Heritable tumor cell division rate heterogeneity induces clonal dominance

Margriet M. Palm[a], Marjet Elemans[a] and Joost B. Beltman[a]

[a]Division of Drug Discovery and Safety, Leiden Academic Center for Drug Research, Leiden University, The Netherlands

November 21, 2017

Abstract

Tumors consist of a hierarchical population of cells that differ in their phenotype and genotype. This hierarchical organization of cells means that a few clones (i.e., cells and several generations of offspring) are abundant while most are rare, which is called *clonal dominance*. Such dominance also occurred in published *in vitro* iterated growth and passage experiments with tumor cells in which genetic barcodes were used for lineage tracing. A potential source for such heterogeneity is that dominant clones derive from cancer stem cells with an unlimited self-renewal capacity. Furthermore, ongoing evolution within the growing population may also induce clonal dominance. To understand how clonal dominance developed in the iterated growth and passage experiments, we built a computational model that accurately simulates these experiments. The model simulations reproduced the clonal dominance that developed in *in vitro* iterated growth and passage experiments when the division rates vary between cells, due to a combination of initial variation and of ongoing mutational processes. In contrast, the experimental results can neither be reproduced with a model that considers random growth and passage, nor with a model based on cancer stem cells. Altogether, our model suggests that *in vitro* clonal dominance develops due to selection of fast-dividing clones.

Introduction

Intratumoral heterogeneity, the genotypic and phenotypic differences within a single tumor, is a well known feature of cancer [1] and strongly influences the effectiveness of cancer therapy [2]. Genotypic heterogeneity is the result of random mutations, and while most of these mutations are neutral passenger mutations, some are functional mutations that add to phenotypic heterogeneity. Phenotypic differences may also be caused by phenomena such as differential signaling from the local tumor micro-environment, epigenetic changes, and stochastic gene expression [3]. Another proposed source of intratumoral, phenotypic heterogeneity is the presence of so-called *cancer stem cells* (CSCs) that have an unlimited potential to renew and can give rise to *differentiated cells* (DCs) with a limited potential to renew [4]. The presence of CSCs would result in a tumor containing a mixture of CSCs, and DCs that all derive from a small number of CSCs.

For a long time, evidence for the presence of CSCs was primarily based on xenograft models in which transplantation of tumor cells into immunodeficient mice resulted in tumor growth in

only a small fraction of the mice [1, 5], suggesting that only a subset of the tumor cells has the ability to sustain long-term growth. However, the lack of success of initiating tumor growth in immunodeficient mice may also be related to the incomplete inhibition of the immune response [6], or to the dramatic change in tumor micro-environment upon transplantation [5]. An alternative approach to identify the existence of CSCs is to perform *lineage tracing* by fluorescent marking of a subpopulation of cells [7, 8]. For example, Schepers *et al.* [9] managed to trace the lineage of CSCs by fluorescently labeling cells expressing stem cell markers in mice developing intestinal adenomas and thereby showed that all cells in small adenomas descended from a single stem cell. However, fluorescent labeling of stem cells is not possible in all cancer types, and for those cancer types an alternative approach is taken by labeling a small fraction of the tumor cells in animal models. Studies employing this strategy showed that the number of colored patches reduced during tumor growth, while the size of these patches increased [10–12]. These observations are compatible with the hypothesis that tumor cells descend from a small number of CSCs.

Another, high-resolution, approach to lineage tracing is the application of unique genetic tags, also called *cellular barcodes*, to a population of tumor cells [13–19]. Tumors grown in immunodeficient mice injected with barcoded tumor cells are dominated by cells that express only a small subset of the barcodes [15, 16]. Serial implantation of barcoded leukemic cells showed that rare clones in one mouse can develop dominance after transplantation into a second mouse, indicating that clonal dominance is not a predetermined property of certain clones [18]. Porter *et al.* [13] used cellular barcoding to follow the development of clonal dominance over time in an *in vitro* setup, thereby controlling the external factors that could affect clonal dominance. Populations derived from several polyclonal cell lines were barcoded and grown for three days after which a fraction of the cells was passed on to the next generation. By repeating this process (Fig 1**A**) and analyzing the intermediate clone distributions, Porter *et al.* [13] showed that clonal dominance progressed over time (Fig 1**B**). Interestingly, similar experiments with a monoclonal K562 cell line resulted in a minimal progression of clonal dominance, hinting that an intrinsic variability in the cell population may cause the progression of clonal dominance. However, Nolan-Stevaux *et al.* [15] did not observe a strong development of clonal dominance during 8 days of *in vitro* growth without passaging, indicating that random passage could also play a role in developing clonal dominance.

Although the studies discussed above represent a strong base of evidence for the development of clonal dominance *in vivo* as well as in *in vitro* tumor cell populations, the mechanism that drives this dominance remains unknown. The presence of CSCs is consistent with the induction of clonal dominance, but only in some cancers direct evidence for the involvement of CSCs is available [5]. Alternatively, evolution may cause clones with a higher division or survival rate to dominate the tumor. Hence, it is necessary to further investigate the role of these mechanism in the development of clonal dominance. One way to do this in a formal way is to construct computational models that incorporate different hypotheses and compare the outcome of computer simulations to spatio-temporal clonal dynamics observed experimentally. Such an approach has been used before in numerous studies addressing the temporal and spatial evolution of tumor cell populations with CSCs, which are thoroughly reviewed in [20]. Several of these modeling studies focused on the development of spatial tumor heterogeneity by employing cell-based models in which intratumoral heterogeneity is induced by CSCs [21, 22], by CSCs and epigenetic changes [23], or by mutations [24]. Here, we built a computational model that simulates the iterated growth and

passage experiments described in [13]. By incorporating different hypotheses for the development of clonal dominance in our model, we show that heritable variability of the rates of cell divisions amongst tumor cells, which is due to a combination of initial variation and mutational processes, is sufficient to induce clonal dominance in iterated growth and passage experiments.

Results

Re-analysis of published sequencing data

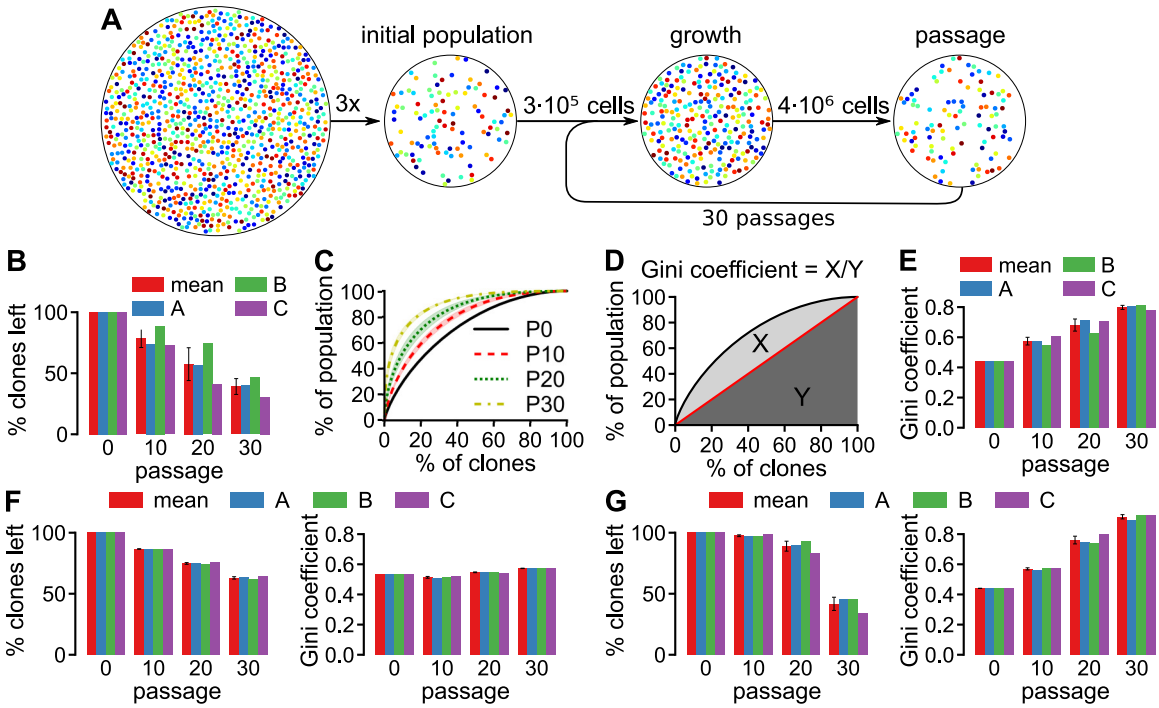


Figure 1: Setup and results of the in vitro iterated growth and passage experiment previously described by Porter et al. [13] A Experimental setup. **B-E** Development of the clone size distribution of polyclonal K562 cells, as obtained from our own analysis of the FASTQ files published by Porter *et al.* [13]. Shown are the percentage of clones that remain after each passage (**B**), the percentage of clones versus the percentage of the population taken up by those clones (**C**, mean ± SD and 3 biological replicates shown) and the Gini coefficient (**E**; ratio of areas X and Y in **D**). **F** Clone loss (left) and Gini coefficient (right) for the *in vitro* experiments with the monoclonal K562 cell line. **G** Clone loss (left) and Gini coefficient (right) for the *in vitro* experiments with HeLa cells. All error bars depict the SD of the 3 replicates.

