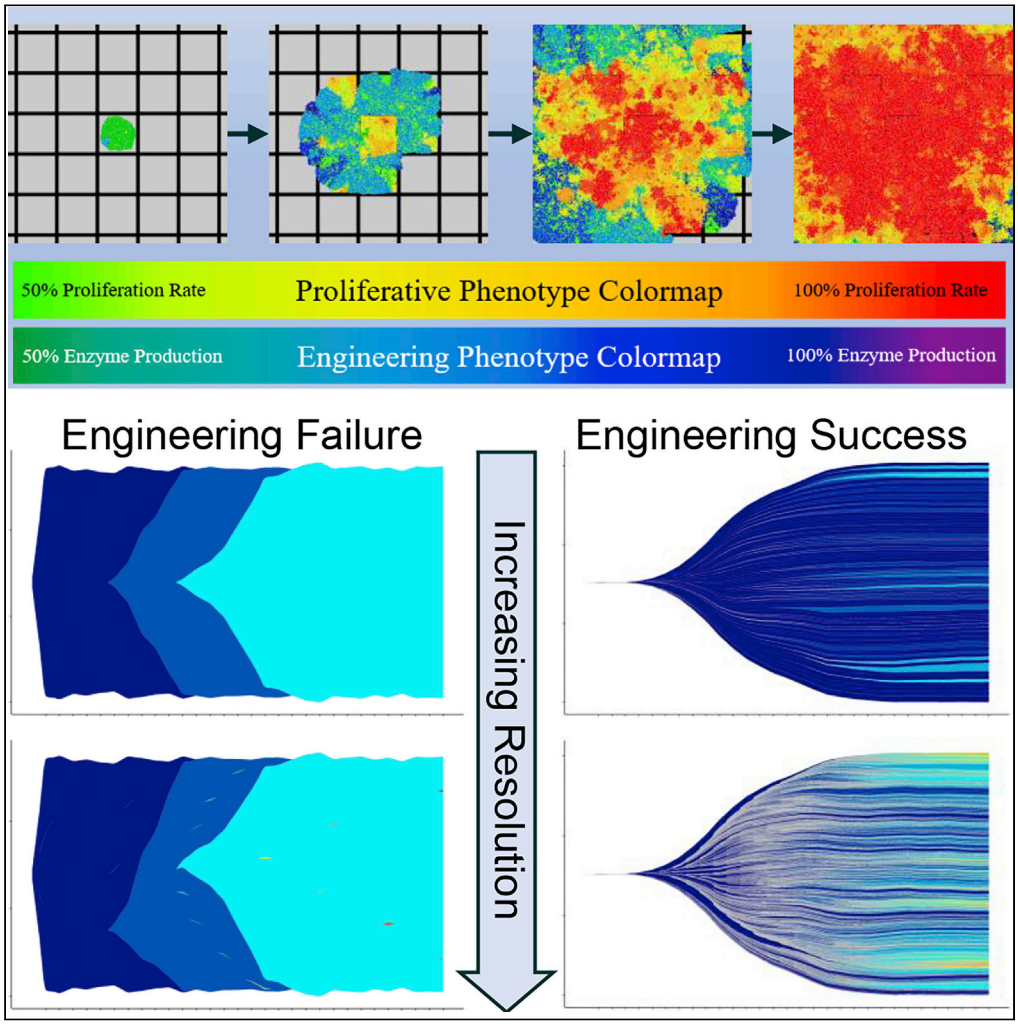
iScience

CellPress
OPEN ACCESS

Article

Selection-driven tumor evolution with public goods leads to patterns of clonal expansion consistent with neutral growth



Jack Edwards,
Andriy Marusyk,
David Basanta

jackedwards1@usf.edu (J.E.)
david@cancerevo.org (D.B.)

