Heritable tumor cell division rate heterogeneity induces clonal dominance

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

[a]Division of Toxicology, Leiden Academic Center for Drug Research, Leiden University, The Netherlands

January 2, 2017

Abstract

Tumors frequently develop clonal dominance, meaning that most cells descend from a small fraction of initial clones. Such dominance also occurred in published *in vitro* iterated growth and passage experiments with tumor cells in which lineage tracing was performed using genetic barcodes. A potential source for such heterogeneity is that dominant clones derive from cancer stem cells with an unlimited self-renewal capacity. Alternatively, clonal dominance may be caused by heterogeneity of division rates between clones. To test which hypothesis can explain the development of clonal dominance, we build a computational model that accurately simulates the iterated growth and passage experiments in which clonal dominance emerged. With this model we show that clonal dominance does not develop in a model where all cells grow at a uniform rate, even when the population is subdivided in cancer stem cells and differentiated cells. However, when the division rates vary between clones and are inherited from parent cells, clonal dominance develops and increases over time. The experimentally observed evolution of clone loss, clonal dominance, and the clonal overlap across biological replicates are closely matched by our simulations. Thus, our findings suggest that tumor cells exhibit a heritable variation in the division rates of individual cells.

Major Findings

The clonal dominance that developed in *in vitro* iterated growth and passage experiments can be reproduced with a computational model when the division rates vary between cells and are inherited to their offspring. 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. These results suggest that clonal dominance, which is frequently also observed in *in vivo* tumors, is not always the result of cancer stem cells.

Quick Guide to Equations and Assumptions

The simulations closely follow the experimental procedure of iterated growth and passage: during growth cells divide randomly with a given rate and during passage a random selection is made of cells that are passed on to the next generation. Because we aim to understand the dynamics of sizes of clones, we follow the clone sizes rather than single cells over time by

employingGillespie's Stochastic Simulation Algorithm (SSA) [1]. We implemented the SSA using the τ-leaping algorithm [2], 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 $n = \sum_{i=0}^{C} N_i$ reaches the critical population size n_{crit} that matches the *in vitro* population size right before passage.

Cancer stem cell growth

To test how cancer stem cell (CSC) driven growth affects the clone size evolution, we incorporated a previously published model of CSC driven growth [3]. In this model cells are either CSCs that can divide indefinitely, or differentiated cells (DCs) that divide a limited number of times. CSCs proliferate at a net growth 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 net growth rate r_{DC} until they reach their maximum number of divisions M and then, following Weekes *et al.* [3], they die with a rate r_{DC}. 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. Then, for each clone i there are 5 possible transitions:

1. CSC $\rightarrow$ 2CSC: $N_{\text{CSC},i}(t+\tau) = N_{\text{CSC},i}(t) + k_i$;
2. CSC $\rightarrow$ CSC + DC: $N_{\text{DC}_0,i}(t+\tau) = N_{\text{DC}_0,i}(t) + k_i$;
3. CSC $\rightarrow$ 2DC: $N_{\text{CSC},i}(t+\tau) = N_{\text{CSC},i}(t) - k_i$ and $N_{\text{DC}_0,i}(t+\tau) = N_{\text{DC}_0,i}(t) + 2k_i$;
4. $\text{DC}_m \rightarrow \text{DC}_{m+1}$: $N_{\text{DC}_m,i}(t+\tau) = N_{\text{DC}_m,i}(t) - k_i$ and $N_{\text{DC}_{m+1},i}(t+\tau) = N_{\text{DC}_{m+1},i}(t) + 2k_i$ for $0 \leq m < M$;
5. $\text{DC}_M \rightarrow$: $N_{\text{DC}_M,i}(t+\tau) = N_{\text{DC}_M,i}(t) - k_i$,

where the means for the Poisson distributions used to obtain k_i are computed in a similar manner as above, except for transition 1-3 where the CSC division rate is multiplied by the respective transition probability.

Division rate variability

To test the effect of division rate heterogeneity on the development of the clone distribution, the default division rate r of the basic model, or the division rates r_{CSC} and r_{DC} for the CSC model, are multiplied by a value X_i chosen from a normal distribution with mean 1 and standard deviation σ_r. X_i is linked to clone i, so division rates are inherited upon division. Furthermore, each simulation replicate uses the same random seed to initialize the division rates and therefore the r_i values of individual clones are identical across all replicates of a single simulation.

Introduction

Intratumoral heterogeneity, the genotypic and phenotypic differences within a single tumor, is a well known feature of cancer [4] and strongly influences the effectiveness of cancer therapy [5]. 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 [6]. 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 differentiate into cells with a limited potential to renew [7]. The presence of CSCs would result in a tumor containing a mixture of CSCs, and populations of cells with varying levels of differentiation.

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 [4, 8], 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 [9], or to the dramatic change in tumor micro-environment upon transplantation [8]. An alternative approach to identify the existence of CSCs is to perform lineage tracing by fluorescent marking of a subpopulation of cells [10]. For example, Schepers *et al.* [11] 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 [12–14]. 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 [15–21]. Tumors grown in immunodeficient mice injected with barcoded tumor cells are dominated by cells that express a only small subset of the barcodes [17, 18]. 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 [20]. Porter *et al.* [15] 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.* [15] showed that clonal dominance progressed over time (Fig 1**B**). Interestingly, similar experiments with a monoclonal K562 cell line did not result in 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.* [17] showed that clonal dominance does not develop 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 [8]. Alternatively, evolution may cause clones with a higher division or survival rate to dominate the tumor. Hence, it is necessary to further investigate the CSC hypothesis and to explore alternative hypotheses. 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 [22]. 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 [23, 24], by CSCs and epigenetic changes [25], or by mutations [26]. In this work we built a computational model that simulates the iterated growth and passage experiments described in [15]. By incorporating different hypotheses for the development of clonal dominance in our model, we show that the presence on CSCs alone does not suffice to reproduce *in vitro* clonal dynamics, while simulations with division rate heterogeneity do closely resemble *in vitro* results.