We started by re-analyzing the data for the *in vitro* experiments with the lentivirally barcoded cell lines, in order to subsequently make quantitative comparisons with simulation results (see Methods section for more details). In these previously published experiments, the barcode-transduced cells were grown and aliquots containing $3 \cdot 10^5$ cells were used to initialize three biological replicates of the iterated growth and passage experiment. The cells were then, iteratively, grown for 3 days

after which $3 \cdot 10^5$ cells were passed on to the next generation (Fig 1**A**).

In agreement with Porter *et al.* [13], iterated growth and passage with the polyclonal K562 cell line resulted in clone loss (Fig 1**B**) and progressing clonal dominance (Fig 1**C**). To evaluate clonal dominance, we plotted clone size, sorted from large to small, against the cumulative population fraction for clones of that size (Fig 1**C**). In the remainder of this paper, we quantify clonal dominance based on this graph (1**D**) by employing the *Gini coefficient* [25]:

$$G = \frac{1}{2C_0} \sum_{i=1}^{C_0} \sum_{j=1}^{C_0} \frac{|N_i - N_j|}{n}$$

with C_0 the number of clones at initialization, N_i the size of clone i, and n the number of cells. As expected from the observed clone distributions, the Gini coefficient is already greater than zero at the start of the experiment (indicating mild clonal dominance) and further increases over time (Fig 1**E**). We also processed the data for the *in vitro* experiments with monoclonal K562 cells (Fig 1**F**) and with HeLa cells (Fig 1**G**) and clone loss was in good agreement with those published in [13] (S1 Fig). The monoclonal K562 cell line exhibits a more limited clone loss and development of clonal dominance (at time points beyond the initial measurement) compared to the polyclonal K562 cell line. The HeLa cells show a similar trend as the K562 cells: the number of unique clones declines while clonal dominance increases. However, clone loss occurred later for the HeLa cells than it did for the polyclonal K562 cells.

Iterated growth and passage reduces the number of clones, but does not cause progressive clonal dominance

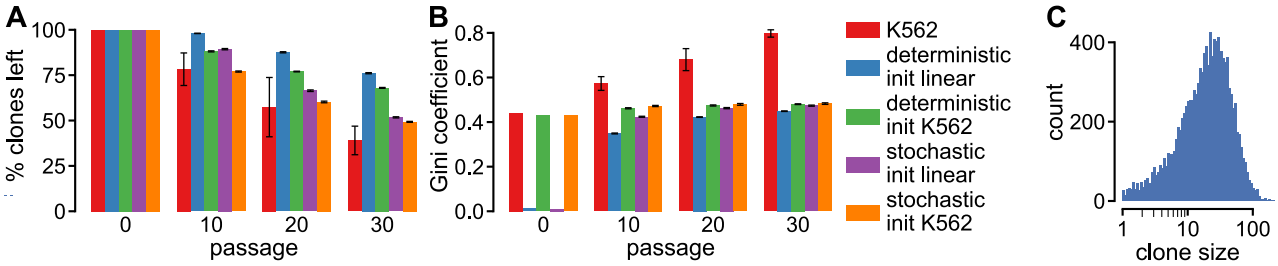


Figure 2: Clonal dominance does not develop during passaging of cells that divide at a fixed rate. A-B Clone loss (**A**) and Gini coefficient (**B**) during iterated growth and passage with either deterministic or stochastic growth and initialized either with a uniform clone size distribution or the initial distribution for polyclonal K562 cells. All values are the mean of 10 simulations and the error bars represent the SD. **C** histogram of the initial clone sizes of polyclonal K562 cells.

The limited development of clonal dominance during passaging in monoclonal K562 cells compared to polyclonal K562 cells indicates that chance could play a role in the development of clonal dominance, hence the most straightforward explanation for the development of clonal dominance is that the iterated passages cause small clones to completely disappear while larger clones remain and grow. This hypothesis is supported by Porter's *in vivo* experiments in which no passage occurred and no loss of clones nor clonal dominance was observed [13]. To test whether clone

loss during passage explains clone loss and progressive clonal dominance, Porter *et al.* employed a computational model of iterated deterministic growth phases and passage steps in which a subset of the cells were selected at random, showing that this could partially explain the experimentally observed clone loss but not the development of clonal dominance (see Methods section of [13]). In these simulations, all clones grow according to $N_i(t+t_{\text{growth}}) = N_i(t)e^{rt_{\text{growth}}}$, where $N_i(t)$ represents the size of clone i at time t, t_{growth} is the duration of the growth phase, and r is the division rate of the cells. At the end of the passage interval, n_{pass} cells are selected randomly and passed to the next generation (Fig 1**A**). As a first step, we confirmed the Porter simulation results using the code that was published with Porter's paper [13] (`https://github.com/adaptivegenome/clusterseq/`). We ran simulations starting with C=14,000 clones uniformly distributed over $3 \cdot 10^5$ cells and $r = \frac{24}{19}\log(2)$ to match the experimentally observed doubling time of 19 hours. As expected, these simulations resulted in a moderate reduction of the number of clones (Fig 2**A**; blue bars). In contrast to the *in vitro* results, clonal dominance developed only slightly and hardly increased over time beyond passage 10 (Fig 2**B**; blue bars). To test if a realistic initial clone distribution improved the resemblance between the experimental observations and the simulation results, we extended the simulations by employing the initial clone distribution of the polyclonal K562 cell line (Fig 2**C**) in which clonal dominance is already slightly developed. In this setting the expected number of clones left still remained larger than that for the polyclonal K562 cells while the clone distribution was hardly affected (Fig 2**A**-**B**; green bars).

Porter *et al.* [13] noted that in simulations in which more cells were passaged, fewer clones disappeared, indicating that the random process of passage causes some clones to become smaller and finally disappear. However, because cell division is modeled as a deterministic process, the clone sizes, relative to other clones, remain similar over time, and the probability to disappear for individual clones remains unchanged over time. The assumption of deterministic growth would be reasonable for large clones, but for small clones, probabilistic events could strongly affect the simulation outcome. As shown in Fig 2**C**, a large part of the clones contain only a few cells, e.g., $\sim 25\%$ of the clones have less than 10 cells. Therefore, we next explored the effect of stochastic cell division by replacing deterministic growth by a stochastic growth model using Gillespie's Stochastic Simulation Algorithm (SSA) [26] (see Methods section for details and Table 1 for the model parameters). In contrast with Porter's deterministic growth model, passage occurs when the population size reaches the critical population size $n_{\text{crit}} = 4 \cdot 10^6$, which corresponds to approximately 3 days of growth with a population doubling time of 19 hours. Because growth continues until the critical population size is reached, the division rate no longer affects the simulation results and we therefore set it to the arbitrary value of 1. As before, n_{pass} cells are then chosen randomly and passed on to the next generation.

We performed stochastic growth simulations initialized with either a uniform clone distribution (Fig 2**A**-**B**; purple bars) or the initially observed clone distribution of polyclonal K562 cells (orange bars). In both cases the results were similar to those with deterministic growth, albeit with a larger decrease in the number of clones. For simulations with stochastic growth that were initialized with the polyclonal K562 clone distribution, the clone loss resembles the average *in vitro* clone loss for passages 10 and 20, but for passage 30 the *in vitro* clone loss overtook the simulated clone loss. This suggests that the early clone loss is dominated by the effects of stochastic division and passage, and that at a later stage another, unknown, mechanism further increases clone loss. Altogether, these results indicate that passage and growth can cause clone loss during passage,

Table 1: Parameters of the stochastic growth and passage models.

parameter	symbol	default value	unit
Iterated growth & passage			
population size after passage	n_{pass}	$3 \cdot 10^5$	cells
maximum population size	n_{crit}	$4 \cdot 10^6$	cells
number of passages		30	
Tau leaping algorithm			
tau leap step	τ	.0005	days
Stochastic growth model			
division rate	r	1	day^{-1}
CSC growth model			
symmetric division probability	p_1	0.5	
asymmetric division probability	p_2	0.5	
symmetric commitment probability	p_3	0	
initial CSC fraction	CSC_0	5%	
maximum DC age	M	10	divisions
CSC division rate	r_{CSC}	1	day^{-1}
DC division rate	r_{DC}	$\frac{24}{19}$	day^{-1}
Agent-based model			
mean initial division rate	r_0	1	day^{-1}
initial division rate SD	σ_0^*	0	
mutation SD	σ_m^*	0	

but neither of these mechanisms can induce progressive clonal dominance.

The presence of CSCs does not induce clonal dominance

The presence of cancer stem cells is thought to induce tumor heterogeneity because they generate a, hierarchically organized, population of cancer stem cells and differentiated cells [27]. To introduce cancer stem cells in our growth model we replaced the unlimited growth in the stochastic model with a previously published model of cancer stem cell driven growth [28]. In this model cells are either cancer stem cells (CSCs) that can divide indefinitely, or differentiated cells (DCs) that divide a limited number of times. CSCs proliferate at a division rate of r_{CSC} and division can result either in two stem cells (with probability p_1), in a stem cell and a differentiated cell (with probability p_2), or in two differentiated cells (with probability p_3) (Fig 3**A**). Differentiated cells proliferate at a division rate r_{DC} until they reach their maximum number of divisions M, after which, following Weekes *et al.* [28], they die with a rate r_{DC}. We did not consider random cell death of cancer stem cells and differentiated cells, because this process only affects the population growth rate.