HIGHLIGHTS

Using agent-based models, we explored eco-evolutionary dynamics of tumors in space

We examined two phenotypes: one for accessing new space and one for consuming space

We observed ecological succession in simulations, resulting in tumors with high ITH

Model encodes selection explicitly but results could be misconstrued as neutral

Edwards et al., iScience *24*, 101901
January 22, 2021
https://doi.org/10.1016/j.isci.2020.101901

iScience

CellPress
OPEN ACCESS

Article

Selection-driven tumor evolution with public goods leads to patterns of clonal expansion consistent with neutral growth

Jack Edwards,[1,3,*] Andriy Marusyk,[2] and David Basanta[1,*]

SUMMARY

Cancers are the result of eco-evolutionary processes fueled by heritable phenotypic diversification and driven by environmentally dependent selection. Space represents a key growth-limiting ecological resource, the ability to explore this resource is likely under strong selection. Using agent-based modeling, we explored the interplay between phenotypic strategies centered on gaining access to new space through cell-extrinsic degradation of extracellular matrix barriers and the exploitation of this resource through maximizing cell proliferation. While cell proliferation is a cell-intrinsic property, newly accessed space represents a public good, which can benefit both producers and non-producers. We found that this interplay results in ecological succession, enabling emergence of large, heterogeneous, and highly proliferative populations. Even though in our simulations both remodeling and proliferation strategies were under strong positive selection, their interplay led to sub-clonal architecture that could be interpreted as evidence for neutral evolution, warranting cautious interpretation of inferences from sequencing of cancer genomes.

INTRODUCTION

Cancer starts when a somatic cell breaks away from the rules of homeostatic cooperation imposed by multicellularity and starts acting as an independent evolutionary entity. As this cell divides, its progeny (clone) acquires genetic and epigenetic alterations generating novel phenotypic variants (subclones). This subclonal diversification in the context of competition for growth-limiting resources between cancer cells enables Darwinian evolution, which can lead to the emergence of key tumor phenotypes, defined by hallmarks of cancers, culminating in lethal malignancies (Axelrod et al., 2006; Greaves and Maley, 2012; Hanahan and Weinberg, 2000, 2011; Korolev et al., 2014; Nowell, 1976). Consistently, genetic and phenotypic intratumor heterogeneity (ITH) correlates with poor prognosis and contributes to treatment failure (Gerlinger et al., 2012; Swanton, 2012). Understanding the basic rules that govern the somatic evolution and how ITH is maintained in the face of subclonal competition is thus key to a more effective tumor prevention and therapies (Basanta and Anderson, 2013; Sottoriva et al., 2015).

Mathematical modeling has been used to understand the behavior of complex dynamic systems, such as changes in cell populations in space and time during the emergence and progression of cancers, which are difficult to intuit without the help from rigorous theoretical frameworks. Properly parameterized and experimentally tested mathematical models can provide a deeper understanding of these complex processes and also serve as predictive tools for therapy optimization. Ordinary differential equations, the most commonly used tools to model somatic evolution, typically assume well-mixed populations (Altrock et al., 2015; Gatenby and Gillies, 2008). This assumption fails to account for heterogeneity in space and tissue architecture as a key growth limiting resource, which has a profound impact on evolutionary dynamics. Agent-based models (ABMs), which can consider cells as individual agents rather than populations, can be used to account for space and spatial structures. The importance of space and the applicability of ABMs has been explored showing that an increase in available space results in higher ITH and can affect evolutionary mode (Chkhaidze et al., 2019; Noble et al., 2019; West et al., 2019).

Epithelial tissues, which give rise to the majority of human cancers, are organized as layered structures, constrained by basal membranes and the extracellular matrix (ECM) (Bissell et al., 2002; Bissell and Radisky,

[1]Integrated Mathematical Oncology Department, H. Lee Moffitt Cancer Center & Research Institute, 12902 USF Magnolia Drive, Tampa, FL 33612, USA

[2]Cancer Physiology Department, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, SRB 4 Rm 24000H, Tampa, Florida 33612, USA

[3]Lead contact

*Correspondence: jackedwards1@usf.edu (J.E.), david@cancerevo.org (D.B.)

