Mathematical Modeling of Clonal Interference by Density-Dependent Selection in Heterogeneous Cancer Cell Lines

Thomas Veith 1,2, Andrew Schultz 1,‡, Saeed Alahmari 3,‡, Richard Beck 1, Joseph Johnson 4 and Noemi Andor 1,2,*

1 Moffitt Cancer Center, Integrated Mathematical Oncology, USF Magnolia Drive, Tampa, FL 33612, USA; thomas.veith@moffitt.org (T.V.); andrew.schultz@moffitt.org (A.S.); richard.beck@moffitt.org (R.B.)
2 Department of Cell Biology, Microbiology, and Molecular Biology, University of South Florida, 4202 E Fowler Ave, Tampa, FL 33612, USA
3 Department of Computer Science, Najran University, King Abdulaziz Road, Najran 61441, Saudi Arabia; saeed3@usf.edu
4 Moffitt Cancer Center, Analytic Microscopy Core, USF Magnolia Drive, Tampa, FL 33612, USA; joseph.johnson@moffitt.org
* Correspondence: noemi.andor@gmail.com
‡ These authors contributed equally to this work.

Abstract: Many cancer cell lines are aneuploid and heterogeneous, with multiple karyotypes co-existing within the same cell line. Karyotype heterogeneity has been shown to manifest phenotypically, thus affecting how cells respond to drugs or to minor differences in culture media. Knowing how to interpret karyotype heterogeneity phenotypically would give insights into cellular phenotypes before they unfold temporally. Here, we re-analyzed single cell RNA (scRNA) and scDNA sequencing data from eight stomach cancer cell lines by placing gene expression programs into a phenotypic context. Using live cell imaging, we quantified differences in the growth rate and contact inhibition between the eight cell lines and used these differences to prioritize the transcriptomic biomarkers of the growth rate and carrying capacity. Using these biomarkers, we found significant differences in the predicted growth rate or carrying capacity between multiple karyotypes detected within the same cell line. We used these predictions to simulate how the clonal composition of a cell line would change depending on density conditions during in-vitro experiments. Once validated, these models can aid in the design of experiments that steer evolution with density-dependent selection.

Keywords: mathematical oncology; tumor evolution; intratumoral heterogeneity; density-dependent selection; life history theory; r/K selection

1. Introduction

Cellular heterogeneity is a defining feature of most cancers, and it is critical to tumor progression and treatment failure [1,2]. Advances in sequencing techniques have provided for an unprecedented depth of genetic profiling and ushered in a new era of individualized, data-driven cancer genetics [3]. Despite this progress, the relationship between genetic and phenotypic heterogeneity remains a significant gap in our current understanding.

One contributor to a cancer cell’s phenotype includes large-scale somatic copy number alterations (SCNAs) of 10 mega base pairs or more. Studies show that SCNAs correlate with progression rates and overall survival, with cancer cells grouped by copy number landscape exhibiting the same resistance to chemotherapy [4–7]. The SCNAs of whole chromosomes or chromosome arms, also known as aneuploidy, are a defining feature of many cancers [8]. Chromosomal instability (CIN) is a hallmark of cancer [9] and accounts for the vast majority of genetic material with an altered copy number state. CIN has been shown to promote metastasis and tumor evolution, particularly in cancers which have common aneuploidy patterns, such as gastric cancer [2,10]. Studies have shown that SCNAs activate
oncogenes, disrupt tumor suppressor genes [11], correlate with cancer phenotype [12–15], and can be spatially segregated within the tumor [16]. Aneuploidy fuels rapid phenotypic evolution and drug resistance, with similar karyotype profiles displaying resistance to the same drug [17]. The correlation between karyotypic and phenotypic divergence is not surprising: two cells with different karyotypes will differ in the expression of thousands of genes, which in turn will have broad phenotypic effects. However, how selection acts upon karyotypes remains poorly understood.

Here, we aim to characterize a targeted subset of the phenotypic differences between co-existing karyotypes, which are defined by their ability to out-compete each other at low vs. high cell densities. Cell densities naturally vary in the tumor over space and time, thus creating niches with distinct selection pressures. A cell that is in a densely packed environment will be under more pressure to overcome contact inhibition than a cell that finds itself in sparse conditions. This variation in evolutionary pressures likely contributes to the coexistence of cancer cells with heterogeneous phenotypes. Density-dependent selection occurs when fitness is a function of population density [18]. A related concept in ecology is life history theory [19] and the r/K selection framework, which investigates trade-offs between the number of offspring a species produces (growth rate or ‘r’) and the ability to compete in dense ecological niches (carrying capacity or ‘K’) [20,21].

We present a framework for examining how density-dependent selection acts on coexisting clones defined by karyotypes and apply it to a set of stomach cancer cell lines. Hereby, we focus on carrying capacity as defined by spatial limitations rather than metabolic constraints.

2. Results

2.1. Identifying Biomarkers of Growth Rate and Carrying Capacity

By sequencing the DNA and RNA of >36,000 cells from nine stomach cancer cell lines, our prior work classified cells into groups with unique karyotypes [22], which are further referred to as superclones [23] or clones. To compare the growth dynamics across cell lines, we grew eight of these stomach cancer cell lines in a T25 flask until they reached confluence (6–23 days); we then imaged them every day to count cells (Section 4).

We changed the culture media every 3.05 days on average, with media changes becoming more frequent as cells became more confluent. We compared cell counts derived from a Countess (Life Technologies Countess II FL Automated Cell Counter) to cell counts derived from live cell imaging to confirm segmentation accuracy (Appendix A Figure A6), thus supporting feasibility of monitoring growth dynamics during routine in-vitro experiments.

We fit the Gompertz growth model and two instances of the generalized logistic growth model (Richards and Verhulst) [24] to this time-series data (Figure 1A; see also Section 4.4). The resulting $R^2$ was high (Adj-$R^2 > 0.95$) for each of the models and cell lines (Appendix A Table A1). Using the Akaike Information Criterion, we determined that the Richards model slightly outperformed the Verhulst model, which in turn slightly outperformed the Gompertz model (Appendix A Figure A3). We thus eliminated the Gompertz model from consideration for model selection. Using likelihood profiling [25] to assess practical identifiability, we concluded that the noise levels in the data collected for 4/8 cell lines were too high to infer all the parameters of the Richards model (Appendix A Figure A4). Therefore, we used the simpler Verhulst model (commonly referred to as the “logistic function”) to infer growth dynamics from time-series data for all eight cell lines (see Figure 1, Appendix A Table A2).
Figure 1. Identifying biomarkers of growth and carrying capacity: (A) Logistic growth curves fit to the cell counts of eight gastric cancer cell lines at various stages of their growth. Fits were shifted along the x-axis such that the midpoint of each curve lay above $x = 0$. The eight cell lines differed in their maximum growth rate ($r$), as well as their maximum sustainable population size ($K$). (B) Summary statistics of linear regression models used to correlate KEGG pathway activity levels with logistic function growth parameters. The top five models and their performance in the training cell lines are shown for growth rate ($r$: light gray rows) and carrying capacity ($K$: dark gray). Columns: Pearson = Pearson correlation coefficient; $P$ = FDR corrected p-value; $R^2$ = adjusted $R^2$. (C,D) Relation between growth parameters ($r, K$) and the model with best performance in the validation dataset: ‘Amoebiasis’ (C) and ‘Arginine and proline metabolism’ (D), respectively. Pathway activity quantified using AUCell with scRNA-seq data. Error bars represent median absolute deviation. (E) Clonal composition confirmed by both scDNA- and scRNA-seq in eight gastric cancer cell lines (data taken...
Overall, there was no significant correlation between the growth rate and the carrying capacity across cell lines (Pearson’s $r = -0.37$, $p = 0.54$). In order to investigate the potential r/K trade-offs between clones within a cell line, we sought to find transcriptomic biomarkers of the inferred growth parameters in the scRNA-seq data available from our prior study [22]. Recently, investigators induced r/K-selection in HeLa cells [20]. The authors found that genes that were differentially expressed between r- and K-selected cells cultured at low densities were enriched among 25 pathways defined in the KEGG database [26]. We tested these pathways for their potential as biomarkers of the growth rate and carrying capacity in five of the eight gastric cancer cell lines (further referred to as the training set; Appendix A Figure A2). For each of the 25 pathways, we fitted a linear regression model to predict the growth parameters inferred for a given cell from the cell’s respective pathway activity level. Because the cell cycle is one of the strongest modulators of pathway activity [22,27,28], we also grouped the cells according to their assigned cell cycle state, thereby calculating the median pathway activity across the following: (i) G0G1 cells, (ii) S cells, (iii) G2M cells, and (iv) all cell cycle states combined. The pathways that had the strongest predictive value for the growth rate ($r$) included the ‘Amoebiasis during G2M’ (adj-$R^2 = 0.89$, $p < 0.02$), ‘Epstein-Barr virus infection’, and ‘AMPK signaling pathway during S-phase’. For the carrying capacity ($K$), the ‘PI3K-Akt signaling pathway’ (adj-$R^2 = 0.91$, $p < 0.01$) and ‘Arginine and proline metabolism’ had the strongest signals (Appendix A Figure A2). This process was repeated to include the remaining three cell lines (further referred to as the validation set), but only for these top performing pathways and cell cycle states (Figure 1B). Of these top pathways, 40% were confirmed in the validation set (FDR adjusted $p < 0.05$; Figure 1B), including ‘Amoebiasis during G2M’ and ‘Arginine and proline metabolism’, which were further used as proxies of the growth rate ($r$) and carrying capacity ($K$), respectively.