We fine-tuned the parameters of the CSC growth model such that the population growth rate is consistent with the 19 hour doubling time reported by Porter *et al.* [13]. For this we exploited the analytical solution of the CSC model elegantly derived by Weekes *et al.* [28]. This solution predicts that the population of cells initially grows, and then develops according to one of three growth regimes determined by $\beta = (p_1 - p_3)r_{\text{CSC}}$. When $\beta > 0$, the population continues to grow,

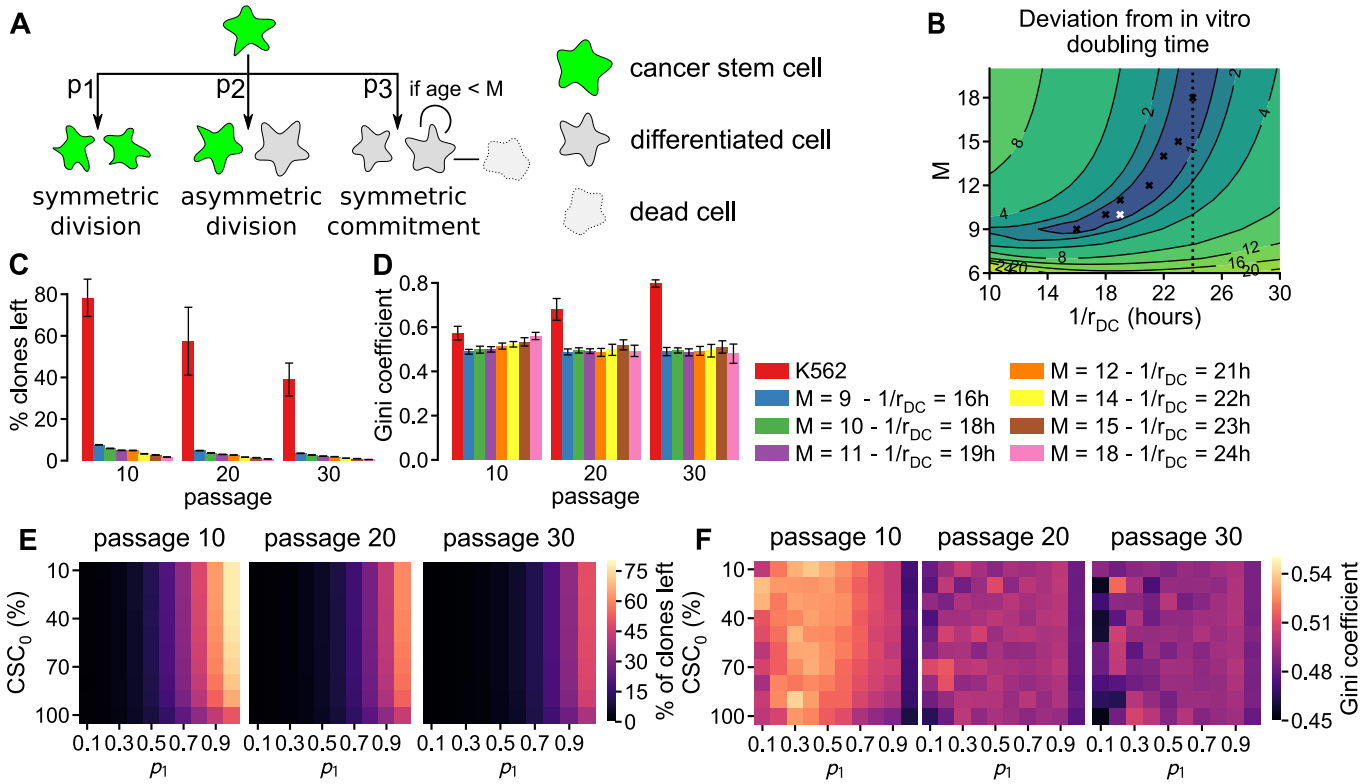


Figure 3: Simulations with CSC model result in massive clone loss and no development of clonal dominance. A Scheme illustrating the divisions and cell death in the CSC growth model. **B** Heatmap showing the difference between *in vitro* and simulated population doubling time (19 hours) depending on the maximum number of DC divisions (M) and DC division rates (r_{DC}) in the CSC growth model. The white cross denotes the default model settings and the black crosses depict several alternative parameter sets that result in a similar population doubling time. **C-D** Clone loss (**C**) and Gini coefficient (**D**) for the parameter sets highlighted in **B** and all other parameters as in Table 1. **E-F** Effect of symmetric CSC division probability (p_1) and initial CSC percentage (CSC_0) on clone loss (**E**) and Gini coefficient (**F**), with all other parameters as in Table 1. All values are the mean of 10 simulation replicates with the error bars depicting the SD.

when $\beta = 0$ the population reaches an equilibrium, and when $\beta < 0$ the population will eventually go extinct. Because the *in vitro* cells were reported to be in "*log phase growth*" [13], we limited the parameter space to $\beta > 0$ by setting p_3 to 0, which ensures a monotonically growing cell population. A complicating factor is that the parameters determining β do not affect the growth in the first couple of days. Instead, during this time interval, growth is determined by the division rate of DCs (r_{DC}) relative to that of CSCs (r_{DC}), and the maximum age of DCs (M). Therefore, we explored how much the simulated population doubling time deviates from the experimental doubling time of 19 hours when M and r_{DC} were varied while the other simulation parameters remained at their default value according to Table 1 (Fig 3**B**). This exploration resulted in a well-defined parameter range for which the *in vitro* doubling time could be reproduced *in silico*.

To illustrate the clonal development of the CSC growth model, we tested the model for several parameter sets from the region that matches with the experimentally observed population growth rate (Fig 3**B**; crosses). The simulations showed that the maximum age of DCs or the CSC division rate hardly affect clonal development (Fig 3**C**-**D**). Furthermore, an at first sight surprisingly strong reduction in clone number occurred in the simulations while the clone distribution remained virtually unchanged (Fig 3**C**-**D**). The dramatic reduction in clone number can readily be explained: clones frequently loose all CSCs during passage, which takes away the long-term self-renewal capacity of such clones. We tested this explanation by varying the initial percentage of CSCs and the probability of a symmetric CSC differentiation while keeping the other parameters constant (Table 1 and white cross in Fig 3**B**). In line with our explanation for the massive clone loss, increasing the probability of symmetric CSC division strongly reduced clone loss (Fig 3**E**). Increasing the initial percentage of stem cells had a less pronounced effect on clone loss (Fig 3**E**), which fits with the observation in [28] that the proportion of CSCs in the population will become constant in the long term and does not depend on the initial percentage of CSCs. Altogether, increasing the probability of a symmetric CSC to 1 and the initial CSC percentage to 100% did not lead to clone loss matching that observed experimentally. Moreover, changing these parameters had only little effect on clonal dominance (Fig 3**F**). The lack of progression of clonal dominance over time can be explained by a lack in difference between the clones: although in the long run all cells descend from only a few clones, there is no difference in the speed at which each clone generates offspring. In conclusion, incorporating CSC growth into our passaging simulation resulted in a poor match to the experimental observations because far too many clones disappeared and the distribution of the remaining clones did not exhibit clonal dominance.

A model with evolving division rates results in clone loss and increasing clonal dominance

Because neither simulations with random growth and passaging of tumor cells nor simulations incorporating CSC growth can explain the development of clonal dominance, we next considered intrinsic variability of growth characteristics as a source of clonal dominance development. To this purpose, we let go of the distinction between cell types and let all cells divide with their own division rate r_i, that is subject to mutations. Note that the term mutation is used here to describe any change in a cell's phenotype that can be inherited, which includes both genetic and epigenetic changes. Mutation of cellular properties requires the explicit representation of single cells, therefore we can no longer use Gillespie's SSA because clones rather than single cells are explicitly described in that type of simulation. Therefore, we implemented an agent based

model (ABM) in which each cell is represented explicitly, and both a barcode and division rate are associated to each cell.

In the ABM there are two sources for division rate variability. First, the division rate of cells at initialization may vary due to prior mutations. To implement this, the division rate of each cell is initialized to $r_i = X_i r_0$, with r_i the cell's division rate, r_0 the mean division rate of all cells and X_i taken from a normal distribution with mean 1 and SD σ_0^*. Second, mutations are included by setting the division rate of each offspring to $r_i = Y_i r_p$, with Y_i taken from a normal distribution with mean 1 and SD σ_m^*, and r_p the division rate of the parent cell. Further details on the implementation of the ABM can be found in the Methods.

