to reach the order of thousands. We will add a discussion on these aspects in the reviewed version of the manuscript.



Author response image 3. Standard deviation and mean value of a distribution of points sampled from a Gaussian distribution with mean 0 and standard deviation 1, versus the number of samples, N. Increasing N leads to a closer approximation of the expected values. In orange is highlighted the Microscopy Working Region (Microscopy WR) which corresponds to the number of samples we are able to reach with microscopy experiments. In yellow the region we would have to reach to lower the estimating error, which is although very expensive in terms of analysis time.

> *(8) It would be helpful to show flow cytometry plots including the identified subpopulations for all cell lines, currently, they are shown only for HCT116 cells. More generally, very little raw data is shown.*

We will provide the requested plots for the other cell lines together with additional raw data coming from simulations in the Supplementary Material.

> *(9) The title of the manuscript could be tailored more to the considered problem. At the moment it is very generic.*

We see the Reviewer point. The proposed title aims at conveying the wide applicability of the presented approach, which ultimately allows for the assessment of the levels of fluctuations in the levels of the cellular components at division. This in turn reflects the asymmetricity in the division.

https://doi.org/10.7554/eLife.104528.1.sa0