2.2. Towards Informing Future Experiments: Steering Clonal Evolution

The identification and characterization of co-existing clones within a cell line requires high-throughput assays, such as single-cell sequencing. Repeated measurements of a cell line’s clonal compositions at a high temporal resolution are thus cost-prohibitive. This underscores the need to identify biomarkers that can be used to estimate clonal growth parameters.

We previously identified 39 clones, which were each defined by distinct SCNAS and confirmed by both scDNA- and scRNA-seq data across the eight cell lines [22] (Figure 1E). Using pathway activity levels from the KEGG pathways ‘Amoebiasis’ and in ‘Arginine and proline metabolism’, we predicted the growth rate and carrying capacity for every sequenced cell in each cell line. By grouping cells according to their clone membership (Figure 1FG) we concluded that up to two r/K trade-offs may exist between clone pairs in two of the eight analyzed cell lines (Online Methods and Appendix A Table A3), namely in NCI-N87 and SNU-668 (Figure 2A).

While the growth conditions for SNU-668 can be modified to favor the r-selected clone, by the time it exceeds the K-selected clone, its frequency in the cell line is less than 0.001%. In this case, the r/K trade-off between the two clones has little practical relevance, because other clones in the cell line take over the population faster than the two clones can outcompete each other (Figure 2B). In contrast, the r/K trade-off identified in NCI-N87 did have practical relevance, with most cell culture schedules resulting in populations where at least one of the two clones maintains high cell representation (>5%; Figure 2C).
Figure 2. Using biomarkers of growth and contact inhibition to steer clonal evolution across passages: (A) Clone pairs with significant differences in both growth rates (r) and carrying capacities (K) were detected in two cell lines. Each color encodes for a single pair of clones (legend). Each bar is calculated as the ratio of r or K between the pair of clones shown in the legend. Horizontal line at 1 indicates identical parameters for both clones. Clone pairs with potential r/K trade-offs are represented by bars on both sides of the horizontal line. P-value of difference in r or K between a given pair of clones is indicated on the top of each bar. (B) Heat map showing outcome of clonal competition of the two clones shown in (A) for the SNU-668 cell line after 30 passages when grown at different seeding densities (x-axis) and passaging intervals (y-axis). Positive entries represent K selection (CloneK wins) and negative entries represent r selection (Cloner wins). The magnitude represents the size of the winning clone. Entries calculated as $\text{sgn}(\text{Clone}_K - \text{Clone}_r) \ast \log_{10}(\max(\text{Clone}_r, \text{Clone}_K))$. (C) Same as in (B), but for the NCI-N87 cell line. Note the presence of both large negative and large positive values for NCI-N87, thus indicating the practical relevance of the r/K trade-off for the former. Highlighted in red are the conditions used for simulations in (E). (D) Simulated growth of the three largest clones (color-coded) detected in the NCI-N87 cell line over 40 days (smallest clone excluded due to insufficient G2M cell representation). (E) Change in frequency of the two NCI-N87 clones shown in (A) over multiple passages (x-axis). Changing the seeding density and the timing of splitting the cells between passages (stars vs. circles in legend), we predict will accelerate or delay the decline of clone 119967 (dark color shades), from passage 4 (intersection of star-shaped curves) to after passage 30 (circle-shaped curves). Clonal frequencies at harvest set the initial frequency conditions for each subsequent seeding.

Optimizing the seeding density and the timing for splitting cells should thus enable evolutionary steering of the cell line’s clonal composition across passages. A condition for such an optimal time to exist is that growth of the population can never be negative, which
we prove analytically (Appendix A.2). We used the growth parameters predicted for the clones identified in the NCI-N87 cell line (Appendix A Table A3) to parameterize a multi-compartment ODE model to simulate how clonal composition changes over time (Figure 2B):

$$\frac{dN_i}{dt} = N_i \cdot r_i \left(1 - \sum_{i=1}^{n} \frac{N_i}{K_i}\right),$$

where $N_i$ is the cell representation of clone $i$, $K_i$ and $r_i$ are the biomarker-inferred carrying capacity and maximum growth rate of each clone, respectively (see also Section 4.4).

The simulations shown in Figure 2D predict that shifts in clonal composition will likely occur as NCI-N87 cells grow in-vitro, with the magnitude of these shifts depending on the seeding density and the timing of splitting cells (Figure 2C–E). For example, with a low seeding density and a splitting interval of 7 days, the r-selected clone outcompeted the K-selected clone after only four passages (Figure 2E). By contrast, a high seeding density and a splitting interval of only 1 day maintained a relative dominance of the K-selected clone over >30 passages (Figure 2E). This suggests that, for a subset of cell lines, density conditions during in-vitro experiments can be optimized to either accelerate or delay changes in clonal composition.

3. Discussion

Our work rests on the shoulders of Uri Ben-David et al.’s landmark paper [29], which examined the responsiveness of 21 different variants of the same cell line with 321 anti-cancer agents. They found that 75% of the tested compounds that strongly inhibited some variants were inactive in others and that copy number changes (which are dominated by differences in karyotype) explained most of this differential phenotypic response. Two follow-up questions their work raises are: “what is the mechanism of clonal interference for a specific culture condition?” and “can we predict the clonal composition that will emerge from a given culture condition?”. Our work contributes to begin answering these multi-faceted questions. Through a re-analysis of existing scRNA-seq data from eight stomach cancer cell lines [22], we confirmed the biomarkers of growth rate and contact inhibition as previously identified in breast cancer cell lines [20], thus suggesting similarities in the mechanisms of contact inhibition exist across multiple tissue and cancer types. Differential expression of biomarkers for growth rate and contact inhibition/carrying capacity between clones within a cell line (Figure 1) suggest that different population densities will alter a cell line’s clonal composition. However, the presence of significant $r/K$ trade-offs within cell lines was limited to only two clone pairs (Figure 2A). Most experimental protocols will not see the cells spend a long time at high density and will thus tend to select for cells with higher growth rates in the exponential phase, thus potentially explaining relative lack of significant $r/K$ trade-offs within cell lines. Despite that, the predicted intra-cell line variability in both carrying capacity and maximum growth rate (Figure 1F,G) heralds ongoing changes in cellular composition, which are in line with changes in clonal evolution we observed previously over only 5–7 passages in a sub-set of these cell lines [22].