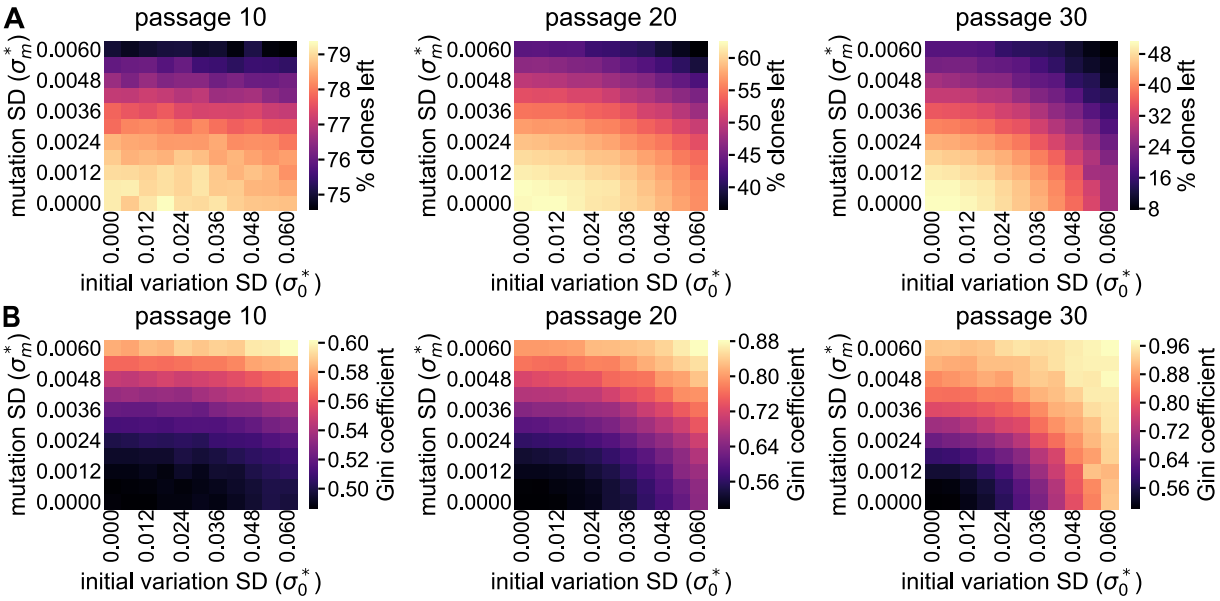


Figure 4: The ABM that describes evolution of division rate variability induces clone loss and clonal dominance A-B Clone loss (**A**) and Gini coefficient (**B**) for a range of initial division rate SDs (σ_0^*) and mutation SDs (σ_m^*), with all other parameters as in Table 1 and all data points representing the mean for 10 simulations.

To test if the ABM can induce similar clone loss (Fig 4**A**) and clonal dominance (Fig 4**B**) as observed *in vitro*, we ran simulations for a range of initial division rate SDs (σ_0^*) and mutation SDs (σ_m^*), all initialized using the initial distribution of the polyclonal K562 cells. Note that we did not vary the initial mean division rate (r_0) because this parameter does not affect the simulation results (S2 Fig). The resulting ranges of clone loss and Gini coefficient do indeed include those observed for the polyclonal K562 cells. Furthermore, simulations with division rates that vary due to either initial variation ($\sigma_0^* > 0$) or mutations ($\sigma_m^0 > 0$) result in higher levels of clone loss and a higher Gini coefficient compared to simulations without any variation ($\sigma_0^* = \sigma_m^* = 0$). Thus, the ABM that describes evolution of division rate variability has the potential to match the *in vitro* results.

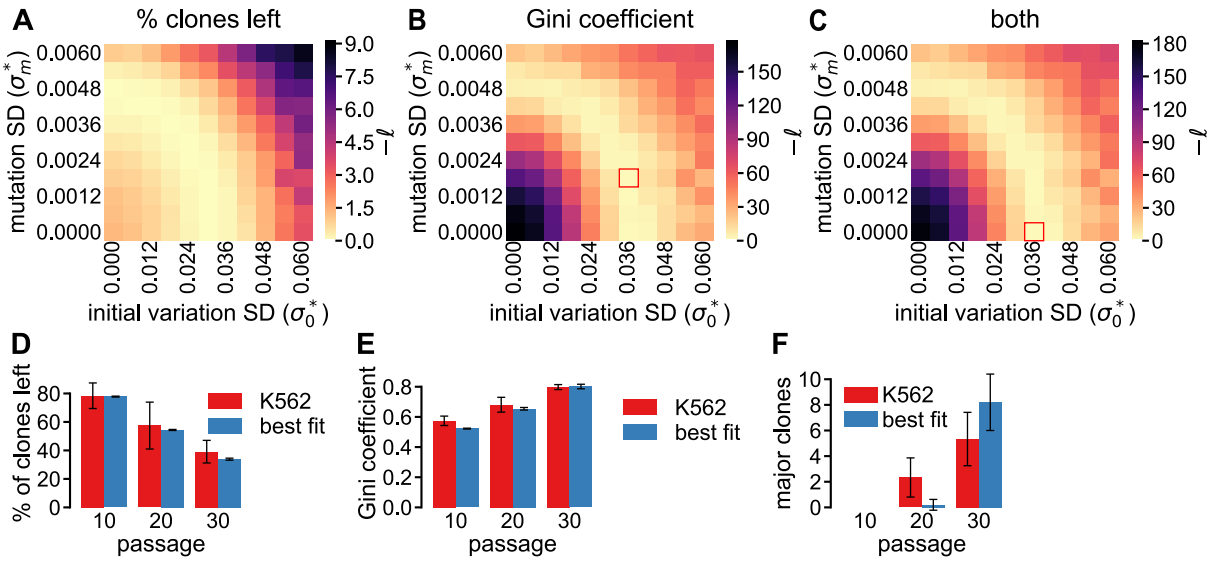


Figure 5: ABM simulations describing evolution of division rate variability match results for polyclonal K562 cell line. A-B Maximum Likelihood estimator (ℓ) based on the percentage of clones lost (**A**), on the Gini coefficient (**B**), and on both metrics (**C**) for a range of initial division rate SDs (σ_0^*) and mutation SDs (σ_m^*). Note that we plot $-\ell$ in these plots and thus its minimum value is sought. **D-E** comparison of clone loss (**D**) and clonal dominance (**E**) observed in simulations with the best fitting parameter values for the Gini coefficient (red rectangle in **B**) and in the experiments with polyclonal K562 cells. **F** Comparison of the number of major clones, i.e. clones representing more than 1% of the population, developing in simulations with the parameter set highlighted by the red rectangle in **B** and in the experiments with polyclonal K562 cells. All simulation results are the mean of 10 simulations and the results for the polyclonal K562 cells are the mean of 3 replicates, with the error bars representing the SD.

Fitting the ABM describing division rate evolution to the K562 data

To fine-tune the model for the polyclonal K562 cell line measurements, we perform a maximum likelihood estimation (MLE) for the simulations shown in Fig 4 to identify the best fitting values for σ_0^* and σ_m^*. For this we consider the errors of all metrics included in the MLE to be normally distributed, so we can use the following definition of the loglikelihood:

$$\ell(\sigma_0^*, \sigma_m^*) \sim \sum_{m \in \text{ metrics}} \sum_{P \in \text{passages}} -\frac{(\bar{m}(P) - \mu_{m,P})^2}{2\sigma_{m,P}^2}$$

with $\bar{m}(P)$ the mean of 10 simulations for metric m at passage P, and $\mu_{m,P}$ and $\sigma_{m,P}$ respectively the mean and SD of the three polyclonal K562 replicates for metric m at passage P. Using passages 10, 20, and 30, and either the fraction of clones left, the Gini coefficient, or both metrics, we obtained heatmaps of the likelihood (Fig 5**A**-**C**). The simulated clone loss is quite similar to that for the polyclonal K562 cells (Fig 5**A**), independent of σ_m^* and σ_0^*. This fits with the previous observation that a model with stochastic growth, but with identical division rates for all cells, already closely fits clone loss (Fig 4**A**). In contrast, the Gini coefficient only fits well for a defined region of parameter values (Fig 5**B**) and as a result this metric dominates the MLE based on both parameters (Fig 5**C**).

The area of best matching parameter sets also includes sets without mutation or initial division rate variation, suggesting that the source of the division rate variability cannot be determined. However, it seems most likely that both mechanisms are active because mutations before the experiment started would already induce initial division rate variation and these mutations continue to happen upon cell division. Consistent with this, the best fit corresponds to a model that includes both mechanisms ($\sigma_0^* = 0.036$ and $\sigma_m^* = 0.0018$) and the simulation results are close to the experimentally observed Gini coefficient (Fig 5**D**) and clone loss (Fig 5**E**). Furthermore, major clones, i.e., clones representing more than 1% of the population, do develop in the ABM simulations (Fig 5**E**).

To test the effect of initially present heterogeneity on clone loss and the development of clonal dominance, Porter *et al.* [13] created a monoclonal K562 cell line and showed that clone loss was reduced and clonal dominance only slightly developed (Fig. 1 F). Because this cell line was derived from a single cell, it is expected to have a lower initial variation yet similar mutation dynamics as other K562 cells. We tested this hypothesis by running a parameter sweep for σ_0^* and σ_m^* in which each simulation starts with the initial distribution of the monoclonal K562 cells.

Performing an MLE for either clone loss (Fig 6**A**) or the Gini coefficient (Fig 6**B**), we found that our model fails to produce a good fit for the clone loss (the maximum ℓ in Fig 6**A** is -780), while the fit for the Gini coefficient is much better (the maximum ℓ in Fig 6**B** is -40). In line with our expectations, the optimal fit of the Gini coefficient occurs at a σ_0^* that is clearly lower than the optimal fit for the polyclonal K562 data (cf. Fig 5**B**). Moreover, the best-fitting σ_m^* values for mono- and polyclonal K562 data are similar.