https://doi.org/10.1016/j.isci.2020.101901

2001). Carcinogenesis involves breaking from the growth limitations imposed upon by epithelial layers and gaining access to the new space through the degradation and remodeling of the ECM (Cawston and Wilson, 2006). From an ecological perspective, accessing this new space constitutes niche engineering; thus, we will refer to the phenotypes of cancer cells which possess the ability to remodel the ECM as "engineers" (Lloyd et al., 2016; Myers et al., 2020). ECM degradation is achieved through the secretion of matrix-degrading enzymes, which require a sufficiently high local concentration to achieve their effects. Since the newly accessed space can benefit not only engineers but also non-engineer cells, the space, unlocked by engineers, represents a "common good" resource (Kagel and Roth, 1995). Optimization for use of space can be achieved by the increased consumption of this resource through elevated proliferation rates; thus, we will refer to the phenotype that optimizes space consumption as "proliferators". To understand eco-evolutionary dynamics resulting from selection for optimal phenotypic strategies toward creating (engineering) and consuming (proliferating) space, we consider an abstract on lattice grid, where the domain is divided up into multiple subdomains by degradable ECM barriers.

Using a cellular automaton to implement our ABM, we examined the eco-evolutionary dynamics emerging under selection pressures that act to optimize mutually exclusive engineering and proliferating phenotypic strategies within domains with different spatial organization (Gatenbee et al., 2019a; Margolus and Toffoli, 1987; Poleszczuk and Enderling, 2014). We found that gaining access to new space through non-cell autonomous effects of engineers can dramatically increase ITH and lead to patterns which, in the context of ecology, can be described as ecological succession. Surprisingly, we found that this selection-driven dynamic could lead, under sampling resolutions commonly used in cancer genomics, to subclonal diversification patterns that could be misinterpreted as evidence for neutral evolution.

RESULTS

Model description

We initiate simulations with 25 cells with the basic phenotype seeded in a subdomain, surrounded by ECM barriers. The subdomain itself is located in the center of the large domain, composed of the grid of subdomains. Cell growth is constrained to the subdomain, unless the ECM separating the population from the neighboring subdomain is degraded. ECM barriers can be degraded through action of engineer cells if a sufficiently high concentration of ECM-degrading enzymes is produced – which is a function of both the number of engineers, as well as their efficiency (Figure 1A). When a cell divides, it can acquire a "driver" mutation that permanently increases either expression of ECM-degrading enzymes (for the engineer phenotype) or proliferation rate (for proliferator phenotype). After each division, a check is made to remove any ECM which has met the removal criteria. A cell can die through apoptosis (with a rate of 15% each timestep), freeing up space that can be explored by its neighbors. There is phenotype plasticity in the model, so engineers can become proliferators (at a rate of 0.2%) and vice versa. Although these engineering and proliferative phenotypes are mutually exclusive, the mutational history of each cell enables them to start from the previous rates of engineering and proliferation if they switch back (Figure 1B). Simulations are initialized within an 601x601 domain, separated into smaller sub-domains, ranging from 25x25 to 301x301 (shown in Figure 1C), and run for 2500 timesteps, which allows for simulations to run for sufficiently long time so that the relevant evolutionary dynamics can be examined.

Impact of environmental engineering on phenotypic and mutational diversity

Regardless of the subdomain size, we observed two scenarios characterized by engineering success. In some simulations, the engineering phenotype was driven to extinction, which left the population confined within the initial subdomain (failed engineering). In others, the population was capable of accessing the majority or all of the subdomains of the larger domain (Figure 1C). As a control, we examined the impact of domain size on phenotypic and clonal heterogeneity in the complete absence of engineering phenotype (Figure S1A). In the absence of successful engineering, phenotypic diversity peaked early and then decreased as more effective proliferators were selected. Increase in the subdomain size lead to an increase in the peak phenotypic diversity. In the presence of successful engineering, phenotypic heterogeneity peaked at later time points but at significantly higher levels. Similar to the outcome in the absence of engineering, the phenotypic diversity decreased after reaching its peak, indicating convergent evolution toward highest proliferation rates (Figures 2A and S1A). Clonal heterogeneity increased directly with subdomain size, reflecting an increase in population size (Figures 2B and S1B). After peaking, upon population reaching carrying capacity within spatial constrains, clonal heterogeneity remained high, with the effect more easily observed in larger subdomains. Successful engineering led to a delay in reaching maximal