Our study focuses on a naturally fluctuating feature of the cellular microenvironment: population density. We therefore define phenotypic differences between karyotypes as their ability to out-compete each other at low vs. high population densities. However, we do not expect these traits to be the only phenotypic difference between the identified karyotypes. The expression magnitude of thousands of genes distinguishes any two karyotypes, and, with them, we expected multiple phenotypes to also differ. However, in the absence of any other differences in growth conditions (such as drug exposure), most phenotypic differences between karyotypes will be latent [29,30], thus allowing us to observe their relatively subtle differential growth under variable density conditions. We consider these differences to be only the tip of the iceberg among all phenotypic differences between co-existing karyotypes. However, the realization that even the most basic cell culture habits (seeding density and passaging frequency) can influence long-term cellular dynamics, we expect, will prompt
the scientific community to prioritize developing databases that routinely record detailed cell culture protocols.

Clones that are far from each other in r–K space may also be spatially segregated and have differential treatment sensitivities. For example, the slower growth rate of r-selected cells may render them more resistant to cytotoxic therapies [31]—an effect that should be amplified by their spatial segregation [32,33]. We would expect K-selected sub-populations to be closer to the center of the tumor mass, where they are also more protected from drug penetration. By contrast, the enhancing tumor should be dominated by r-selected cells [31]. This scenario opens the door to an emerging concept of cancer treatment called adaptive therapy [34]. Adaptive regimens aim to optimize doses and dose schedules to maintain a fixed population of therapy-sensitive cells such that these, in turn, can suppress the growth of resistant cells [34]. Biomarkers that quantify the representation of r- and K-selected sub-populations could thus contribute to the design of adaptive regimens that spare a narrow border of r-selected cells such that the K-selected population stays enclosed [31].

The growing field of spatial transcriptomics makes it possible to measure the expression levels of thousands of genes throughout tissue space [35]. The ability to predict a cell’s placement along the r–K phenotype continuum from its transcriptome could thus facilitate the spatial de-lineation of r- from K-selected sub-populations in primary tumor tissue sections. Testing these ideas would require expanding our models into a spatial in-vivo framework. However, an understanding of the spatial distribution of r- and K-selected sub-populations could facilitate the design of therapy regimens, such as adaptive therapy or double bind therapy [36], aiming to stabilize the overall tumor burden to prevent or at least delay therapy resistance.

Our in silico results add to the evidence accumulating from many other studies [22,29,37] that every cell division is an opportunity for the cells to mutate and adapt to their environment. This insight underscores the need for routine tracking of the pedigree of evolving cell populations over decades, along with potentially changing environments (e.g., cell culture habits and therapy). Such efforts could help reveal long-term trends in the evolution of cell lines that remain elusive at shorter time-scales.

4. Online Methods
4.1. Cell Culture

The identity of each cell line was determined through independent karyotyping and mycoplasma contamination assessment. Cells were cultured in their recommended media conditions at 37 °C. For, HGC-27, EMEM (Quality Biological Inc., Gaithersburg, MD, USA); KATOIII, RPMI-1640, and EMEM (1:1 mix) were used; for NCI-N87, RPMI-1640 (ATCC modified) was used; the remaining five cell lines (MKN-45, NUGC-4, SNU-601, SNU-638, and SNU-668) used RPMI-1640. All cell lines were grown in the aforementioned media with 10% fetal bovine serum (Gibco, Carlsbad, CA, USA) and 1% penicillin—streptomycin (Gibco). For cell passaging, Trypsin EDTA (0.25%) with phenol red (Gibco) was added, followed by inactivation using the respective growth media. Cells were seeded at a low initial density (range: 4.98–38.64 × 10^3 cells/cm^2) in order to allow for the population’s exponential growth and subsequent plateau at the carrying capacity (range: 1.92–15.32 × 10^5 cells/cm^2).

The end-point for each experiment was determined by fitting a generalized logistic growth curve to the in-vitro data at various points during the time in culture. Across 8 cell lines, the time in culture ranged from 6–21 days, with media being refreshed on average every 48 h until the cells reached confluence, whereupon media was refreshed every 24 h. The experiment was stopped once the fold change in the inferred carrying capacity was less than 4% when removing any subset of the last 20% of time points.

4.2. Microscopy

Cells were seeded in a T25 flask (Fisherbrand), and sets of 4 phase contrast images were taken at 20× magnification for NCI-N87, and 10× magnification for the aforementioned cell lines on an Evos FL. Segmentation using Cellpose was utilized to quantify the cell
count and cellular features (number of detections, centroid X µm, centroid Y µm, ROI, area m², and perimeter µm). Growth curves were generated to examine the carrying capability of each cell line.

**Image preprocessing:** The microscopy image acquisition and light settings resulted in variations in the brightness of acquired images. The automatic segmentation and counting pipeline applied to dark images showed inaccurate results (i.e., mostly false negatives). Therefore, we applied pre-processing steps to our pipeline, which consisted of the following: (i) gamma correction to reduce darkness; (ii) histogram equalization to improve image contrast, and (iii) Gaussian blurring for smoothing.

**Cell segmentation:** Fully automated cell segmentation of phase contrast images was performed using Cellpose software.

Cellpose is a deep learning approach based on the U-Net architecture for cell segmentation [38,39], where vertical and horizontal spatial gradients of cells are predicted. Furthermore, Cellpose predicts a binary map of a cell location either inside or outside a region of interest (ROI). Using both the combined vertical and horizontal gradients and the binary map, the cell localization and generation of binary masks for every cell are performed. Our pipeline for cell segmentation using Cellpose used 2D microscopic images as input. An ROI annotation rectangle was applied to the phase contrast image (left corner index of the rectangle: (100,100); rectangle width and height: (1800 px, 1100 px). This ROI was used for subsequent image segmentation and analysis. We fine-tuned a pre-trained Cellpose model called (Cytotorch_2) for learning to segment cells on given microscopic images of different cell lines. After learning the segmentation using Cellpose, the trained model was tested on a hold-out set for evaluation. Then, feature extraction of each segmented cell was performed for further analysis of cell growth at a given time point. The features extracted from each cell segmentation included area, perimeter, roundness, and centroid. The pipeline saved the extracted features and visualization of the segmentation onto a user-desired folder for cell growth estimation and analysis.

**4.3. Determination of Clonal Growth Parameters**

Determining relative clonal carrying capacities, growth rates, and loss of contact inhibition parameter values is a necessary first step in modeling density-dependent selection in heterogeneous cell lines. To achieve this, we used gene expression signatures as surrogates of growth parameters as previously described [40,41].

**Quantification of single cell pathway activity from gene expression:**

In order to identify cells with active gene sets, we utilized AUCell version 3.18 for R version 4.3.1 [42]. AUCell takes as its input the scRNA sequencing data for the cells of interest and a list of gene sets. The output is the gene set activity for each cell. AUCell uses the area under the curve across the rankings of all genes of a particular cell, where genes are ranked by their expression value. A rank-based scoring method means AUCell is not affected by units or normalization methods of the gene expression data. Genes in the top 5% of the ranking are considered active in a given gene set. In order to account for potential batch effects, scRNA sequencing data for all cell lines were analyzed together. Cells with less than 200 features were excluded. Seurat version 2.3.4 was used to create a Seurat object for input to AUCell to quantify the activity of more than 2000 pathways from the KEGG database. For all further analysis, we focused on only a subset of 25 KEGG pathways, which were previously identified as being differentially expressed between r-selected and K-selected HeLa cells [20].

**Clonal growth parameters:**

We tested these 25 KEGG pathways quantified by AUCell as biomarkers of growth rate (r) and carrying capacity (K). The test takes the form of a linear model:

$$\delta \sim a \ast p + c,$$

(2)
where $\delta$ is the parameter value ($r$ or $K$), and $p$ is the pathway activity. The pathways are then ranked by adjusted $R^2$, and the top five best-fitting pathways are prioritized as potential biomarkers of a given parameter.

With the linear models correlating pathway activity with growth parameters built at the cell line level, we can predict growth parameters for all clonal populations using their pathway activity as input (Figure 1E,F, Appendix A Table A3). Once parameters have been predicted, clonal growth can be simulated as systems of paired ODEs (see Section 4.4).