Although the optimal fitting parameters for the Gini coefficient show a close fit for the Gini coefficient (Fig 6**C**), this same parameter set produces a clone loss that follows a similar pattern as the monoclonal K562 cells, but the simulated clone loss is much higher (Fig 6**D**). However, the overestimation of clone loss by our model may actually be an underestimation of *in vitro* clone loss that is quantified by the PCR and sequencing procedure. Because any clone that had a barcode in the reference library is counted, so-called *spurious reads* may cause false positives

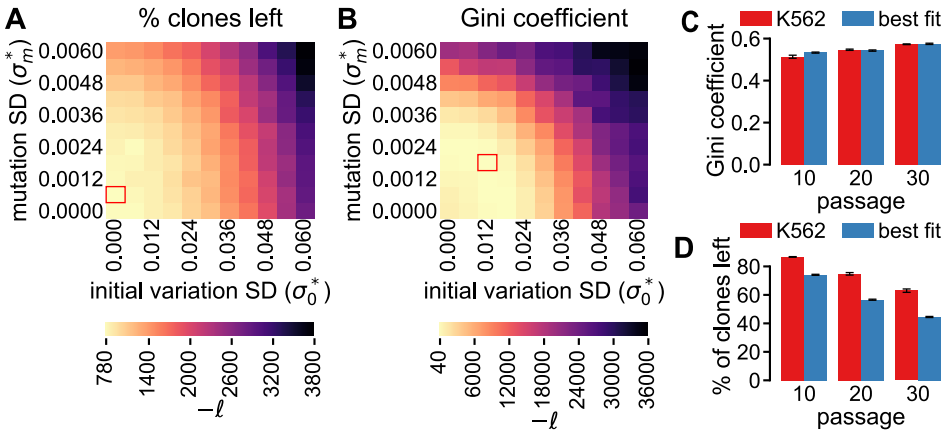


Figure 6: ABM simulations match the limited clonal dominance development for the monoclonal K562 cell line. A-B Maximum Likelihood estimator (ℓ) based on clone loss (**A**) or Gini coefficient (**B**), for a range of initial division rate SDs (σ_0^*) and mutation SDs (σ_m^*). **C-D** Clone loss (**C**) and clonal dominance (**D**) in simulations, with the parameters from the red rectangle in **B**, and in the experiments with monoclonal K562 cells. All simulation results are the mean of 10 runs and the results for the monoclonal K562 cells are the mean of 3 replicates, with the error bars representing the SD.

and thus underestimate actual clone loss. These reads are at least partly the result of systematic errors that are common with next-generation sequencing methods [29]. Porter *et al.* [13] elegantly showed that the sequencing results were reproducible by pair-wise comparing the clone frequencies in four samples of the plasmid library. Nevertheless, this analysis does not completely rule out spurious reads because read errors are correlated to specific base sequences [30] and such errors are expected to occur at similar frequencies in each replicate [31]. To test the effect of spurious reads we artificially 'contaminated' the clone sizes returned by the simulations with spurious reads (see details in S1 File). This analysis showed that such spurious reads can decrease clone loss to such an extent that the model matches *in vitro* clone loss. The effect of spurious reads on the Gini coefficient was less pronounced, arguing in favour of fitting to the Gini coefficient rather than to clone loss data or to a combination of the two metrics. Furthermore, results were similar for the HeLa cells, i.e. for that cell line clone loss was not matched well whereas the Gini coefficient and the number of major clones was (for details see S2 File). Altogether, these results indicate that variation in division rates explains the dynamics of the clone size distribution described by Porter *et al.* [13].

Discussion

Employing a simple model of stochastic cell division and passage, we showed that most of the clone loss taking place during passage, observed by Porter *et al.* [13], is due to the random loss of clones during passage. In contrast, the progressive clonal dominance that developed in the same experiments, cannot at all be explained by random clone loss. Extending the simulations with indefinitely dividing CSCs and DCs with limited division capacity did not induce progressive clonal dominance and led to a much larger clone loss than observed experimentally. However,

a model in which variability of division rates was present and was subject to evolution resulted in a progressive clonal dominance matching the dominance observed experimentally. The level of correspondence between clone loss and the simulations varied per tested cell line, which we hypothesize to result from errors during sequencing as these are known to occur frequently and at a reproducible rate [29, 31]. We showed that indeed such errors have a much larger effect on the measured clone loss than on the Gini coefficient, which we used to characterize clonal dominance.

Altogether, our model provides strong evidence that heritable division rate variation rather than the presence of CSCs induces the changes in clonal distribution observed by Porter *et al.* [13]. Note that our model was set up to mimic *in vitro* growth and cannot be directly extrapolated to predict the development of heterogeneity within real tumors, because in that case spatial effects such as contact inhibition of growth and signals from the tumor micro-environment are likely to play a role. Nevertheless, our simulations elucidate which effect evolution of division rates can have on the clonality of a cell population, and such effects are likely to also be important *in vivo*. For example, several *in vivo*, microscopy-based, lineage tracing studies [10, 11], ascribe the development of large monochromatic patches from a mosaic pattern to CSCs. Based on these observations, Driessens *et al.* [10] proposed a mathematical model with CSCs and DCs that closely fitted the clone sizes they observed in their experiments with skin papillomas. However, their model considered CSCs to divide twice as fast as DCs, while CSCs are typically thought to divide slower or at a similar rate as DCs [4]. It would therefore be interesting to investigate whether division rate variability could also contribute to such *in vivo* lineage tracing data.

Although there is in general ample evidence for the existence of CSCs [1, 3, 4], our model results do not point to a role for CSCs in the experiments of Porter *et al*. Indeed, in the CSC growth model with division rate heterogeneity, clonal dominance appeared in combination with a massive clone loss. This clone loss occurred because only clones that had a CSC at initialization had any chance of generating offspring in the long term, and even those clones could disappear when CSCs were accidentally lost during passage. Consistent with this explanation, removing the distinction between CSCs and DCs in our model, led to clone loss closely matching the experimental observations. Recent studies showed that the CSC fate is plastic, meaning that differentiated cancer cells sometimes can become CSCs [4, 32]. This CSC plasticity might provide an alternative mechanism to prevent clone loss, by enabling clones to (re)acquire CSCs. Adding such plasticity to our CSC growth model would give all clones the potential to generate offspring indefinitely, making it similar to a model in which all cells divide indefinitely, and should thus be able to fit *in vitro* clone loss. To test if such a model with clonal evolution of division rates would develop clonal dominance we extended the CSC model with heritable heterogeneous division rates (see Methods). Simulation with this model indeed showed that clonal dominance develops (see Fig S3 Fig). However, in order to limit clone loss, it was required to have a high probability of symmetric CSC division and a large fraction of initial CSCs in the simulations. Therefore, we expect that a model with CSC plasticity can only reproduce the *in vitro* findings when the transition from DC to CSC is common.

Reproducing the *in vitro* results in our simulation was possible when we considered the division rates of tumor cells to vary between clones and to be fully and directly inherited from the parent cell (although the rate could be changed by mutation). Whereas classical studies have provided ample evidence for division rate heterogeneity among tumor cells [33–35], we are unaware of observations of its full and direct inheritance. Nevertheless, Gray *et al.* [33] showed that when

melanoma cells are iteratively grown in mice, isolated, and transferred into new mice, the tumor growth speed increased every generation, which indicates that fast dividing cells generate offspring that also divide fast. More recent work further supports the assumption of heritable division rates by showing a strong, positive, correlation in the division rate of B-cell siblings [36]. However, other studies have shown that the division times of breast cancer cells correspond less between parents and offspring than between siblings [37], and that in lymphoblasts the division times of parents and offspring do not correlate at all [38]. These findings indicate that the child's division rate is not a direct copy of the parent's division rate. Sandler *et al.* [37] propose a *kicked cell cycle* model where the cell cycle length is determined by the level of an oscillating protein which is inherited from the parent and the phase of this protein determines the time between birth and division. Hence, such a model results in similar division times for siblings, while the correlation between division times of parents and offspring depends on the cell cycle duration [37]. These observations show the need for a better understanding of how the division rate of child cells depend on the parent, which can be achieved from lineage tracing studies employing imaging of multiple divisions over time.

While our model explains the strong clonal dominance evolving over time *in vitro*, it does not perfectly match all *in vitro* observations of Porter *et al.* [13]. The main discrepancy is that our model overestimates the clone loss observed with monoclonal K562 cells or HeLa cells (and is also slightly off for polyclonal K562 cells). Clone loss is minimal when we omit all division rate variation (Fig 2**A**), but even in such simulations clone loss is overestimated. Clone loss can also be reduced by increasing the percentage of clones that is passed (S4 Fig**A**), however this also reduces clonal dominance (S4 Fig**B**). Alternatively, it could be that passaging occurs with a bias for small clones to survive passaging, for instance when clones closely attach to each other *in vitro* and small groups of attached cells are more likely to be selected for passaging than large groups. However, this explanation seems unlikely considering that trypsinization was sufficient to detach any adhering cells. Therefore, we consider it likely that *in vitro* clone loss was actually underestimated in the experimental data due to the presence of spurious reads. Our analysis (S1 File) confirmed that our model would better fit the *in vitro* data if such spurious reads are included in the simulation results.