Methods

Analysis of sequencing data

The sequencing data was 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 using ClusterSeq [15] (`https://github.com/adaptivegenome/clusterseq`), which extracts barcodes and their counts from the FASTQ files and removes ambiguous barcodes. To generate a reference library for the cell lines barcoded with the lentiviral vector, the 4 plasmid library samples were merged, excluding any barcode with a frequency smaller than 0.0002% or that appeared in only 1 biological replicate. The resulting reference library, containing 13237 barcodes, was then used to select only known barcodes from experimental data.

Computational model

The computational model of stochastic growth and passage consists of iterative growth and passage steps during which the size $N_{i,s}$ of each clone i of cell type s changes. Each simulation is initialized with $3 \cdot 10^5$ cells. During the growth phase, cell divisions and cell death (if applicable) occur at random intervals, but with an overall rate r_s. When the critical population size $N_{\text{crit}} = 4 \cdot 10^6$ is reached, $n_{\text{pass}} = 3 \cdot 10^5$ clones are selected randomly and taken to the next generation. More details can be found in the Quick Guide. All simulations were performed with a single, custom C++ program, which is available at github (link will be made available upon publication).

Results

Re-analysis of published sequencing data

We started by re-analyzing the data for the *in vitro* experiments with the lentivirally barcoded polyclonal K562 cell line (see Methods section for more details). The 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**). Iterated growth and passage with the K562 cell line resulted in a loss of clones and a progressive clonal dominance (Fig 1**B**-**C** and

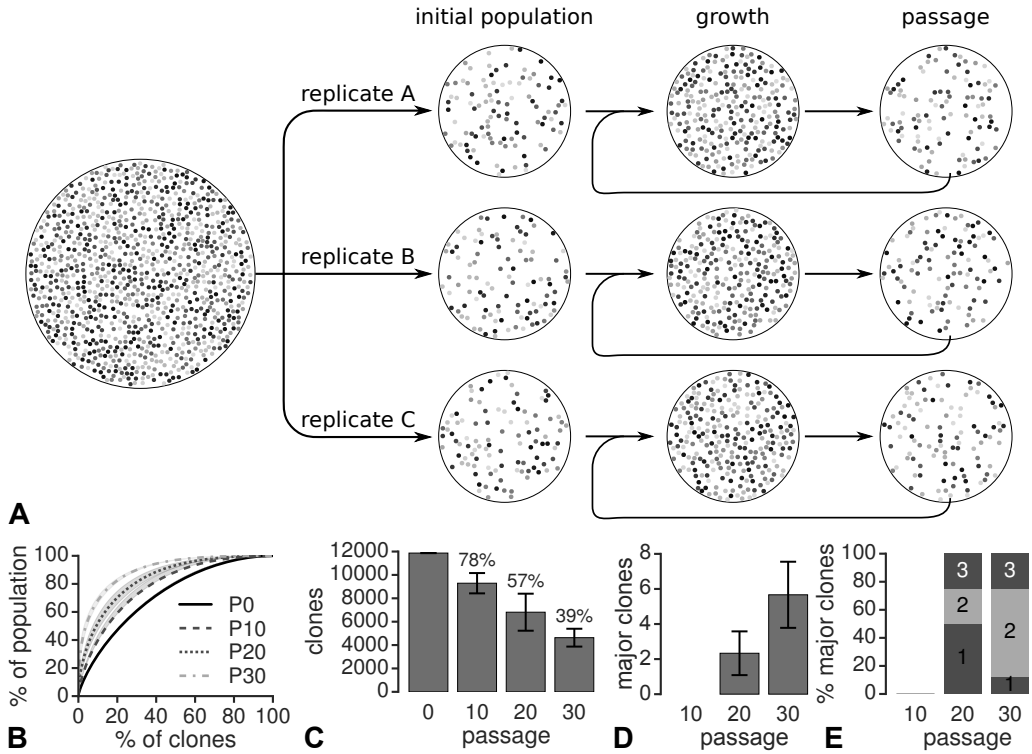

Figure 1: *In vitro* iterated growth and passage experiments result in clone loss and clonal dominance. **A** Experimental setup of Porter *et al.* [15]. **B**-**D** Evolution of the clone distribution visualized by the percentage of clones versus the percentage of the population represented by those clones (**B**), the number of clones (**C**), and the emergence of major clones (clones representing >1% of the cells) (**D**). All values are the average of the three biological replicates and the error bars or colored areas mark the standard deviations. **E** Similarity of major clones across the 3 replicates expressed as the percentage of major clones occurring in in one or more replicates indicated with the colors/numbers.

Fig S1). While the qualitative results are similar for all three replicates, there some variation in the percentage of remaining clones and the clone distribution. Major clones, clones representing $> 1\%$ of the cell population, emerged in all biological replicates (Fig 1**D**). Comparing the barcodes of the major clones in the 3 replicates shows that the majority of major clones coexists in exactly 2 out of 3 replicates (Fig 1**E**), suggesting that dominance depends on a combination of intrinsic clonal properties and chance. In the following, we will develop model simulations based on the experimental setup and compare the model results to the experimental findings with respect to clonal dominance, clone loss, and the development of major clones.

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

The experimental results support the importance of chance 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

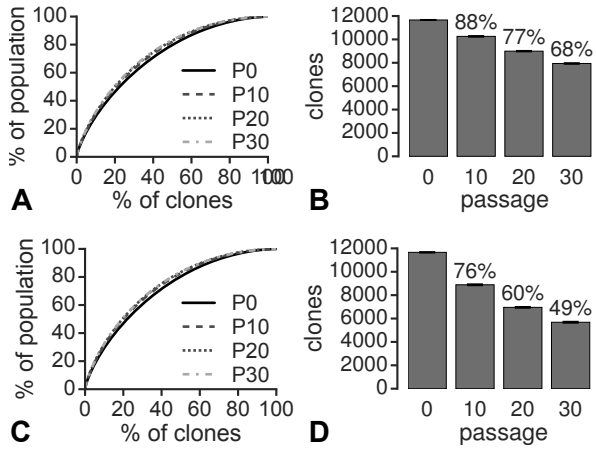

Figure 2: Clonal dominance does not develop for indefinitely dividing cells. **A**-**B** Evolution of clone distribution for Porter's [15] model with stochastic passage and deterministic division, with $t_{\text{growth}} = 72$ hours and $n_{\text{pass}} = 3 \cdot 10^5$ cells. **C**-**D** Evolution of clone distribution for a model with stochastic division and passage implemented with Gillespie's τ-leaping method. Both models were initialized using the experimentally observed initial clone distribution of the polyclonal K562 cell line that was barcoded using a lentiviral vector. All values are the mean of 10 simulations and the error bars in **A** and **C** represent the standard deviation.