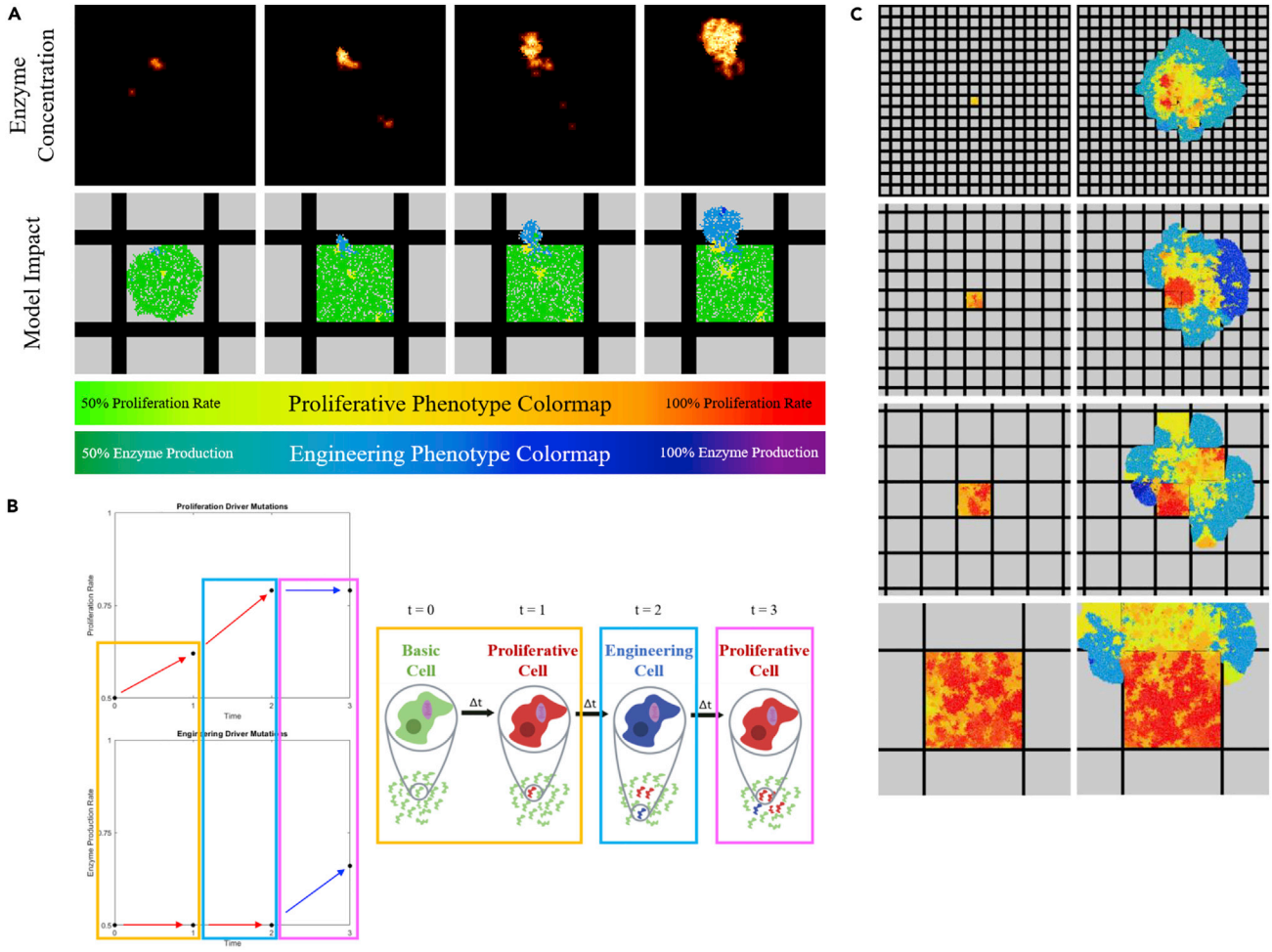


Figure 1. Model description

(A) A depiction engineers degrading an ECM barrier in the model. For the ECM to be degraded, a sufficient concentration (~45% enzyme per lattice unit) of matrix-degrading enzyme is required near the ECM. The top images show the concentration of enzyme and the bottom images show the population of engineers producing them (with phenotypic color map below).