Altogether, in this work we used a computational approach to test different hypotheses for the development of clonal dominance and showed that only the presence and evolution of division rate heterogeneity, can reproduce the experimental observations. Hence, this study showcases the value of computational modeling in the interpretation of experimental results. In the future, the model could be further extended to improve its power, especially for comparison with *in vivo* data. For this, the model should be extended with an explicit representation of space and physical interactions between cells [39]. With such a model it becomes possible to explore the consequences of division rate variability while comparing with intra-vital images studies.

Methods

Analysis of sequencing data

The sequencing data were downloaded from the NIH Sequence Read Archive (`https://www.ncbi.nlm.nih.gov/sra/SRX535233`) using the SRA Toolkit. The downloaded FASTQ files were processed with our own code (S3 File), which collects all barcodes for which the read quality is

56 or higher and counts the occurrence of each barcode. We used our own code rather than the ClusterSeq code (i.e., the code developed and used in Porter *et al.* [13]), because ClusterSeq clusters barcodes for each dataset separately, which may result in identical barcodes ending up in a different cluster at different time steps of the same experiment.

To generate a reference library for the cell lines barcoded with the lentiviral vector, we merged the four plasmid library samples, excluding any barcode with a frequency smaller than 0.0002% or that appeared in only one biological replicate. The resulting reference library, containing 13325 barcodes, was then used to select only these known barcodes from the experimental data (S1 Dataset). Finally, we processed the FASTQ files containing the reads for the experiments with the polyclonal K562, monoclonal K562, and HeLa cell lines. The full processing of these files is outlined in S3 File.

SSA model of stochastic growth & passage

In order to understand the clone size dynamics of tumour cells, we employed Gillespie's Stochastic Simulation Algorithm (SSA) [26] to simulate the growth and passage of a cell population. The SSA allows to follow the size N_i of each clone i over time rather than that of single cells and we applied this to the polyclonal K562 cell line. The simulations were initialized based on the read counts for this cell line at passage 0 (see S2 Dataset), which contains the counts for $n_{\text{library}} = 13325$ unique barcodes. Hence, we initialized N_i for $1 \leq i \leq n_{\text{library}}$ as $N_i(0) = \frac{n_{\text{pass}}}{n_{\text{library}}} N_{\text{K562},i}$, where $N_{\text{K562},i}$ is the read count for clone i in the polyclonal K562 sample and $n_{\text{pass}} = 3 \cdot 10^5$ is the number of cells that are selected during passaging (i.e., also for the initial passage).

The simulations closely follow the experimental procedure of iterated growth and passage: the initial population of $3 \cdot 10^5$ cells grows until the critical population size $n_{\text{crit}} = 4 \cdot 10^6$ is reached, then n_{pass} cells are selected randomly and taken to the next generation. We implemented the SSA using the τ-leaping algorithm [40], in which size N_i of clone i is described by:

$$N_i(t + \tau) = N_i(t) + k_i \tag{1}$$

with time step τ, and k_i taken from a Poisson distribution with mean $rN_i(t)\tau$, where r denotes the division rate. These τ leaps are performed until the population size $n = \sum_{i=0}^{C} N_i \geq n_{\text{pass}}$, where C is the number of clones. We set $\tau = 0.0005$, which resulted in a simulation time of typically ~100 seconds (for ~12,000 clones with one species and r=1) with similar results as for lower τ values (S5 Fig). The source code of the simulations can be found in S4 File.

To test how CSC driven growth affects the clone size evolution, we incorporated a previously published model of CSC driven growth [28]. In this model CSCs proliferate at a division rate of r_{CSC} and division can result either in two CSCs with probability p_1, in a CSC and a DC with probability p_2, or in DCs with probability p_3 (Fig 3**A**). DCs proliferate at a division rate r_{DC} until they reach their maximum number of divisions M and then, following Weekes *et al.* [28], they die with a rate r_{DC}. For simplicity, we did not consider random cell death of CSCs and DCs, because this process only affects the population division rate. This division scheme is incorporated in the growth model by defining for each clone i the number of CSCs $N_{\text{CSC},i}$ and the number of DCs $N_{\text{DC}_m,i}$ of age m (where m can take values ranging from 0 to the maximum age M). Then, for each clone i there are 5 possible transitions:

1. CSC → 2CSC: $N_{\text{CSC},i}(t + \tau) = N_{\text{CSC},i}(t) + k_i$;

2. CSC → CSC + DC: $N_{\mathrm{DC}_0,i}(t+\tau) = N_{\mathrm{DC}_0,i}(t) + k_i$;

3. CSC → 2DC: $N_{\mathrm{CSC},i}(t+\tau) = N_{\mathrm{CSC},i}(t) - k_i$ and $N_{\mathrm{DC}_0,i}(t+\tau) = N_{\mathrm{DC}_0,i}(t) + 2k_i$;

4. $\mathrm{DC}_m \rightarrow 2\mathrm{DC}_{m+1}$: $N_{\mathrm{DC}_m,i}(t+\tau) = N_{\mathrm{DC}_m,i}(t) - k_i$ and $N_{\mathrm{DC}_{m+1},i}(t+\tau) = N_{\mathrm{DC}_{m+1},i}(t) + 2k_i$ for $0 \leq m < M$;

5. $\mathrm{DC}_M \rightarrow$: $N_{\mathrm{DC}_M,i}(t+\tau) = N_{\mathrm{DC}_M,i}(t) - k_i$,

where k_i is obtained as described above, except for transitions 1-3 where the CSC division rate is multiplied by the respective transition probability. At initialization, the number of CSCs in each clone is set to $N_{\mathrm{CSC},i} = \mathrm{CSC}_0 N_i$, rounded to the nearest integer. The remaining cells are distributed evenly over the M DC species, while rounding to the lowest integer: $N_{\mathrm{DC},i} = \left\lfloor \frac{1}{M}(1-\mathrm{CSC}_0)N_i \right\rfloor \ \forall\ 0 \leq i \leq M$. When, due to rounding, $N_{\mathrm{CSC},i} + \sum_{i=0}^{M} N_{\mathrm{DC},i} < N_i$, the remaining cells are randomly distributed over all species of clone i.

To test the effect of division rate heterogeneity on the development of the clone size distribution, the division rates $r_{\mathrm{i,CSC}}$ and $r_{\mathrm{i,DC}}$ of each clone i are multiplied by a randomly chosen value X_i that is obtained from a normal distribution with mean 1 and SD σ_r^*, and X_i is set to 0 when X_i <0. Note that σ_r^* is the division rate SD relative to the mean division rate; the actual division rate SD is $\sigma_r^* \cdot r$. Because X_i is linked to clone i, the division rates are inherited upon division.

ABM of stochastic growth & passage

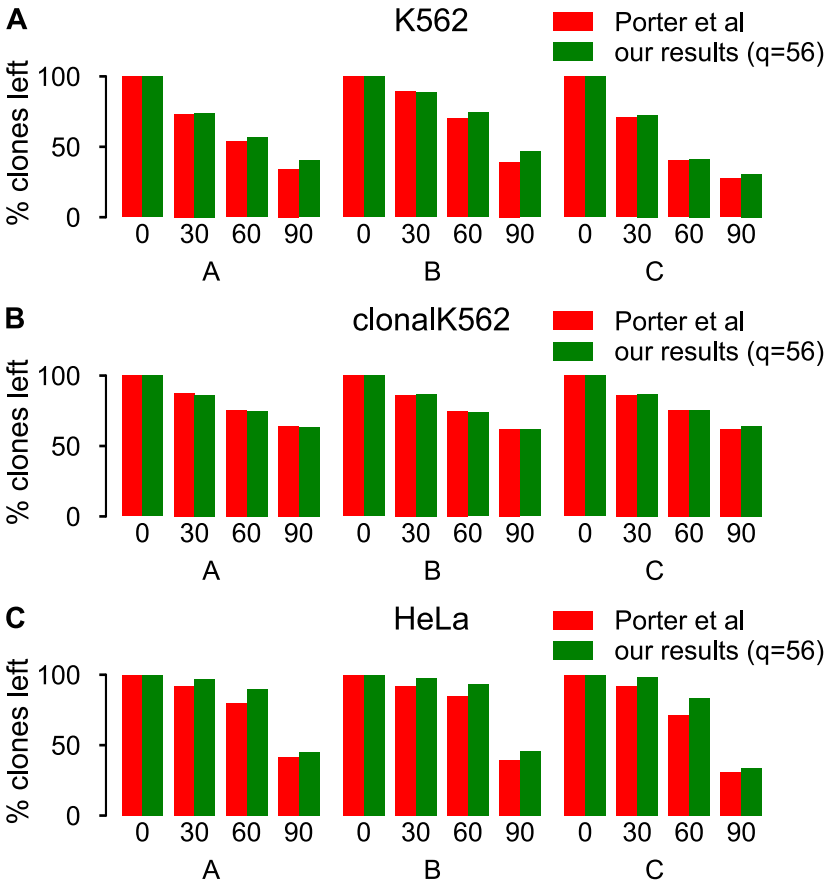
The agent based model (ABM) differs from the SSA model in the explicit representation of individual cells rather than of clones. In the ABM, each cell i has a barcode b_i and a division rate r_i. All simulations are initialized based on sequencing data of either the polyclonal K562 cells, the monoclonal K562 cells, or the (polyclonal) HeLa cells, which can all be found in S2 Dataset. First, we create a *master population* from which the initial population of each replicate is selected. The master population consists of $5 \cdot 10^6$ cells, each of which contains a barcode following read abundance in the sequencing data, such that the relative abundance of the barcodes is initially correct. Then, a division rate $r_i = r_0 \max(X, 0)$ is assigned to each cell in the *master population*, where X is randomly drawn from a normal distribution with mean 1 and SD σ_0^*, and r_0 is the initial mean division rate. The seed of the random number generator used for division rate selection is kept constant across simulation replicates such that the same division rate is associated with the same cell and the same barcode in every replicate. Finally, n_{init} cells are taken from the initial population to be used in the simulation of iterated growth and passage. Note that the initialization omits the clonal dynamics during the 8 to 9 days before the growth & passage experiments started. During this time interval, a within-clone correlation of the division rates likely develops, which is missing in the simulated cell population. As a result the model does not reproduce the overlap between major clones that was observed by Porter *et al.* [13].