### 4.4. Mathematical Models of In-Vitro Cell Growth

The nature of tumor growth is not well known, and the exact laws which govern the growth of tumor cells will likely be context-dependent (cancer type, location in the body, etc.). However, even when many of these dependencies are held constant, it is difficult to discern between various models of cancer cell growth [43–46]. We fit generalized logistic growth models [24] to the time series cell growth data for eight gastric cancer cell lines (Appendix A Table A1):

$$\frac{dN_i}{dt} = N_i * r_i \left(1 - \left(\frac{\sum_{i=1}^{n} N_i}{K_i}\right)^p\right),$$

where $i$ is either the entire cell line population (i.e., $n = 1$) or one of multiple clones within a cell line (i.e., $n \geq 3$). $K_i$ and $r_i$ are the carrying capacity and maximum growth rate of each population, respectively. $v$ is the loss of contact inhibition, which is assumed to be identical for all clones within a cell line. Two instances of Equation (3) were fitted to the data: one representing Richards growth model where $v$ was kept variable and the other representing the logistic growth model with $v := 1$.

In addition to the two models mentioned above, we also fitted cell line population growth with the Gompertz model [24]. These models were chosen due to their prevalence in the literature for describing the growth of tumors and their biological interpretation. The growth data was read into R and fit using the package growthrates [47]. In order to compare the quality of growth-model fits, we calculated the adjusted-$R^2$ for each of the three models across all eight cell lines (Appendix A Table A1) We then compared Akaike Information Criterion scores (Appendix A Figure A3) and found the Richards model to have the lowest score in 7/8 cell lines. However, identifiability analysis revealed that we were unable to confidently infer the growth rate ($r$) and/or loss of contact inhibition ($v$) parameters in 4/8 cell lines (Appendix A Figures A4 and A5). Thus, we modeled in-vitro growth using the logistic growth model. To model clonal growth dynamics, we used the ODE45 solver for MatLab version 9.7.0.1216025 to solve Equation (3) after being parameterized using the values in Appendix A Table A3.

### 4.5. Identifying r/K Trade-Offs between Co-Existing Clones within a Cell Line

Let $C = \{N_1, ... N_n\}$ be the set of $n$ clones identified within a given cell line. For each pair $\{x, y| x, y \in C \land x \neq y\}$ we use a Student’s $t$-test to compare the growth rates predicted for cells assigned to clone $x$ ($r_x$) vs. cells assigned to clone $y$ ($r_y$). We do the same for the carrying capacity predicted for cells of the two clones ($K_x, K_y$). For clone pairs with a $p$-value $\leq 0.1$ for both $r$ and $K$, we further calculate $\tau_x(x, y) = \frac{r_x}{r_y}$ and $\tau_y(x, y) = \frac{K_x}{K_y}$. We define $(x^*, y^*)$ as clones with potential $r/K$ trade-offs:

$$(x^*, y^*) = \{x, y| (\tau_x(x, y) < 1 \land \tau_y(x, y) > 1) \lor (\tau_x(x, y) > 1 \land \tau_y(x, y) < 1)\}. \quad (4)$$

**Author Contributions:** Conceptualization, T.V. and N.A.; methodology, T.V., S.A., A.S., J.J. and N.A.; software, S.A. and N.A.; validation, T.V., S.A., R.B. and N.A.; formal analysis, T.V., S.A. and N.A.; investigation, T.V. and N.A.; resources, N.A.; data curation, T.V., S.A., A.S. and N.A.; writing—original draft preparation, T.V., S.A. and N.A.; writing—review and editing, T.V. and N.A.; visualization, T.V., S.A. and N.A.; supervision, N.A.; project administration, N.A.; funding acquisition, N.A. All authors have read and agreed to the published version of the manuscript.
Funding: This work was supported by the National Institute of Health at the National Cancer Institute (1R37CA266727-01A1 to N.A.) and by the Moffitt Cancer Center Evolutionary Therapy Center of Excellence (CET 30-20458-03-20 to N.A.). The funders had no role in the study design, data collection and analysis, decision to publish, or the preparation of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Supporting data for the reported results, as well as the code that generates the main figures in this manuscript, can be found at https://www.github.com/MathOnco/densityDependentSelection.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Identifiability Analysis

Identifiability analysis is a group of methods used to determine how well the parameters of a model are estimated [25]. The likelihood function is one method that can be used to identify parameter values that are more likely than others. In other words, it tells us which parameter values could be expected to produce the data to which we are fitting the model. Profiling likelihoods is achieved by constructing a function that minimizes the negative log likelihood for a fixed value of a parameter of interest, and it returns $z$, which is the signed square-root deviance from the minimum. At the maximum likelihood estimation, $z = 0$ by definition.

Let the vector $\theta$ be the group of parameters used in Equation (3) to make predictions. Then, the likelihood function can be written as $L(\theta| x) = f_\theta(x)$. Suppose that $\theta$ can be decomposed as $\theta = (\delta, \xi)$, where $\delta$ is the parameter of interest, and $\xi$ is a vector containing the nuisance parameters. We can re-write our likelihood function as $L(\delta, \xi| x) = f_{\delta, \xi}(x)$. By profiling, we are concentrating the likelihood function for a subset of parameters by expressing the nuisance parameters as a function of the parameter of interest and replacing the nuisance parameter in the likelihood function. Thus, the profile likelihood is calculated as follows:

$$L_p(\delta) = \sup_{\xi} L(\delta, \xi, x).$$ (A1)

Let $\epsilon$ be the error between observations and model fit, as defined above, and let the growth rate ($r$) be our parameter of interest so that $\delta := r$ and $\xi := (K, v)$. Then, in each direction away from $\arg\max L(\delta)$, we fix $\delta$ by $\pm s.d.(\epsilon)$, and we adjust $\xi$ such that $z$ is minimized. This process continues until $z$ stops changing for new values of $\delta$ or is equal to the MLE fit error. Then, $z$ approximates a chi-squared distribution on which we can compute the confidence intervals. We define parameters with confidence intervals that have a $>2$-fold difference between the 2.5% and 97.5% quantiles as unidentifiable. This analysis revealed that the growth rate ($r$) or loss of contact inhibition ($v$) parameters were unidentifiable in the Richards model in 4/8 cases (Appendix A Figure A4, Appendix A Figure A5). Thus, we excluded the Richards model and selected the logistic model, given its better fits and lower AIC scores compared to the Gompertz model (Appendix A Table A1 and Figure A3).

Appendix A.2. Analytical Results Optimizing Cell Passaging for Logistic Growth Model (Two Clone Case)

We identified two r/K trade-offs between clone pairs across eight cell lines (Figure 2A, Appendix A Table A3). This suggests that, for these cell lines, an optimal time for splitting the cells exists where the trade-off is balanced. Optimizing the timing for splitting the cells should thus stabilize a cell line’s clonal composition over multiple passages. A condition for such an optimal time to exist is that the growth of the population can never be negative. Here, we prove that a heterogeneous cell line with a positive growth rate, consisting of two clones with an r/K trade-off, will never have a negative growth rate.
Let \( i \in \{1 \ldots n\} \) be a clone within a cell line consisting of \( n \) clones. Let each clone have its own specific growth rate, \( r_i \), and carrying capacity, \( K_i \). Then, the change in the number of cell members, \( N_i \), of clone \( i \) can be modeled as follows:

\[
\frac{dN_i}{dt} = (N_i \cdot r_i) \cdot \left(1 - \frac{\sum_{j=1}^{n} N_j}{K_i}\right),
\]

which has the following solution:

\[
N_i(t) = \frac{N_i \cdot K_i}{\sum_{j=1}^{n} N_j + (K_i - \sum_{j=1}^{n} N_j) \cdot e^{-r_i t}}.
\]

In order to arrive at the solution above in (A3), we start by separating the variables in (A2), thereby obtaining our \( dt \) on the LHS and the \( dN_i \) on the RHS.

\[
(r_i)dt = \frac{dN_i}{N_i \cdot (1 - \frac{\sum_{j=1}^{n} N_j}{K_i})}.
\]