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 [15]. To test whether clone loss during passage explains clonal loss and progressive clonal dominance, Porter *et al.* [15] employed a computational model of iterated deterministic growth phases and random passage, showing that this could only partially explain the experimentally observed clone loss but not the development of clonal dominance. 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} the duration of the growth phase, and r 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**). To confirm the simulation results, we used the available code [15] and ran simulations starting with $C = 14,000$ clones uniformly distributed over $3 \cdot 10^5$ cells, which indeed resulted in a reduction of the number of clones (Fig S2**B**). However, in contrast to the *in vitro* results, clonal dominance only developed slightly and did not increase over time, clone loss was small, and no major clones appeared (Fig S2**A**-**B**). To test if a realistic initial clone distribution improved the resemblance between the experimental observations and the simulation results, we repeated the simulations using the initial clone distribution of the K562 cell line (Fig 1**B**) in which clonal dominance is already slightly developed. In this setting the clone distribution remained unchanged (Fig 2**A**) and number of remaining clones remained larger than that for the K562 cells (Fig 2**B**).

Porter *et al.* [15] noted that in simulations where 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. Therefore, we replaced deterministic growth by a stochastic growth model and performed simulations with stochastic growth initialized with a uniform clone distribution (Fig S2**C**-**D**), or with a the clone distribution derived from the initial clone distribution of the K562 cells (Fig 2**C**-**D**). 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 K562 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 (see Fig 1**C** and Fig 2**D**). 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 stochastic passage and growth can cause clone loss during passage, but neither of these mechanisms can induce progressive clonal dominance.

The presence of CSCs does not induce clonal dominance

The presence of CSCs is thought to induce tumor heterogeneity because they generate a, hierarchically organized, population of CSCs and differentiated cells (DCs) [27]. To introduce CSCs in our growth model we added a previously published model of CSC driven growth [3]. In this model cells are either CSCs that can divide indefinitely, or DCs that divide a limited number of times. CSCs proliferate at a net growth 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 two DCs with probability p_3 (Fig 3**A**). DCs proliferate at a net growth rate r_{DC} until they reach their maximum number of divisions M and then, following Weekes *et al.* [3], 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 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.* [15]. For this we exploited the analytical solution of the CSC model elegantly derived by Weekes *et al.* [3]. This solution predicts that the population of cells initially grows, and then develops according to one of three growth regimes determined by $\beta = (p_1 - p_3)r_{\mathrm{CSC}}$. When $\beta > 0$, the population continues to grow, 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*" [15], we limited the parameter space to $\beta > 0$, 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}) and the maximum number of divisions of DCs (M). Therefore, we explored how much the simulated population doubling time deviates from the experimental doubling time of 19 hours for a range of parameter combinations (Fig 3**B**). Interestingly, there is a minimum value for the division capacity that DCs need to have in order to reach the *in vitro* doubling time. To illustrate the clonal development of the CSC growth model, we choose an arbitrary parameter set from the region that matches with the experimentally observed population growth rate: $M = 10$ divisions, and $r_{\mathrm{DC}} = \frac{19}{24}$ day $^{-1}$ (white cross in Fig 3**B**), with the remaining parameters set to $r_{\mathrm{CSC}} = 1$ day $^{-1}$, $p_1 = 0.5$, $p_2 = 0.5$, $p_3 = 0$. The evolution of single clones is strongly affected by the number of CSCs present in a clone, so it is important to take the initial percentage of CSCs, CSC_0, into