(B) A depiction of how driver mutations interact with phenotypic switching in the model. The top plot shows the proliferation driver mutation space, and the bottom plot shows the engineering driver mutation space. For all basic cells, their first driver mutation is randomly selected, which is the only time a mutation determines the cell's phenotype (orange box). Afterward, driver mutations will increase the cell's phenotype at the time of the mutation. However, phenotypic switches occur at the end of a time frame. The change boxed in blue shows the cell acquiring a proliferative driver (as it was the cell's phenotype at the time of division), yet it stochastically switched phenotypes at the end of the timestep. Similarly, the cell acquires an engineering mutation in t = 3 and again stochastically switches the phenotype at the end of the timestep (boxed in pink). The cell retains a memory of its mutational history, so the fitness does not reset upon switching phenotypes, yet only one phenotypic characteristic can be expressed at a time.

(C) Images taken from simulations of the four subdomain sizes in the model, showing the two types of characterized simulations. All parameters are the same except subdomain size. The left column shows simulations, where engineering failed, and the right column shows simulations where engineering succeeded.

clonal heterogeneity but enabled populations to reach much higher levels of clonal heterogeneity. Notably, engineering limited reduction in clonal ITH, observed shortly after the peak. Therefore, environmental engineering resulting in production of public goods (space) enabled tumor populations, not only to gain access to space and achieve larger population size but also to gain increased levels of phenotypic and clonal heterogeneity.

Successful engineering was associated with noticeable patterns of phenotypic succession. After breaking the initial ECM barrier, engineers exploited the newly accessed space, growing outward the next barriers. As engineers continued to degrade the ECM and grow outward, proliferative cells began to outcompete vestigial engineers near the core of the tumor (Figure 3A, Video S1). Once all of the ECM barriers in the