Next, we expand the RHS of (A4) through partial fraction decomposition, where \( \Gamma \) and \( \Psi \) are unknown constants.

\[
\frac{1}{N_i \cdot (1 - \frac{\sum_{j=1}^{n} N_j}{K_i})} = \frac{\Gamma}{N_i} + \frac{\Psi}{(1 - \frac{\sum_{j=1}^{n} N_j}{K_i})}.
\]

We multiply both sides of (A5) by the LHS denominator to simplify, giving us

\[
1 = \Gamma \cdot [1 - \frac{\sum_{j=1}^{n} N_j}{K_i}] + (\Psi \cdot N_i).
\]

Then, by solving for \( \Psi \) and \( \Gamma \) we obtain the following:

\[
\Psi = \frac{1}{N_i} - \frac{\sum_{j=1}^{n} N_j}{N_i} \cdot \frac{\Gamma \cdot \sum_{j=1}^{n} N_j}{N_i \cdot K_i} \Rightarrow \Gamma = 1.
\]

Now, we can return to integrating the following:

\[
\int r_i dt = \int \frac{dN_i}{N_i} + \int \frac{\frac{\sum_{j=1}^{n} N_j}{N_i} \cdot \frac{\Gamma \cdot \sum_{j=1}^{n} N_j}{N_i \cdot K_i}}{1 - \frac{\sum_{j=1}^{n} N_j}{K_i}}.
\]

The LHS and first term of RHS are easy enough to solve. For the second term of the RHS, we need to integrate through \( u \) substitution, where \( u = 1 - \frac{\sum_{j=1}^{n} N_j}{K_i} \) and \( du = \frac{-1}{K_i} \). We then obtain the following:

\[
\int \frac{-du}{u} = -ln[u] = -ln[1 - \frac{\sum_{j=1}^{n} N_j}{K_i}].
\]

Then, by putting all the terms together and accounting for the constant of integration \( C \), we obtain the following:

\[
(r_i \cdot t) + C = \ln[\frac{N_i}{1 - \frac{\sum_{j=1}^{n} N_j}{K_i}}].
\]

Next, we exponentiate both sides:

\[
C \cdot e^{(r_i \cdot t)} = \frac{N_i}{1 - \frac{\sum_{j=1}^{n} N_j}{K_i}}.
\]
We find how $C = e^C$ relates to the initial conditions:

$$C = \frac{N_i \cdot K_i}{K_i - \sum_{j=1}^{n} N_j}.$$  \hfill (A12)

We plug $C$ into (A11) and solve for $N$:

$$N_i(t) = \frac{N_i \cdot K_i}{\sum_{j=1}^{n} N_j + (K_i - \sum_{j=1}^{n} N_j) \cdot e^{-(r_i t)}}.$$  \hfill (A13)

Now, we can move from population sizes to frequencies by letting $p = \frac{N}{\sum_{i=1}^{n} N_i}$ and $T = \sum_{i=1}^{n} (N_i)$, where $p$ is the proportion of one of our subclones $N_i$, and $T$ is the total population. Thus, the equations that describe how $p$ and $T$ change over time, in the two clone cases, are as follows:

$$\frac{dp}{dt} = p \cdot (1 - p)[r_a \cdot (1 - \frac{T}{K_a}) - r_b \cdot (1 - \frac{T}{K_b})]$$  \hfill (A14)

$$\frac{dT}{dt} = T \cdot [(p \cdot r_a)(1 - \frac{T}{K_a}) + (1 - p)(r_b)(1 - \frac{T}{K_b})],$$  \hfill (A15)

where $K_a$ and $K_b$ are the carrying capacity of clone (a) and (b), respectively, and $r_a$ and $r_b$ are the respective growth rates.

Next, we sought to prove that Equation (A15) is always positive. To do so, we started out by assuming the opposite—that there does exist some time point where $T' < 0$—and arrived at the proof by contradiction. Please note that we are most interested in the situation where there is an $r/K$ trade-off. That is, for the rest of this exercise, please assume $r_a > r_b$ and $K_a < K_b$. Let us also assume that our two populations are well-mixed and initially comprise no more than 60 percent of $\max(K_a, K_b)$.

If we let

$$\frac{dT}{dt} = T \cdot [(p \cdot r_a)(1 - \frac{T}{K_a}) + (1 - p)(r_b)(1 - \frac{T}{K_b})] = 0,$$  \hfill (A16)

and solve for $T$, we obtain the following:

$$T = \frac{(K_a \cdot K_b)[(r_b \cdot p) - r_b - (r_a \cdot p)]}{(K_a \cdot r_b \cdot p) - (K_a \cdot r_b) - (K_b \cdot r_a \cdot p)}.$$  \hfill (A17)

Equation (A17) is the $T$ nullcline. It is easy to see that this function, $T(p)$, is one-to-one. If $T(p_1) = T(p_2)$, then $p_1 = p_2$. The nullcline being a one-to-one function implies that every input that can bring the system to zero will only bring the system to zero. Now, we have $T$ totally in terms of $p$, $r_a, r_b, K_a$, and $K_b$. Next, we can plug it into Equation (A14) to obtain it defined in terms of $p$ and the aforementioned $r$ and $K$ parameters:

$$\frac{dp}{dt} = \{p \cdot (1 - p)[r_a \cdot (1 - \frac{K_a[(r_b \cdot p) - r_b - (r_a \cdot p)]}{(K_a \cdot r_b \cdot p) - (K_a \cdot r_b) - (K_b \cdot r_a \cdot p)} - r_b \cdot (1 - \frac{K_a[(r_b \cdot p) - r_b - (r_a \cdot p)]}{(K_a \cdot r_b \cdot p) - (K_a \cdot r_b) - (K_b \cdot r_a \cdot p)})] - (K_a \cdot K_b)[(r_b \cdot p) - r_b - (r_a \cdot p)]]/(K_a \cdot r_b \cdot p) - (K_a \cdot r_b) - (K_b \cdot r_a \cdot p),$$  \hfill (A18)

which is always less than or equal to zero. Thus, the system can never cross the nullcline given by $T' = 0$. We know that our system exists in the $T' > 0$ space due to the initial conditions and positive $r$ and $K$ terms, but it can never cross into $T' < 0$ space. This contradicts our original assumption that there existed a timepoint $t$ such that $T'(t) < 0$. Hence $T'$ is always positive. The vector field given by $T'$ and $p'$ with the nullcline $T' = 0$ (Appendix A Figure A1) shows that the $T$ nullcline passes the Horizontal Line Test, thus further demonstrating its injective nature. For visualization purposes, the total population,
$T$ (the x-axis), has been scaled to $\max(K_a, K_b) = K_b = 1$. Accordingly, each of our variables, $p$ and $T$, are scaled by $p = p * K_b$ and $T = T * K_b$.

Figure A1. Growth dynamics for the two clone case. Vector field given by Equations (A17) and (A18). The red curve shows where there is no change in total population (given by the nullcline $T' = 0$). Arrows indicate total population growth ($T'$). Velocities calculated as $\sqrt{T'(t)^2 + A'(t)^2}$. Velocity is always zero at the nullcline, thus showing that the system can never cross from $T' > 0$ into $T' < 0$ space.

Table A1. Adjusted-$R^2$ for model fits across cell lines.

<table>
<thead>
<tr>
<th></th>
<th>HGC-27</th>
<th>KATOIII</th>
<th>MKN-45</th>
<th>NCI-N87</th>
<th>NUGC-4</th>
<th>SNU-601</th>
<th>SNU-638</th>
<th>SNU-668</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>1.00</td>
<td>0.97</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Richards</td>
<td>1.00</td>
<td>0.97</td>
<td>0.99</td>
<td>0.99</td>
<td>0.97</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>Gompertz</td>
<td>1.00</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
<td>0.98</td>
<td>0.98</td>
<td>0.95</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table A2. Inferred parameters for best logistic models across cell lines.