Division is modeled using the dynamic Monte Carlo method [41] implemented with a first reaction method. Whenever a cell is created, either during initialization or after division, the next division time for that cell is assigned according to $t_i = \frac{-\ln R}{r_i}$, where r_i is the cell's division rate and R is a random number from a uniform distribution between 0 and 1. Then, by creating an ordered list of next division times, population growth can be simulated efficiently. When division occurs, a new division rate r_c is assigned to each newly created cell according to $r_c = r_p \max(Y, 0)$ with r_p

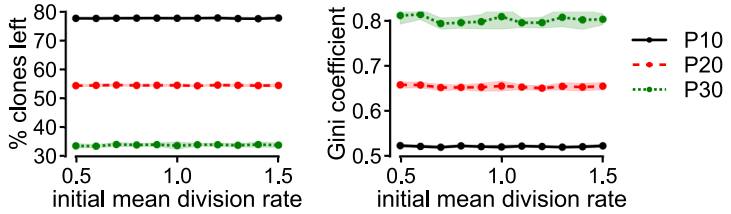
the parent's division rate and Y taken from a normal distribution with mean 1 and SD σ_m^*. Note that σ_0^* is the initial division rate SD relative to r_0; the actual SD of the initial division rates is $\sigma_0^* \cdot r_0$. As in the SSA, division continues until the population size reaches its maximum n_{crit} after which n_{pass} cells are randomly selected and passed to the next generation. The source code of the simulations can be found in S5 File.

Note that there are some implementation differences between the aforementioned SSA model and the ABM, which result in small quantitative differences in the simulation results. These differences occur for two reasons: First, growth is slightly underestimated in the SSA model, and the exact underestimation depends on the clone size. Second, a clone in the ABM can include cells with different division rates, which is not the case in Gillespie simulations.

Supporting information

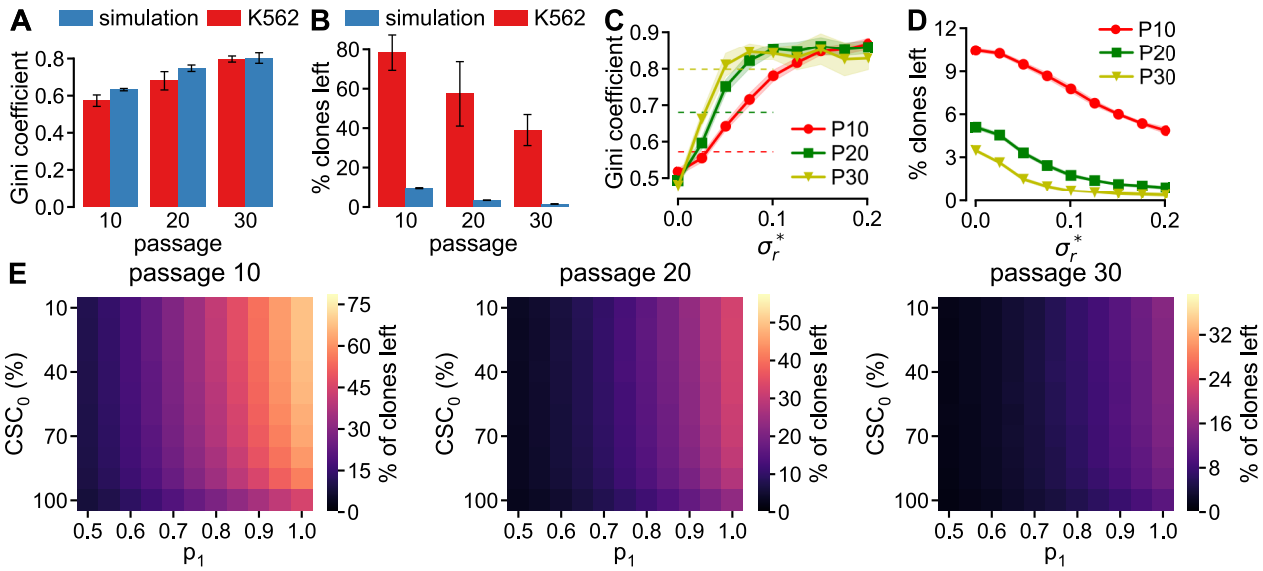


S1 Fig.
Comparison of clone number in [13] and in our analysis. **A** Polyclonal K562 cells. **B** Monoclonal K562 cells. **C** HeLa cells. The values in the original publication were retrieved from [13] (Figure 3f, Additional File 5d, and Additional File 12d), whereas a description of our analysis can be found in the Methods.



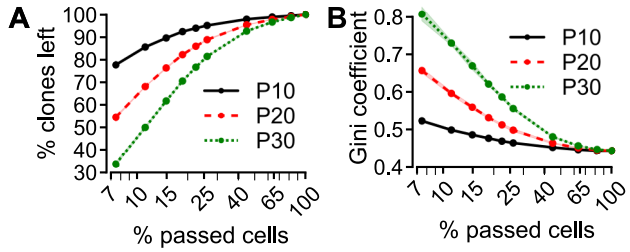
S2 Fig.

The initial mean division rate in the ABM does not affect the clone size distribution. Clone loss (left panel) and Gini coefficient (right panel) for a series of simulations with $\sigma_0^* = 0.036$ and $\sigma_r^* = 0.0018$ and varying values of the initial mean division rate r_0. All data points are the mean of 10 simulations and shaded areas (only visible for the non-solid lines) represent the SD.



S3 Fig.

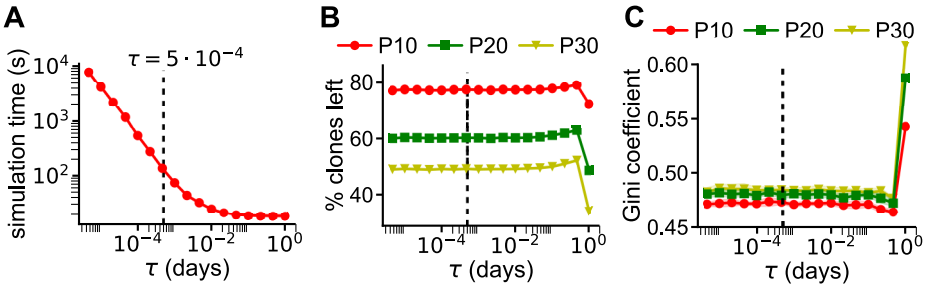
Clonal dominance develops in simulations with the CSC growth model and division rate variability. **A**-**B** Gini coefficient (**A**) and clone loss (**B**) in a simulation with $\sigma_r^* = 0.05$. **C**-**D** Relation between the division rate SD and Gini coefficient (**C**) or clone loss (**D**), with the horizontal lines in **C** denoting the corresponding experimental values for the polyclonal K562 cell line. **E** Clone loss for simulations with a varying initial CSC percentage (CSC_0) and symmetric CSC division probability (p_1). The maximum of each colormap is set to the average clone loss, at the respective passage, of the three biological replicates. All values are the mean of 10 simulation replicates and the error bars (**A**-**B**) and the shaded areas (**C**-**D**) represent the SD.



S4 Fig.

Increasing the number of cells passed reduces clone loss and clonal dominance **A**-**B** Clone loss (**A**) and Gini coefficient (**B**) for a varying percentage of passed cells (relative to n_{crit}). All data points are the mean of 10 simulations and shaded areas (only visible for the non-solid lines)

represent the SD.



S5 Fig.
Effect of the simulation time step on the Gillespie simulations. A-C Effect of the time step (τ) on the simulation time (A), on the clone loss (B) and on the Gini coefficient(C). The vertical dashed line marks the τ used for all simulations. All points are the mean of 10 simulation replicates and the shaded areas (only visible in **C**) represent the SD.

S1 File. Analysis of the effects of spurious clones on simulated clone loss and gini coefficient.

S2 File. Comparison of the ABM describing division rate evolution to the HeLa data

S3 File. Analysis of the FASTQ files. File contains an executable jupyter notebook and a pdf print of that notebook as well as all code needed to process the FASTQ files. *This file will be made available upon publication.*

S4 File. Archive containing the source code for the SSA model. *This file will be made available upon publication.*

S5 File. Archive containing the source code for the ABM. *This file will be made available upon publication. This file will be made available upon publication.*

S1 Dataset. Reference library used for the analysis of the experimental data. *This file will be made available upon publication.*

S2 Dataset. Barcode counts of the polyclonal K562 cell line barcoded with the lentiviral vector, at passage 0. *This file will be made available upon publication.*

Acknowledgments

We thank Huan Yang for helpful discussions on setting up the maximum likelihood estimation.