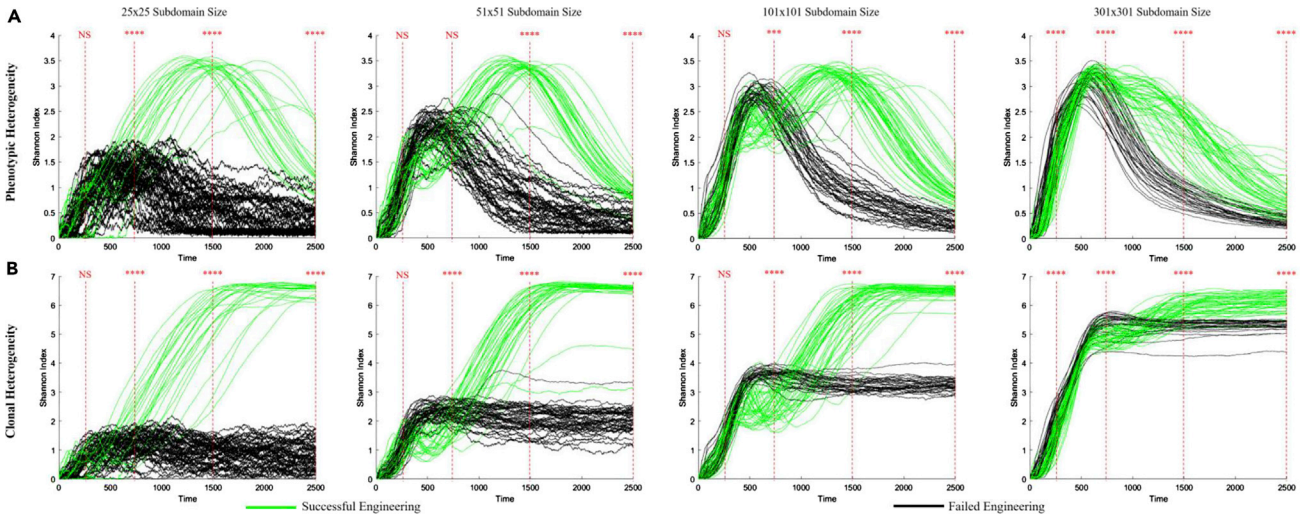


Figure 2. Impact of successful engineering on intra-tumor htheterogeneity

Temporal changes in genetic (A) and phenotypic (B) heterogeneity are captured by the Shannon Index. Green and black lines represent successful and failed engineering, respectively. Each plot depicts the outcomes of 80 simulations. NS, *, **, ***, and **** represent p values of Kolmogorov-Smirnov tests at the indicated time points: NS indicates $p > 0.05$, * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$, and **** indicates $p < 0.0001$.

domain were removed, engineers were succeeded by proliferators, both from the expansion of pre-existing proliferative sub-clones and mutational conversion of engineers to proliferators (Figures 3B and 3C). Notably, this phenotypic succession was not accompanied by a clonal one, as while phenotypic strategy converged toward proliferative phenotype, clonal heterogeneity was maintained at near peak levels. This result implies high levels of convergent evolution, where identical phenotypic solutions are reached by different subclonal lineages.

Selection-driven dynamics can lead to clonal architectures consistent with neutral evolution

Since we recorded mutational history for each of the cell within simulation, these data enabled reconstruction of "true" (within simulations) clonal architecture, in contrast to inferred clonal architecture obtained from analyses of bulk genome sequencing. We visualized clonal architecture using Muller plots (Gatenbee et al., 2019b). In this visualization, subclones with sub-threshold frequency (the majority of subclones in our simulations) are "invisible", instead being grouped with the parental (sub)clone. We chose to visualize changes in clonal architecture using 10% clonal resolution threshold (Figure 4A), corresponding to a common 20x genome sequencing depth (assuming near diploid genome), and 1% clonal resolution threshold, reflecting higher resolution analyses (Figure 4B), and 0.1% threshold (Figure S2) – which provides a more accurate representation of the "true" clonal architecture.

In the absence of successful engineering, in the smallest subdomain size, we observed patterns of clear clonal succession (Figure 4). Increase in the subdomain size resulted in tumors with more complex clonal architecture but still with clear evidence of clonal expansions. Successful engineering dramatically increased clonal heterogeneity, leading to lower rates of expansion of individual sub-clones, indicative of increased clonal interference. Noticeably, increase of the analysis resolution revealed more complex clonal patterns (Figures 4 and S2).

Presence/absence of strong clonal expansions within the Muller plot visualization can be intuitively interpreted as evidence of selection-driven/neutral evolution. However, as such an interpretation might be misleading, we asked whether the apparent differences in patterns of clonal expansions observed at different resolution levels would be interpreted differently when analyzed using quantitative metrics of neutrality. To this end, we examined the relation between the cumulative number of mutations as a function of the inverse frequency of the mutations. This approach has been used to discriminate between neutral (no fitness differences between subclonal lineages) and selection-driven evolution in inferences of evolution modes from bulk sequencing data, where a linear fit between the inverse of the variant allele frequencies (VAFs) of the detected mutations with the cumulative mutation function is interpreted as evidence of evolution neutrality (Williams et al, 2016, 2018). Since we