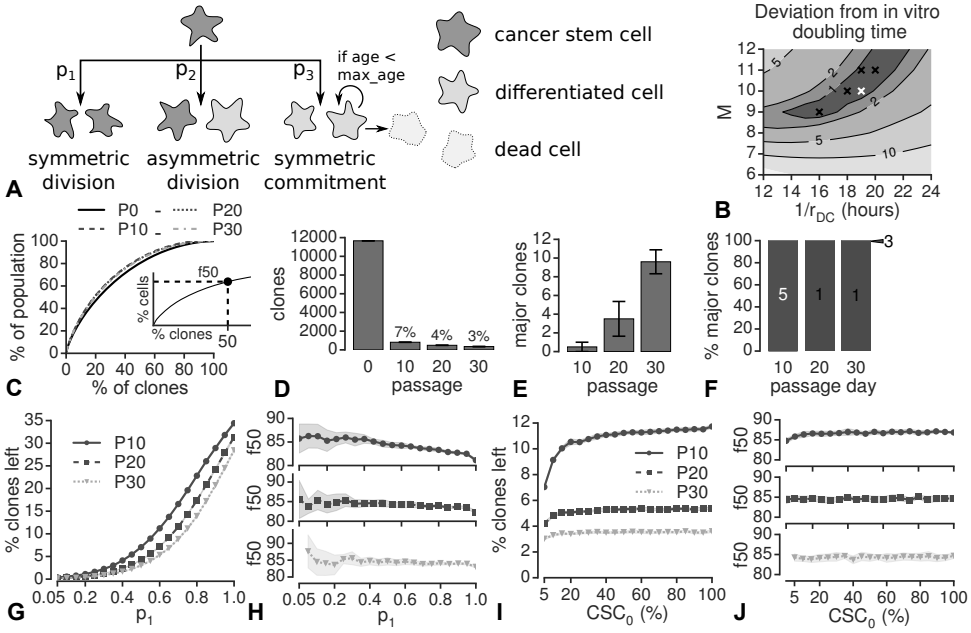

Figure 3: Simulations with the CSC growth model result in massive clone loss and no development of clonal dominance. **A** Scheme illustrating the divisions and cell death in the CSC growth model. **B** Heatmap showing the difference between *in vitro* population growth rate and simulated population growth rate in the CSC growth model depending on the maximum number of DC divisions (M) and DC division rates (r_{DC}). The white cross denotes the default model settings and the black crosses depict a set of alternative settings depicted in Fig S3. **C**-**F** Evolution of the clone distribution (**C**), clone number (**D**), major clones (**E**), and major clone overlap across replicates (**F**) for the CSC growth model. The inset in **C** illustrates how f50 is obtained from the clone distribution. **G**-**H** Effect of symmetric CSC division probability (p_1) on clone loss (**G**) and f50 (**H**). **I**-**J** Effect of the initial CSC percentage (CSC_0) on clone loss (**I**) and f50 (**J**). $p_1 = 0$ and $\mathrm{CSC}_0 = 0\%$ were omitted because, with those settings, the population goes extinct after a few passages. The simulations (**C**-**J**) were performed with the parameter values mentioned in the main text and $\mathrm{CSC}_0 = 5\%$, unless mentioned otherwise. All values in **C**-**E** and **G**-**J** are the mean of 10 simulation replicates and the error bars (**D**-**E**) and colored areas (**G**-**J**) denote the standard deviation.

account. The fraction of CSCs in a tumor is usually thought to be small [7], but recent work has shown that these low values are likely caused by the experimental setup [9]. Therefore, instead of fixing the initial CSC percentage, we evaluated our model for a range of initial CSC fractions.

At first sight surprisingly, our simulations with the CSC growth model resulted in a strong reduction in clone number while the clone distribution remained virtually unchanged (Fig 3**C**-**D**). As a side effect of the massive clone loss, several clones came to represent more than 1% of the population (Fig 3**E**). Nevertheless, clonal dominance did not evolve over time (Fig 3**C**) due to the similar dynamics of surviving clones. Most major clones were unique among the 10 simulation replicates, except for 1 major clone at passage 10 and 1 major clone at passage 30 (Fig 3**F**), which occurred because the clone sizes were initialized from the same, non-uniform, distribution. To

ensure that these results were not caused by the arbitrary choices for M and r_{DC}, we repeated the simulations with several alternative settings from the optimal region in Fig 3**B** (see black crosses), which resulted in nearly identical plots (Fig S3).

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. To easily evaluate the effect of these parameter changes, we devised two metrics. First, to quantify clonal dominance, we determine the percentage of the population that contains the largest 50% of the clones, which is referred to as *f50* (see inset Fig 3**C**). Second, to track clone loss, we measured the percentage of remaining clones at a given passage. In line with our explanation for the massive clone loss, increasing the probability of symmetric CSC differentiation (Fig 3**G**) or increasing the initial CSC percentage reduced clone loss (Fig 3**I**). Nevertheless, increasing 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**H** and 3**J**). 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.

Division rate variability induces clonal dominance

As shown in the previous section, the presence of CSCs does not suffice to induce clonal dominance due to the homogeneous division rate of cells. Classical studies have shown that the division rates of cell lines derived from the same tumor vary *in vitro* and *in vivo* [28–30] and that the growth properties are preserved after transfer [28, 29]. Both these observations suggest that the division rate of individual cells varies which could lead to fast dividing cells becoming more dominant than slowly dividing cells. To test if heterogeneity in the division rate between clones induces progressive clonal dominance, we modified the CSC growth model by replacing the uniform division rates r_{CSC} and r_{DC} by clone-specific growth rates: $r_{\mathrm{CSC},i} = X_i \cdot r_{\mathrm{CSC}}$ and $r_{\mathrm{DC},i} = X_i \cdot r_{\mathrm{DC}}$, where i refers to clone i, and X_i is taken from a Gaussian distribution $\mathcal{N}(1, \sigma_r)$ with σ_r the standard deviation of the division rate. Since we do not keep track of individual cells within the same clone, the division rate is implicitly inherited by all offspring of a clone.

In our simulations, the addition of heterogeneous division rates indeed resulted in progression of clonal dominance (Fig 4**A**), but this was accompanied by a much larger clone loss than observed with the K562 cells (Fig 4**B**). Major clones did develop during these simulations, but nearly all major clones were unique to their replicate (Fig 4**C**-**D**). This indicates that their abundance is not only caused by their division rate, but is also the result of surviving passage by chance. The clonal dominance that emerged in our simulations strongly depended on the division rate variability (Fig 4**E**). This variability also affected the amount of clones lost during the simulation, yet increasing it only led to a poorer match between the simulation and the *in vitro* results (Fig 4**F**). Further exploration of the model parameters revealed that a reduction of clone loss can be achieved when the initial CSC percentage (CSC_0) and the probability of symmetric CSC division (p_1) are increased (Fig 4**G**). In fact, the number of remaining clones best approaches the *in vitro* observations for