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>K</th>
<th>bestFit</th>
<th>adjR Squared</th>
<th>doubling Time</th>
<th>cellLine</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGC-27</td>
<td>0.96</td>
<td>5,095,556.00</td>
<td>Logistic</td>
<td>1.00</td>
<td>1.10</td>
<td>HGC-27</td>
</tr>
<tr>
<td>KATOIII</td>
<td>0.44</td>
<td>4,904,316.00</td>
<td>Logistic</td>
<td>0.97</td>
<td>4.70</td>
<td>KATOIII</td>
</tr>
<tr>
<td>MKN-45</td>
<td>0.63</td>
<td>9,237,232.00</td>
<td>Logistic</td>
<td>0.99</td>
<td>3.00</td>
<td>MKN-45</td>
</tr>
<tr>
<td>NCI-N87</td>
<td>0.45</td>
<td>38,306,195.00</td>
<td>Logistic</td>
<td>0.99</td>
<td>2.50</td>
<td>NCI-N87</td>
</tr>
<tr>
<td>NUGC-4</td>
<td>0.92</td>
<td>6,899,073.00</td>
<td>Logistic</td>
<td>0.98</td>
<td>1.10</td>
<td>NUGC-4</td>
</tr>
<tr>
<td>SNU-601</td>
<td>0.51</td>
<td>7,374,909.00</td>
<td>Logistic</td>
<td>0.99</td>
<td>2.30</td>
<td>SNU-601</td>
</tr>
<tr>
<td>SNU-638</td>
<td>0.71</td>
<td>4,791,667.00</td>
<td>Logistic</td>
<td>0.97</td>
<td>1.20</td>
<td>SNU-638</td>
</tr>
<tr>
<td>SNU-668</td>
<td>0.45</td>
<td>5,048,178.00</td>
<td>Logistic</td>
<td>0.98</td>
<td>1.90</td>
<td>SNU-668</td>
</tr>
<tr>
<td>Pathway/Condition</td>
<td>K: adj.r.squared</td>
<td>K: p.value</td>
<td>r: adj.r.squared</td>
<td>r: p.value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>------------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI3K-Akt signaling pathway: GOG1</td>
<td>0.91</td>
<td>0.01</td>
<td>-0.21</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI3K-Akt signaling pathway: all</td>
<td>0.72</td>
<td>0.04</td>
<td>-0.28</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine and proline metabolism: G2M</td>
<td>0.62</td>
<td>0.07</td>
<td>-0.29</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine and proline metabolism: S</td>
<td>0.62</td>
<td>0.07</td>
<td>-0.27</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine and proline metabolism: all</td>
<td>0.60</td>
<td>0.08</td>
<td>-0.27</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amoebiasis: G2M</td>
<td>-0.33</td>
<td>0.94</td>
<td><strong>0.89</strong></td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epstein-Barr virus infection: GOG1</td>
<td>-0.32</td>
<td>0.88</td>
<td><strong>0.86</strong></td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epstein-Barr virus infection: all</td>
<td>-0.33</td>
<td>0.91</td>
<td><strong>0.80</strong></td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPK signaling pathway: S</td>
<td>0.28</td>
<td>0.21</td>
<td><strong>0.72</strong></td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPK signaling pathway: GOG1</td>
<td>0.07</td>
<td>0.33</td>
<td><strong>0.70</strong></td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPK signaling pathway: all</td>
<td>0.16</td>
<td>0.27</td>
<td>0.68</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epstein-Barr virus infection: G2M</td>
<td>-0.29</td>
<td>0.78</td>
<td>0.66</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA degradation: G2M</td>
<td>-0.33</td>
<td>0.98</td>
<td>0.61</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPK signaling pathway: G2M</td>
<td>0.59</td>
<td>0.08</td>
<td>0.39</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine and proline metabolism: GOG1</td>
<td>0.55</td>
<td>0.09</td>
<td>-0.25</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI3K-Akt signaling pathway: S</td>
<td>0.52</td>
<td>0.10</td>
<td>-0.23</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spliceosome: GOG1</td>
<td>0.31</td>
<td>0.19</td>
<td>-0.26</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spliceosome: S</td>
<td>0.31</td>
<td>0.19</td>
<td>-0.27</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI3K-Akt signaling pathway: G2M</td>
<td>0.31</td>
<td>0.19</td>
<td>-0.33</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathways in cancer: S</td>
<td>0.25</td>
<td>0.22</td>
<td>0.10</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base excision repair: G2M</td>
<td>0.22</td>
<td>0.24</td>
<td>-0.30</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axon guidance: all</td>
<td>0.21</td>
<td>0.25</td>
<td>0.05</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axon guidance: GOG1</td>
<td>0.21</td>
<td>0.25</td>
<td>-0.07</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathways in cancer: G2M</td>
<td>0.20</td>
<td>0.25</td>
<td>0.24</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base excision repair: all</td>
<td>0.20</td>
<td>0.25</td>
<td>-0.25</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spliceosome: all</td>
<td>0.18</td>
<td>0.26</td>
<td>-0.13</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axon guidance: G2M</td>
<td>0.08</td>
<td>0.33</td>
<td>0.15</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base excision repair: GOG1</td>
<td>0.07</td>
<td>0.34</td>
<td>-0.02</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathways in cancer: all</td>
<td>0.05</td>
<td>0.35</td>
<td>0.22</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axon guidance: S</td>
<td>0.03</td>
<td>0.36</td>
<td>0.42</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathways in cancer: GOG1</td>
<td>-0.02</td>
<td>0.41</td>
<td>0.30</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine metabolism: all</td>
<td>-0.04</td>
<td>0.43</td>
<td>-0.02</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influenza A: G2M</td>
<td>-0.05</td>
<td>0.44</td>
<td>-0.05</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteasome: GOG1</td>
<td>-0.06</td>
<td>0.45</td>
<td>-0.04</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spliceosome: G2M</td>
<td>-0.07</td>
<td>0.46</td>
<td>-0.07</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine metabolism: GOG1</td>
<td>-0.07</td>
<td>0.46</td>
<td>0.06</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement and coagulation cascades: S</td>
<td>-0.08</td>
<td>0.47</td>
<td>-0.33</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influenza A: G2M</td>
<td>-0.09</td>
<td>0.47</td>
<td>-0.07</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influenza A: all</td>
<td>-0.09</td>
<td>0.47</td>
<td>-0.02</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine metabolism: G2M</td>
<td>-0.10</td>
<td>0.48</td>
<td>-0.05</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p53 signaling pathway: G2M</td>
<td>-0.10</td>
<td>0.48</td>
<td>0.26</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrimidine metabolism: G2M</td>
<td>-0.10</td>
<td>0.48</td>
<td>0.08</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influenza A: GOG1</td>
<td>-0.10</td>
<td>0.48</td>
<td>0.03</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrimidine metabolism: all</td>
<td>-0.14</td>
<td>0.53</td>
<td>0.08</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p53 signaling pathway: S</td>
<td>-0.14</td>
<td>0.53</td>
<td>-0.29</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell cycle: S</td>