References

[1] Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;501(7467):328–337. doi:10.1038/nature12624.

[2] Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med. 2015;22(1):105–113. doi:10.1038/nm.3984.

[3] Caiado F, Silva-Santos B, Norell H. Intra-tumour heterogeneity - going beyond genetics. FEBS J. 2016;283(12):2245–2258. doi:10.1111/febs.13705.

[4] Clevers H. The cancer stem cell: premises, promises and challenges. Nat Med. 2011;17(3):313–319. doi:10.1038/nm.2304.

[5] Shackleton M, Quintana E, Fearon E, Morrison S. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009;138(5):822–829. doi:10.1016/j.cell.2009.08.017.

[6] Quintana E, Shackleton M, Sabel MS, Fullen DR, Johnson TM, Morrison SJ. Efficient tumour formation by single human melanoma cells. Nature. 2008;456(7222):593–598. doi:10.1038/nature07567.

[7] Blanpain C, Simons BD. Unravelling stem cell dynamics by lineage tracing. Nat Rev Mol Cell Biol. 2013;14(8):489–502. doi:10.1038/nrm3625.

[8] Scheele CLGJ, Maynard C, van Rheenen J. Intravital Insights into Heterogeneity, Metastasis, and Therapy Responses. Trends Cancer Res. 2016;2(4):205–216. doi:10.1016/j.trecan.2016.03.001.

[9] Schepers AG, Snippert HJ, Stange DE, van den Born M, van Es JH, van de Wetering M, et al. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science. 2012;337(6095):589–609. doi:10.1306/12280707023.

[10] Driessens G, Beck B, Caauwe A, Simons BD, Blanpain C. Defining the mode of tumour growth by clonal analysis. Nature. 2012;488(7412):527–530. doi:10.1038/nature11344.

[11] Zomer A, Ellenbroek SIJ, Ritsma L, Beerling E, Vrisekoop N, Van Rheenen J. Brief report: Intravital imaging of cancer stem cell plasticity in mammary tumors. Stem Cells. 2013;31(3):602–606. doi:10.1002/stem.1296.

[12] Tang Q, Moore JC, Ignatius MS, Tenente IMIM, Hayes MN, Garcia EG, et al. Imaging tumour cell heterogeneity following cell transplantation into optically clear immune-deficient zebrafish. Nat Comm. 2016;7(10358). doi:10.1038/ncomms10358.

[13] Porter SN, Baker LC, Mittelman D, Porteus MH. Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. Gen Biol. 2014;15(5):R75. doi:10.1186/gb-2014-15-5-r75.

[14] Bhang HEC, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. Nat Med. 2015;21(5):440–448. doi:10.1038/nm.3841.

[15] Nolan-Stevaux O, Tedesco D, Ragan S, Makhanov M, Chenchik A, Ruefli-Brasse A, et al. Measurement of Cancer Cell Growth Heterogeneity through Lentiviral Barcoding Identifies Clonal Dominance as a Characteristic of In Vivo Tumor Engraftment. PLoS ONE. 2013;8(6):e67316. doi:10.1371/journal.pone.0067316.

[16] Nguyen LV, Cox CL, Eirew P, Knapp DJHF, Pellacani D, Kannan N, et al. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. Nat Comm. 2014;5(5871). doi:10.1038/ncomms6871.

[17] Guernet A, Mungamuri SK, Cartier D, Sachidanandam R, Jayaprakash A, Adriouch S, et al. CRISPR-Barcoding for Intratumor Genetic Heterogeneity Modeling and Functional Analysis of Oncogenic Driver Mutations. Mol Cell. 2016;63(3):526–538. doi:10.1016/j.molcel.2016.06.017.

[18] Klauke K, Broekhuis MJC, Weersing E, Dethmers-Ausema A, Ritsema M, Vilà González M, et al. Tracing dynamics and clonal heterogeneity of Cbx7-induced leukemic stem cells by cellular barcoding. Stem Cell Reports. 2015;4(1):74–89. doi:10.1016/j.stemcr.2014.10.012.

[19] Bystrykh LV, Belderbos ME. Clonal Analysis of Cells with Cellular Barcoding: When Numbers and Sizes Matter. In: Turksen K, editor. Stem Cell Heterogeneity. Springer New York; 2016. p. 57–89.

[20] Enderling H. Cancer stem cells: small subpopulation or evolving fraction? Integr Biol. 2014;7(1):14–23. doi:10.1039/c4ib00191e.

[21] Sottoriva A, Sloot PMA, Medema JP, Vermeulen L. Exploring cancer stem cell niche directed tumor growth. Cell Cycle. 2010;9(8):1472–1479. doi:10.4161/cc.9.8.11198.

[22] Sottoriva A, Verhoeff JJC, Borovski T, McWeeney SK, Naumov L, Medema JP, et al. Cancer Stem Cell Tumor Model Reveals Invasive Morphology and Increased Phenotypical Heterogeneity. Cancer Res. 2010;70(1):46–56. doi:10.1158/0008-5472.CAN-09-3663.

[23] Sottoriva A, Vermeulen L, Tavaré S. Modeling Evolutionary Dynamics of Epigenetic Mutations in Hierarchically Organized Tumors. PLoS Comput Biol. 2011;7(5):e1001132. doi:10.1371/journal.pcbi.1001132.

[24] Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak Ma. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature. 2015;525. doi:10.1038/nature14971.

[25] Ceriani L, Verme P. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. The Journal of Economic Inequality. 2012;10(3):421–443.

[26] Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys. 1976;22(4):403–434. doi:10.1016/0021-9991(76)90041-3.

[27] Shibata M, Shen MM. The roots of cancer: Stem cells and the basis for tumor heterogeneity. BioEssays. 2013;35(3):253–260. doi:10.1002/bies.201200101.

[28] Weekes SL, Barker B, Bober S, Cisneros K, Cline J, Thompson A, et al. A multicompartment mathematical model of cancer stem cell-driven tumor growth dynamics. Bull Math Biol. 2014;76(7):1762–1782. doi:10.1007/s11538-014-9976-0.

[29] Deakin CT, Deakin JJ, Ginn SL, Young P, Humphreys D, Suter CM, et al. Impact of next-generation sequencing error on analysis of barcoded plasmid libraries of known complexity and sequence. Nucleic Acids Res. 2014;42(16):e129. doi:10.1093/nar/gku607.

[30] Meacham F, Boffelli D, Dhahbi J, Martin DIK, Singer M, Pachter L. Identification and correction of systematic error in high-throughput sequence data. BMC Bioinformatics. 2011;12:451. doi:10.1186/1471-2105-12-451.

[31] Beltman JB, Urbanus J, Velds A, van Rooij N, Rohr JC, Naik SH, et al. Reproducibility of Illumina platform deep sequencing errors allows accurate determination of DNA barcodes in cells. BMC Bioinformatics. 2016;17:151. doi:10.1186/s12859-016-0999-4.

[32] Singh AK, Arya RK, Maheshwari S, Singh A, Meena S, Pandey P, et al. Tumor heterogeneity and cancer stem cell paradigm: Updates in concept, controversies and clinical relevance. Int J Cancer. 2015;136(9):1991–2000. doi:10.1002/ijc.28804.

[33] Gray JM, Pierce GB. Relationship between growth rate and differentiation of melanoma in vivo. J Natl Cancer Inst. 1964;32(6):1201–1211. doi:10.1093/jnci/32.6.1201.

[34] Dexter DL, Kowalski HM, Blazar Ba, Blazar A, Vogel R, Gloria H. Heterogeneity of Tumor Cells from a Single Mouse Mammary Tumor. Cancer Res. 1978;38(10):3174–3181.

[35] Danielson KG, Anderson LW, Hosick HL. Selection and Characterization in Culture of Mammary Tumor Cells with Distinctive Growth Properties in Vivo. Cancer Res. 1980;40(6):1812–1819.

[36] Duffy KR, Wellard CJ, Markham JF, Zhou JHS, Holmberg R, Hawkins ED, et al. Activation-Induced B Cell Fates Are Selected by Intracellular Stochastic Competition. Science. 2012;335(6066):338–341. doi:10.1126/science.1213230.

[37] Sandler O, Mizrahi SP, Weiss N, Agam O, Simon I, Balaban NQ. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. Nature. 2015;519(7544):468–471. doi:10.1038/nature14318.

[38] Cornwell JA, Hallett RM, der Mauer SA, Motazedian A, Schroeder T, Draper JS, et al. Quantifying intrinsic and extrinsic control of single-cell fates in cancer and stem/progenitor cell pedigrees with competing risks analysis. Sci Rep. 2016;6(January):27100. doi:10.1038/srep27100.

[39] Van Liedekerke P, Palm MM, Jagiella N, Drasdo D. Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. Comp Part Mech. 2015;2:401–444. doi:10.1007/s40571-015-0082-3.

[40] Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys. 2001;115(4):1716–1733. doi:10.1063/1.1378322.

[41] Fichthorn KA, Weinberg WH. Theoretical foundations of dynamical Monte Carlo simulations. J Chem Phys. 1991;95(2):1090–1096. doi:10.1063/1.461138.