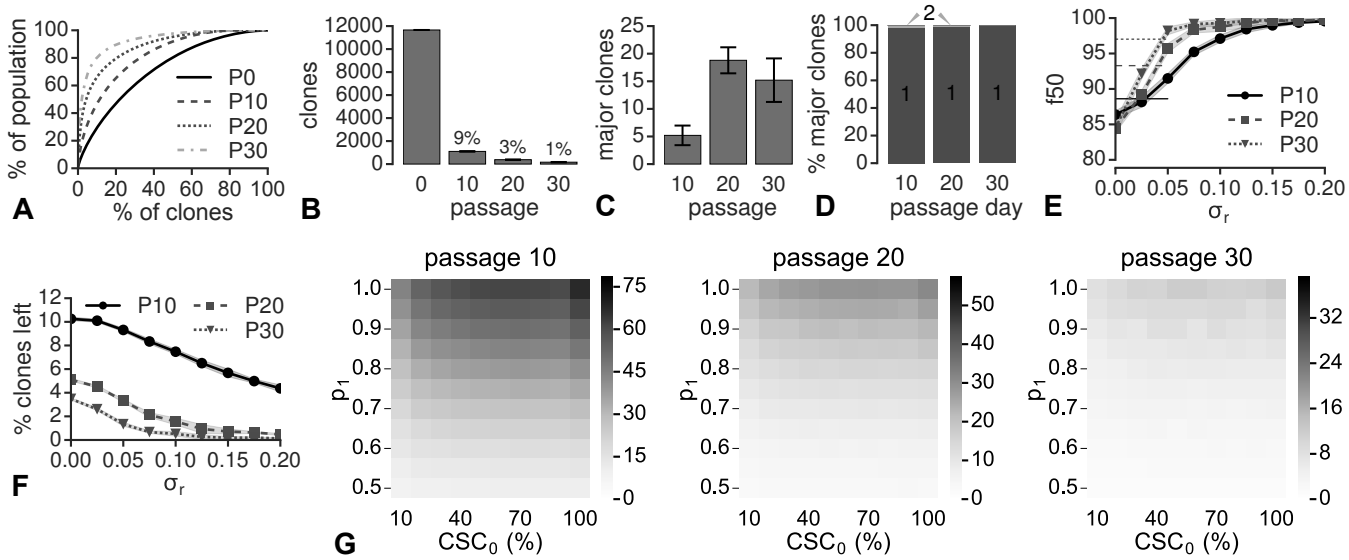

Figure 4: Simulation of the CSC growth model with division rate variability results in clonal dominance and massive clone loss. **A**-**D** Evolution of the clone distribution (**A**), clone number (**B**), major clones (**C**) and major clone overlap across replicates (**D**) for the CSC growth model with division rate variability ($\sigma_r = 0.05$ and CSC_0=20%). **E**-**F** Relation between the division rate standard deviation and clone distribution (**E**) or clone loss (**F**), with the horizontal lines in **E** denoting the corresponding experimental values for the polyclonal K562 cell line. **G** Clone loss, indicated by the heatmap color, for simulations with a varying initial CSC percentage (CSC_0) or with a varying 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 (**B**-**C**) and colored areas (**E**-**F**) represent the standard deviation.

$p_1 = 1$ and $\text{CSC}_0 = 100$, which means that all cells are CSCs that divide indefinitely. Interestingly, as for the *non-CSC* growth model with a uniform division rate, the number of clones left for passage 10 and 20 closely approached the experimental observations, while for passage 30, the simulated clone loss did not match the experimental observations. Together, these observations suggest that a further exploration of the non-CSC growth model with a variable growth rate is needed.

A model with indefinitely growing cells and varying division rates matches in vitro data

When all cells are CSCs that only divide into CSCs, the heterogeneous division rate is the only feature that differentiates the CSC model from the model in which all cells have an unlimited division capacity. This indicates that the presence of CSCs is not at all required for the development of clonal dominance. To test if such a model without CSCs indeed results in progressive clonal dominance, we introduced division rate variability to the earlier model where all cells have an unlimited division potential, according to $r_i = X_i \cdot r$. Indeed, division rate variation alone suffices

to induce clonal dominance that increases over time (Fig 5**A**), accompanied by moderate clone loss (Fig 5**B**) and the development of several dominant clones of which some are unique to their replicate and others occur in multiple replicates (Fig 5**C**-**D**). The speed of clonal disappearance and the level of clonal dominance depends on the standard deviation of the growth rate (Figure 5**E**-**F**), and not on the mean division rate (Fig S4). Besides the polyclonal K562 cell line, Porter *et al.* [15] tested several others, *i.e.* human embryonic kidney (HEK) 293T, HeLa, and human non-small-cell lung cancer (HCC827) polyclonal cell lines, and a monoclonal K562 cell line. Compared to the polyclonally, lentivirally barcoded K562 cell line (Fig 1) the HEK-293T cells exhibited a similar progression of clonal dominance, while the HeLa and HCC827 cell lines showed a more pronounced clonal dominance, and the clonal K562 cell line showed no clonal dominance. The shift towards an increased or decreased dominance can be obtained in the model by altering the standard deviation of the division rates (Fig 5**E**). Hence, it is possible to fine tune our model by searching for the division rate standard deviation that best fits the *in vitro* results for a particular cell line.

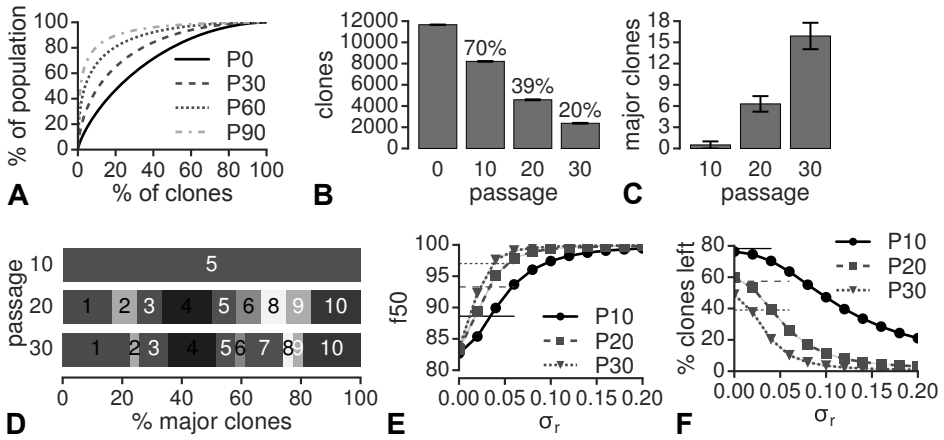

Figure 5: Simulation of growth model with indefinitely dividing cells and division rate heterogeneity results in clonal dominance and moderate clone loss which both depend on the division rate SD. **A**-**D** Evolution of the clone distribution (**A**), clone number (**B**), major clones (**C**), and major clone overlap (**D**) (with $\sigma_r = 0.04$), averaged over 10 simulations. **E**-**F** Relation between the growth rate standard deviation (σ_r) and f50 (**F**) and the number of remaining clones (**F**), with the horizontal lines denoting the corresponding experimental values for the polyclonal K562 cell line. Each data point represents the mean for 10 simulations and the error bars (**B**-**C**) and colored areas (**E**-**F**) represent the standard deviation.