<td>-0.15</td>
<td>0.54</td>
<td>0.47</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure A2. Cont.**
<table>
<thead>
<tr>
<th>Pathway Description</th>
<th>r1</th>
<th>r2</th>
<th>r3</th>
<th>r4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complement and coagulation cascades: G2M</td>
<td>-0.15</td>
<td>0.54</td>
<td>-0.33</td>
<td>0.92</td>
</tr>
<tr>
<td>Pyrimidine metabolism: G0G1</td>
<td>-0.15</td>
<td>0.54</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>Purine metabolism: S</td>
<td>-0.16</td>
<td>0.55</td>
<td>0.00</td>
<td>0.39</td>
</tr>
<tr>
<td>Complement and coagulation cascades: all</td>
<td>-0.16</td>
<td>0.56</td>
<td>-0.32</td>
<td>0.90</td>
</tr>
<tr>
<td>Hepatitis B: G2M</td>
<td>-0.17</td>
<td>0.56</td>
<td>0.01</td>
<td>0.38</td>
</tr>
<tr>
<td>Epstein-Barr virus infection: S</td>
<td>-0.17</td>
<td>0.56</td>
<td>0.45</td>
<td>0.13</td>
</tr>
<tr>
<td>Complement and coagulation cascades: G0G1</td>
<td>-0.18</td>
<td>0.58</td>
<td>-0.32</td>
<td>0.86</td>
</tr>
<tr>
<td>p53 signaling pathway: all</td>
<td>-0.19</td>
<td>0.59</td>
<td>-0.21</td>
<td>0.62</td>
</tr>
<tr>
<td>p53 signaling pathway: G0G1</td>
<td>-0.20</td>
<td>0.61</td>
<td>-0.22</td>
<td>0.63</td>
</tr>
<tr>
<td>Hippo signaling pathway: G2M</td>
<td>-0.21</td>
<td>0.62</td>
<td>0.12</td>
<td>0.30</td>
</tr>
<tr>
<td>Hepatitis B: S</td>
<td>-0.21</td>
<td>0.62</td>
<td>-0.05</td>
<td>0.43</td>
</tr>
<tr>
<td>Cell cycle: all</td>
<td>-0.21</td>
<td>0.62</td>
<td>0.31</td>
<td>0.19</td>
</tr>
<tr>
<td>Pertussis: all</td>
<td>-0.21</td>
<td>0.63</td>
<td>0.08</td>
<td>0.33</td>
</tr>
<tr>
<td>Hippo signaling pathway: G0G1</td>
<td>-0.22</td>
<td>0.63</td>
<td>0.15</td>
<td>0.28</td>
</tr>
<tr>
<td>Pertussis: G0G1</td>
<td>-0.22</td>
<td>0.63</td>
<td>0.08</td>
<td>0.33</td>
</tr>
<tr>
<td>Hippo signaling pathway: S</td>
<td>-0.23</td>
<td>0.64</td>
<td>0.09</td>
<td>0.32</td>
</tr>
<tr>
<td>Small cell lung cancer: S</td>
<td>-0.23</td>
<td>0.64</td>
<td>-0.33</td>
<td>0.53</td>
</tr>
<tr>
<td>Proteasome: S</td>
<td>-0.23</td>
<td>0.64</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>Hippo signaling pathway: all</td>
<td>-0.23</td>
<td>0.64</td>
<td>0.10</td>
<td>0.32</td>
</tr>
<tr>
<td>Ribosome biogenesis in eukaryotes G2M</td>
<td>-0.23</td>
<td>0.65</td>
<td>-0.24</td>
<td>0.67</td>
</tr>
<tr>
<td>Pertussis: S</td>
<td>-0.24</td>
<td>0.66</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>Proteasome: all</td>
<td>-0.24</td>
<td>0.66</td>
<td>0.00</td>
<td>0.39</td>
</tr>
<tr>
<td>Small cell lung cancer: G2M</td>
<td>-0.26</td>
<td>0.70</td>
<td>-0.28</td>
<td>0.74</td>
</tr>
<tr>
<td>Pyrimidine metabolism: S</td>
<td>-0.26</td>
<td>0.71</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>ECM-receptor interaction: G0G1</td>
<td>-0.27</td>
<td>0.72</td>
<td>-0.03</td>
<td>0.42</td>
</tr>
<tr>
<td>ECM-receptor interaction: S</td>
<td>-0.27</td>
<td>0.73</td>
<td>-0.08</td>
<td>0.47</td>
</tr>
<tr>
<td>RNA transport: S</td>
<td>-0.28</td>
<td>0.74</td>
<td>0.46</td>
<td>0.13</td>
</tr>
<tr>
<td>Hepatitis B: all</td>
<td>-0.28</td>
<td>0.75</td>
<td>-0.04</td>
<td>0.42</td>
</tr>
<tr>
<td>Pertussis: G2M</td>
<td>-0.28</td>
<td>0.76</td>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis (ALS): G0G1</td>
<td>-0.29</td>
<td>0.76</td>
<td>-0.21</td>
<td>0.62</td>
</tr>
<tr>
<td>ECM-receptor interaction: all</td>
<td>-0.29</td>
<td>0.78</td>
<td>-0.13</td>
<td>0.51</td>
</tr>
<tr>
<td>Hepatitis B: G0G1</td>
<td>-0.30</td>
<td>0.80</td>
<td>-0.01</td>
<td>0.40</td>
</tr>
<tr>
<td>Base excision repair: S</td>
<td>-0.30</td>
<td>0.81</td>
<td>-0.17</td>
<td>0.57</td>
</tr>
<tr>
<td>RNA transport: G0G1</td>
<td>-0.30</td>
<td>0.81</td>
<td>0.51</td>
<td>0.11</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis (ALS): all</td>
<td>-0.30</td>
<td>0.81</td>
<td>-0.18</td>
<td>0.58</td>
</tr>
<tr>
<td>RNA transport: all</td>
<td>-0.31</td>
<td>0.84</td>
<td>0.56</td>
<td>0.09</td>
</tr>
<tr>
<td>Amoebiasis: G0G1</td>
<td>-0.32</td>
<td>0.85</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>RNA degradation: G0G1</td>
<td>-0.32</td>
<td>0.87</td>
<td>0.48</td>
<td>0.12</td>
</tr>
<tr>
<td>Small cell lung cancer: all</td>
<td>-0.32</td>
<td>0.89</td>
<td>-0.33</td>
<td>0.91</td>
</tr>
<tr>
<td>Cell cycle: G2M</td>
<td>-0.32</td>
<td>0.89</td>
<td>-0.31</td>
<td>0.82</td>
</tr>
<tr>
<td>Ribosome biogenesis in eukaryotes: S</td>
<td>-0.32</td>
<td>0.89</td>
<td>-0.33</td>
<td>0.91</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis (ALS): G2M</td>
<td>-0.32</td>
<td>0.89</td>
<td>-0.13</td>
<td>0.51</td>
</tr>
<tr>
<td>RNA degradation: all</td>
<td>-0.32</td>
<td>0.90</td>
<td>0.52</td>
<td>0.10</td>
</tr>
<tr>
<td>Ribosome biogenesis in eukaryotes: all</td>
<td>-0.33</td>
<td>0.91</td>
<td>-0.33</td>
<td>0.93</td>
</tr>
<tr>
<td>Amoebiasis: S</td>
<td>-0.33</td>
<td>0.91</td>
<td>0.56</td>
<td>0.09</td>
</tr>
<tr>
<td>Cell cycle: G0G1</td>
<td>-0.33</td>
<td>0.92</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis (ALS): G0G1</td>
<td>-0.33</td>
<td>0.92</td>
<td>-0.12</td>
<td>0.50</td>
</tr>
<tr>
<td>RNA transport: G2M</td>
<td>-0.33</td>
<td>0.93</td>
<td>0.52</td>
<td>0.10</td>
</tr>
<tr>
<td>Amoebiasis: all</td>
<td>-0.33</td>
<td>0.93</td>
<td>0.55</td>
<td>0.09</td>
</tr>
<tr>
<td>ECM-receptor interaction: G2M</td>
<td>-0.33</td>
<td>0.94</td>
<td>-0.26</td>
<td>0.70</td>
</tr>
<tr>
<td>Small cell lung cancer: G0G1</td>
<td>-0.33</td>
<td>0.96</td>
<td>-0.29</td>
<td>0.76</td>
</tr>
<tr>
<td>Ribosome biogenesis in eukaryotes: G0G1</td>
<td>-0.33</td>
<td>0.99</td>
<td>-0.33</td>
<td>0.97</td>
</tr>
<tr>
<td>Proteasome: G2M</td>
<td>-0.33</td>
<td>0.99</td>
<td>-0.03</td>
<td>0.42</td>
</tr>
<tr>
<td>RNA degradation: S</td>
<td>-0.33</td>
<td>0.99</td>
<td>0.60</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Figure A2. Statistics evaluating the transcriptional activity of 25 KEGG pathways as biomarkers of growth rate (columns with prefix ‘r’) and carrying capacity (columns with prefix ‘K’). The 25 pathways...
were chosen because they were identified as differentially expressed between r/K-selected HeLa cells under low-density culture conditions [20]. For each pathway, a linear regression model was fitted to predict the respective growth parameter (r or K) from pathway activity level measured in the same cell line (columns with suffix ‘r.squared’ and ‘p.value’). Analysis was performed across five gastric cancer cell lines used as training dataset (HGC-27, KATOIII, NCI-N87, NUGC-4, and SNU-668). The top 5 pathways’ predictive growth rates and top five pathways’ predictive carrying capacities (names in **bold**) were prioritized during training and further used in Figure 1B for validation with an extended dataset. p-values are color-coded depending on whether the respective pathway and growth parameter were positively (purple) or negatively (blue) correlated in the training dataset.

**Table A3.** Predicted growth parameters for clones (rows) identified in eight gastric cancer lines. Parameter predictions come from linear models built between pathway biomarkers and inferred parameter values at the cell line level. The input is clonal pathway activity, and the output is the predicted growth parameter for that clonal population.

<table>
<thead>
<tr>
<th>cloneID</th>
<th>K</th>
<th>r</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>102944</td>
<td>4,690,105</td>
<td>0.49</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102945</td>
<td>4,674,250</td>
<td>0.38</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102948</td>
<td>3,450,777</td>
<td>0.53</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102950</td>
<td>4,331,258</td>
<td>0.66</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102951</td>
<td>5,488,043</td>
<td>0.53</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102952</td>
<td>4,379,298</td>
<td>NA</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102954</td>
<td>5,046,408</td>
<td>0.62</td>
<td>SNU-668</td>
</tr>
<tr>
<td>102955</td>
<td>4,773,230</td>
<td>0.52</td>
<td>SNU-668</td>
</tr>
<tr>
<td>106394</td>
<td>3,166,668</td>
<td>0.56</td>
<td>KATOIII</td>
</tr>
<tr>
<td>106396</td>
<td>3,117,464</td>
<td>NA</td>
<td>KATOIII</td>
</tr>
<tr>
<td>106399</td>
<td>3,274,342</td>
<td>0.53</td>
<td>KATOIII</td>
</tr>
<tr>
<td>106404</td>
<td>3,346,502</td>
<td>0.58</td>
<td>KATOIII</td>
</tr>
<tr>
<td>112380</td>
<td>7,448,893</td>
<td>0.32</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112382</td>
<td>8,148,176</td>
<td>0.34</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112387</td>
<td>7,231,439</td>
<td>NA</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112389</td>
<td>6,627,640</td>
<td>0.32</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112392</td>
<td>6,798,104</td>
<td>0.34</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112399</td>
<td>6,312,316</td>
<td>NA</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112402</td>
<td>8,050,488</td>
<td>0.38</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112404</td>
<td>7,081,546</td>
<td>0.31</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112408</td>
<td>8,255,791</td>
<td>0.37</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112410</td>
<td>7,903,615</td>
<td>NA</td>
<td>SNU-601</td>
</tr>
<tr>
<td>112413</td>
<td>7,196,330</td>
<td>0.32</td>
<td>SNU-601</td>
</tr>
<tr>
<td>114525</td>
<td>11,055,814</td>
<td>0.65</td>
<td>MKN-45</td>
</tr>
<tr>
<td>114530</td>
<td>10,785,452</td>
<td>0.68</td>
<td>MKN-45</td>
</tr>
<tr>
<td>119963</td>
<td>18,156,049</td>
<td>NA</td>
<td>NCI-N87</td>
</tr>
<tr>
<td>119965</td>
<td>22,002,556</td>
<td>0.62</td>
<td>NCI-N87</td>
</tr>
<tr>
<td>119967</td>
<td>20,354,070</td>
<td>0.57</td>
<td>NCI-N87</td>
</tr>
<tr>
<td>119968</td>
<td>18,548,305</td>
<td>0.67</td>
<td>NCI-N87</td>
</tr>
<tr>
<td>122360</td>
<td>5,156,522</td>
<td>0.80</td>
<td>SNU-638</td>
</tr>
<tr>
<td>122361</td>
<td>4,331,370</td>
<td>NA</td>
<td>SNU-638</td>
</tr>
<tr>
<td>122363</td>
<td>4,759,109</td>
<td>0.80</td>
<td>SNU-638</td>
</tr>
<tr>
<td>125616</td>
<td>9,123,026</td>
<td>0.78</td>
<td>NUGC-4</td>
</tr>
<tr>
<td>125618</td>
<td>9,655,395</td>
<td>0.69</td>
<td>NUGC-4</td>
</tr>
<tr>
<td>125619</td>
<td>8,845,460</td>
<td>0.82</td>
<td>NUGC-4</td>
</tr>
<tr>
<td>129343</td>
<td>9,429,235</td>
<td>0.85</td>
<td>HGC-27</td>
</tr>
<tr>
<td>129344</td>
<td>8,996,528</td>
<td>0.82</td>
<td>HGC-27</td>
</tr>
<tr>
<td>129345</td>
<td>8,856,699</td>
<td>0.76</td>
<td>HGC-27</td>
</tr>
<tr>
<td>129346</td>
<td>8,616,362</td>
<td>0.78</td>
<td>HGC-27</td>
</tr>
</tbody>
</table>
Figure A3. Difference in AIC scores across cell lines. Different AIC scores were calculated using the dAIC function from the package bbmle for the statistical programming language R. For 7/8 cell lines, the Richards model had the lowest AIC score. Thus, the added parsimony could be justified by the increase in the goodness of fit.

### Difference in Akaike Information Criterion scores across cell lines

<table>
<thead>
<tr>
<th></th>
<th>HGC-27</th>
<th>KATOIII</th>
<th>MKN-45</th>
<th>NCI-N87</th>
<th>NUGC-4</th>
<th>SNU-601</th>
<th>SNU-638</th>
<th>SNU-668</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richards (3 params)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Logistic (2 params)</td>
<td>151.4</td>
<td>54.6</td>
<td>22.1</td>
<td>17</td>
<td>7.8</td>
<td>3.8</td>
<td>16.8</td>
<td>77</td>
</tr>
<tr>
<td>Gompertz (2 params)</td>
<td>86.5</td>
<td>116.8</td>
<td>101.2</td>
<td>34.2</td>
<td>0</td>
<td>105.6</td>
<td>36.6</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Figure A4. Likelihood profile confidence intervals across parameters and model types. Confidence intervals generated from likelihood profiles for all parameters in Richards and logistic (Verhulst) growth models. Entries with >2-fold difference between the 2.5% and 97.5% quantiles highlighted in red. Growth rate (r) is not identifiable for NUGC-4 or SNU-668 in the Richards model. Additionally, loss of contact inhibition (v) is not identifiable for those two cell lines: NCI-N87, and SNU-638.
Figure A5. Likelihood profiles across cell lines. Graphical representation of the likelihood profiles for both parameters in the logistic growth model for all cell lines. Please note here that mumax is the growth rate (r). Maximum likelihood values are at the center of the x-axis, with the y-axis giving the absolute value of the z statistic from the chi-square distribution these profiles approximate. Confidence intervals given at various percentages. Tight confidence intervals tell us when parameters are identifiable.
Figure A6. Segmentation evaluation using Countess. Comparing Countess-derived cell counts to cell counts derived from live cell imaging for each cell line. Results do not appear to be confounded by flask type (color-code in legend).

References


15. Li, B.; Jiang, Y.; Li, G.; Fisher, G.A.; Li, R. Natural killer cell and stroma abundance are independently prognostic and predict gastric cancer chemotherapy benefit. *JCI Insight* 2020, 5, e136570. [CrossRef] [PubMed]


32. Aubry, M.; de Tayrac, M.; Etcheverry, A.; Clavreul, A.; Saikali, S.; Menei, P.; Mosser, J. ‘From the core to the border beyond the margin’: A genomic picture of glioblastoma intratumor heterogeneity. *Oncotarget* 2015, 6, 12094–12109. [CrossRef]


44. Murphy, H.; Jaafari, H.; Dobrovolny, H.M. Differences in predictions of ODE models of tumor growth: A cautionary example. *BMC Cancer* 2016, 16, 163. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.