We illustrate this fine-tuning procedure with the polyclonal K562 cell line by running a series of simulations for a narrow range of division rate standard deviations (10 replicates per standard deviation) and comparing the simulation results quantitatively with the *in vitro* data on the clone distribution (Fig 1**B**), the number of clones (Fig 1**C**), and the number of major clones (Fig 1**D**). By computing the mean squared error (MSE) for each evaluation criterion and the normalized average of the MSE of the three criteria, we obtained a measure of the deviation between the simulation result and the *in vitro* data. Because the results of the three biological replicates exhibited some variability, we computed the MSEs per biological replicate and interpreted the results as a parameter range (replicate A in Fig 6**A**-**D**, replicate B in S5**A**-**D**, and replicate C in

Fig S5**E**-**H**). The optimal fits range from $\sigma_r = 0.022$, for replicate B, to $\sigma_r = 0.026$, for replicates A and C. Indeed there is an excellent correspondence in the number of clones (Fig 6**E**), the clone distribution (Fig 6**F**), and the number of major clones (Fig 6**G**). Moreover, the overlap of major clones across replicates, which was not a criterion for the optimization of the division rate standard deviation, is close to the experimental observations (Fig 6H). Nevertheless, there are small differences: at passage 30, 87.5% of the experimental major clones were found in more than 1 out of 3 biological replicates (Fig 1**E**), while in the simulations only 31.25% of the major clones were found in more than one third of the replicates. Thus, our model can reproduce the experimentally observed clone distribution, clone loss, and development of major clones, but there are minor deviations with respect to the clonal overlap across replicates.

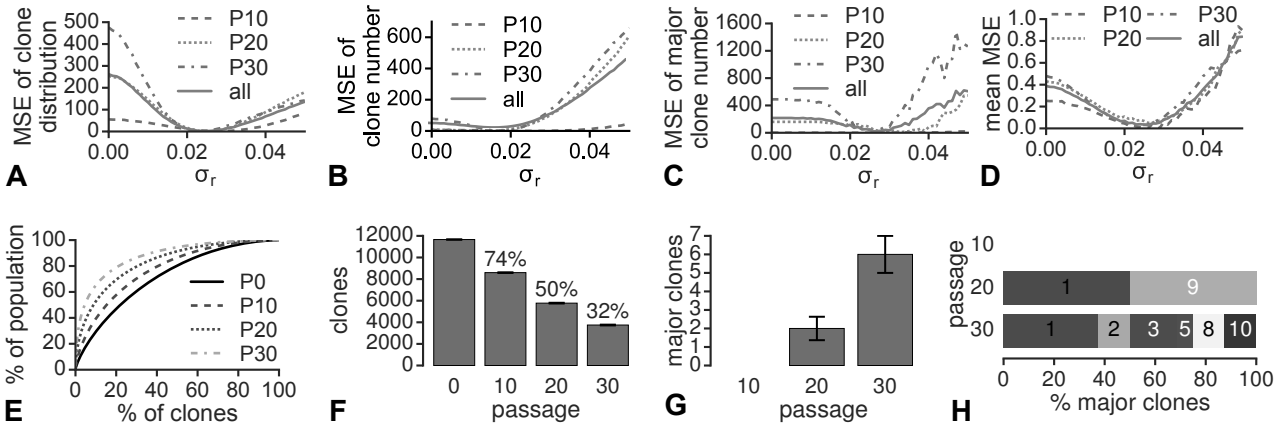

Figure 6: The division rate standard deviation can be fine-tuned to closely match the experimental observations for the polyclonal, lentivirally barcoded K562 cell line. **A**-**D** Mean squared error for the clone distribution (**A**), the number of clones (**B**), the number of major clones (**C**), and the normalized average of all three metrics (**D**) between 10 simulation replicates and biological replicate A. **E**-**H** simulation results for the optimal fit to replicate A ($\sigma = 0.026$), averaged over 10 simulation replicates with the error bars (**E** and **F**) depicting the standard deviation.

Discussion

Employing a simple model of stochastic cell division and passage, we showed that most of the clone loss observed during passage, observed by Porter *et al.* [15], can be explained by 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, the addition of a heterogeneous, heritable division rate resulted in a progressive clonal dominance that closely matched the experimental data. The amount of clone loss and the level of clonal dominance depended solely on the level of division rate heterogeneity in the model, which can be adapted to closely approximate the *in vitro* results for the polyclonal K562 cell line.

Altogether, our model provides strong evidence that CSCs are unlikely to have a role in shaping the changes in clonal distribution observed by Porter *et al.* [15]. This contradicts the conclusions drawn from numerous *in vivo*, microscopy-based, lineage tracing studies [12, 13], in which the development of large monochromatic patches from a mosaic pattern is ascribed to CSCs. Based on these observations, Driessens *et al.* [12] proposed a mathematical model with CSCs and DCs that closely fitted the clone sizes they observed in their experiments with skin papilloma. However, their model assumed that CSCs divided twice as fast as DCs, while CSCs are typically thought to divide slower or at a similar rate as DCs [7]. Furthermore, while such a model may be able to correctly predict the development of *in vivo* clone size, we showed that it could not match the clonal dynamics observed *in vitro*. Altogether, while monochromatic patches could be formed *in vivo* due to the presence of CSCs, we show that this is not the sole explanation for clonal dominance.

Although our model results do not point to a role for CSC in the experiments of Porter *et al.* [15], there is in general ample evidence for the existence of CSCs in other settings [4, 6, 7]. 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, 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 [7, 31]. This CSC plasticity could 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 may also allow for a good fit between simulations and *in vitro* observations.

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. Whereas classical studies have provided ample evidence for division rate heterogeneity among tumor cells [28–30], direct evidence for its full and direct inheritance has not yet been obtained. Nevertheless, Gray *et al.* [28] 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 [32]. However, other studies have shown that the division times of breast cancer cells correspond less between parents and offspring than between siblings [33], and that in lymphoblasts the division times of parents and offspring do not correlate at all [34]. These findings indicate that the child's division rate is not a direct copy of the parent's division rate. Sandler *et al.* [33] 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 [33]. 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 cannot perfectly match all *in vitro* observations of Porter *et al.* [15]. The most obvious difference between our simulations and the experimental data is the lack of variation between the simulations, while the experimental data is highly variable. The variability in the experimental data is readily explained by variable lab conditions, and by the quantification procedure of PCR and sequencing itself. Another discrepancy is the match between the model results and the results of two alternative experiments with K562 cells. The first experiment involves a monoclonal K562 cell line, in which clonal dominance did not develop and clone loss was limited. When we consider that all cells in the monoclonal K562 cell line have the same division rate, we can reproduce the lack of clonal dominance, but the associated clone loss is higher. Interestingly, the results that come closest to the experiments with monoclonal K562 cells are those of Porter's model with deterministic growth (Fig 2**A**-**B**), which could suggest that the cells in the monoclonal K562 cell population have a uniform division rate and synchronized cell cycle. The second experiment concerns the K562 cells barcoded using the zinc-finger technique that, in contrast to the lentiviral vector, allows barcodes to be inserted at a precise *target* location. As a result, the barcodes are inserted at a position where they hardly interfere with cell functions. In comparison with the lentivirally barcoded K562 cells, experiments with the targeted K562 cells quickly developed clonal dominance which barely progressed after the 10th passage, and results in a much stronger clone loss. From our simulations we know that clonal dominance develops faster when the simulation starts with a heterogeneous clone distribution, indicating that the barcode library used for the targeted K562 was likely more heterogeneous than the lentiviral barcode library. In addition, the pronounced clone loss in targeted K562 cells was possibly due to the presence of antibiotics in these experiments.

Altogether, in this work we used a computational approach to test two alternative hypothesis for the development of clonal dominance and showed that only one of the two, division rate heterogeneity, can reproduce the experimental observations. This conclusion contradicts the common thinking that CSCs drive clonal dominance. 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 [35]. With such a model it becomes possible to explore the consequences of division rate variability while comparing with intra-vital images studies.

References

[1] Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys. 1976;22(4):403–434.

[2] Gillespie DT. Approximate accelerated stochastic simulation of chemically\nreacting systems. J Chem Phys. 2001;115(4):1716–1733.

[3] 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.

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

[5] 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.

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

[7] Clevers H. The cancer stem cell: premises, promises and challenges. Nat Med. 2011;17(3):313–319.

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

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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).

[15] 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.

[16] 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.

[17] 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.

[18] 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).

[19] 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.

[20] 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.

[21] 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.

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

[23] Sottoriva A, Sloot PMa, Medema JP, Vermeulen L. Exploring cancer stem cell niche directed tumor growth. Cell Cycle. 2010;9(8):1472–1479.

[24] 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.

[25] Sottoriva A, Vermeulen L, Tavaré S. Modeling Evolutionary Dynamics of Epigenetic Mutations in Hierarchically Organized Tumors. PLoS Comput Biol. 2011 5;7(5):e1001132.

[26] 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.

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

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

[29] Dexter DL, Kowalski HM, Blazar Ba, Blazar A, Vogel R, Gloria H. Heterogeneity of Tumor Cells from a Single Mouse Mammary Tumor of Tumor Cells from a Single Mouse Mammary Tumor1. Cancer Res. 1978;38(10):3174–3181.

[30] Danielson KG, Anderson LW, Hosick HL. Selection and Characterization in Culture of Mammary Tumor Cells with Distinctive Growth Properties in Vivo Selection and Characterization in Culture of Mammary Tumor Cells with Distinctive Growth Properties in Vivo1. Cancer Res. 1980;40(6):1812–1819.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

Supplementary Figures

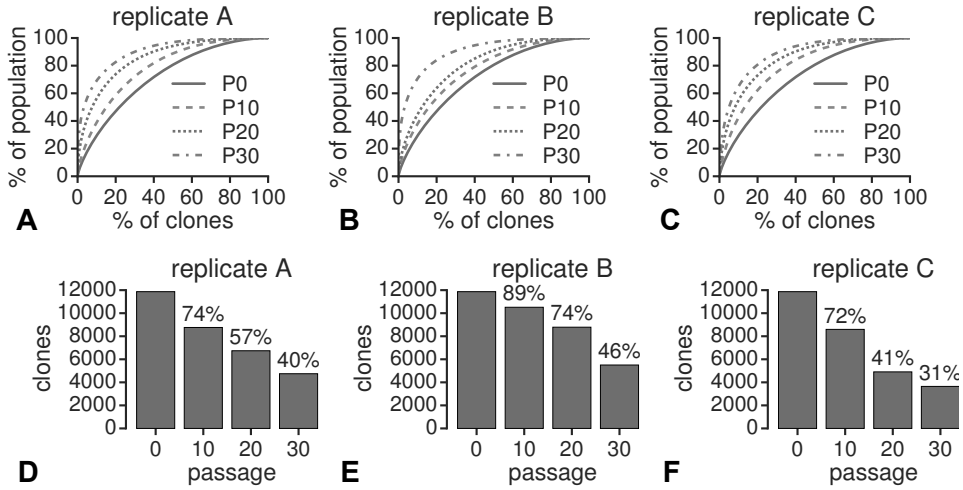

Figure S1: Overview of the three biological replicates of the polyclonal K562 cell line with barcodes inserted using lentiviral vectors. **A**-**C** Number of clones left at passage 10, 20, and 30. **D**-**F** Clone distribution at passage 10, 20, and 30.

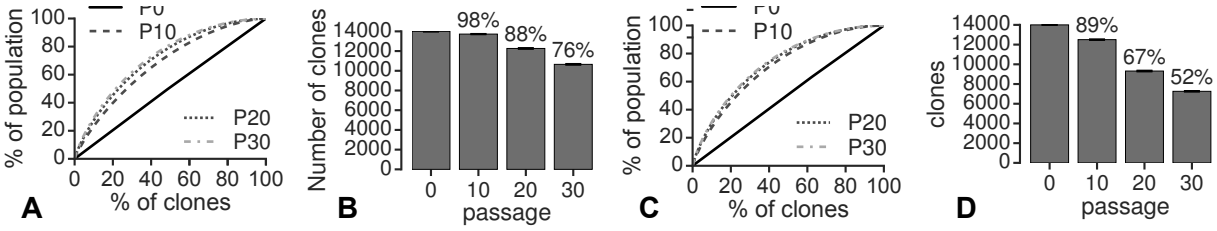

Figure S2: Evolution of the clone distribution with deterministic growth and stochastic passage (**A**-**B**), or with stochastic growth and passage (**C**-**D**). All values are the mean of 10 simulation replicates and the error bars in **A** and **C** represent the standard deviation.

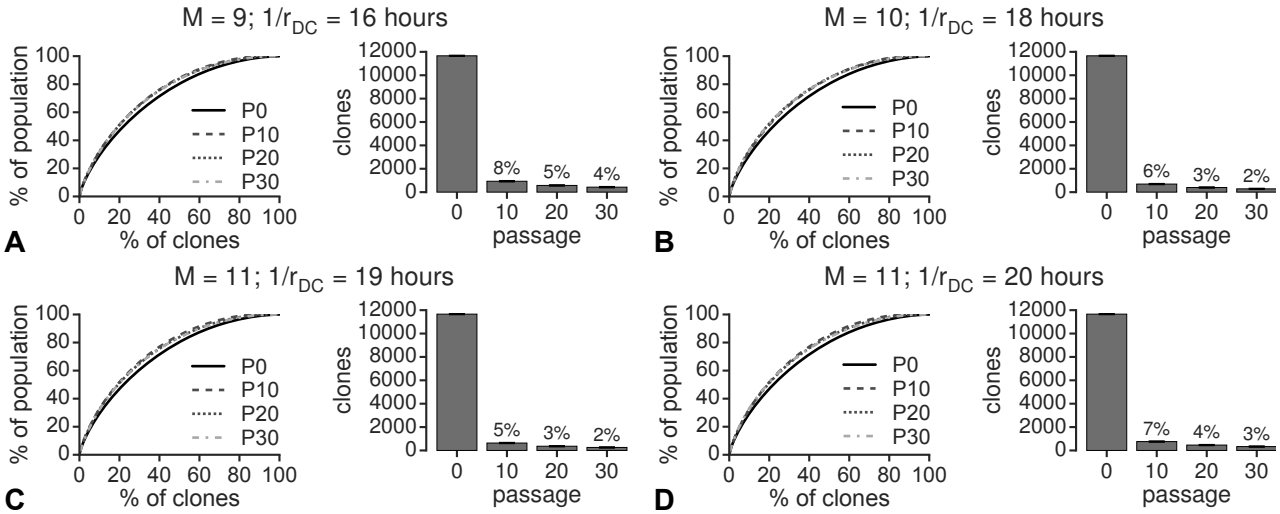

Figure S3: Evolution of the clone distribution and clone loss for the alternative parameters settings for M and r_{DC}, denoted by the black crosses in Fig 3**B** in the main text, with $m = 10$, $r_{\text{CSC}} = 1$ day^{-1}, $r_{\text{DC}} = \frac{19}{24}$ day^{-1}, $p_1 = 0.5$, $p_2 = 0.5$, $p_3 = 0$, $\text{CSC}_0 = 5\%$. All results are the mean of 10 simulation replicates and the error bars denote the standard deviation.

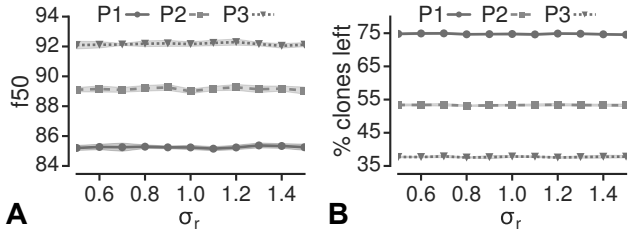

Figure S4: Effect of the mean division rate on the clone distribution (**A**) and clone loss (**B**). All points are the mean of 10 simulation replicates and the colored areas represent the standard deviation.

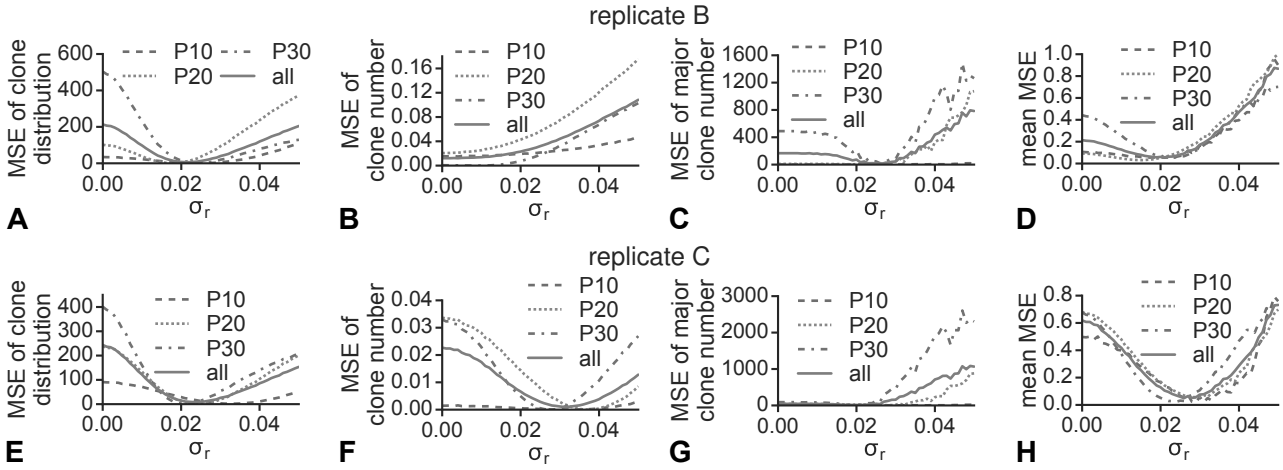

Figure S5: Comparison of model results and experimental data for biological replicates B and C. Mean squared error of the number of clones (**A** and **E**), the clone distribution (**C** and **F**), the number of major clones (**C** and **G**), and the normalized mean of the three metrics (**D** and **H**).