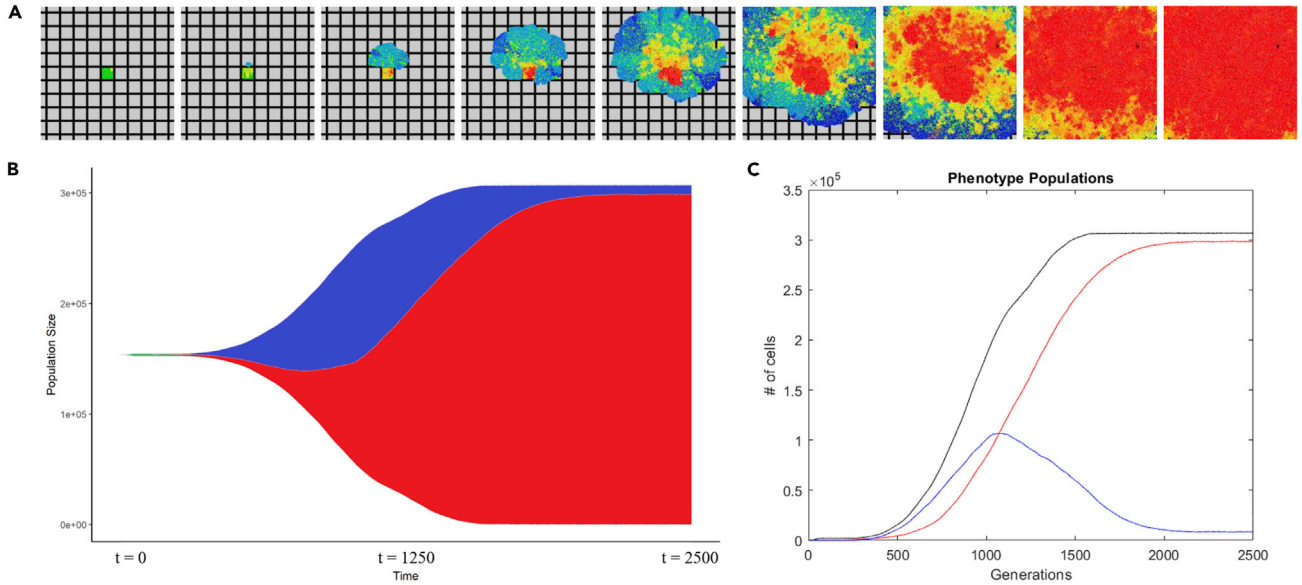


Figure 3. Patterns of ecological succession in the model

(A) Images from the visualization of a simulation outcome showing engineers gaining access to new space and being succeeded by proliferative cells over time.

(B) Frequencies of basic, proliferative, and engineering cells over time.

(C) Populations of proliferative cells, engineering cells, and total tumor over time. Engineers flourish when they break out initially, and over time, proliferative cells can grow into the space created by engineers.

had VAF data recorded from our simulations, we analyzed the goodness of linearity fit of our simulation data at different sub-domain sizes and resolution levels of 1%, 0.1%, and 0.01%. As a reference point, we generated VAF data from simulations that followed neutral dynamics (no changes in proliferation probability), under a scenario of no ECM barriers (lack of subdomains). Strikingly, we found a better linear fit of data from our selection-driven simulations, compared to the neutral control, at both 1% and 0.1% resolution, with the trend reversed at 0.01% (Figure 5A). The impact of resolution level on goodness of fit was not limited to our simulations, as we have also observed marked differences in the goodness of linearity fit within the artificial data set provided with neutralitytestr (Williams et al, 2016, 2018) (Figure S3). Interestingly, comparison of the goodness of linearity fit between simulations with successful and failed engineering revealed significantly higher fit in simulations with successful engineering across all of the resolution levels, with the differences being more pronounced at lower subdomain size (Figure 5B). These results warrant caution in inferring mode of evolution from genomic sequencing data.

DISCUSSION

Our study examined the eco-evolutionary dynamics of tumor cell populations using an ABM where cells evolve under explicit pre-defined rules, specifying the impact of mutations and epigenetic switches on phenotypes that impact Darwinian competition for a limited ecological resource – space. In addition to the commonly used consideration of evolutionary fitness as a cell-intrinsic proliferation probability, our simulations included a second evolutionary strategy, which specified the ability of cells to unlock access to the "public good" of new space via non-cell autonomous action of environmental engineering, mediated by the action of secreted ECM-remodeling enzymes. We found that non-cell autonomous production of common good had a profound impact on both phenotypic and clonal evolution, leading to higher heterogeneity and the emergence of patterns of phenotypic succession, while at the same time producing clonal architecture that might be consistent with a hypothesis of evolution neutrality.

Consistent with published reports of convergent evolution in cancers, where mutations with similar functional consequences are observed in distinct subclonal sub-populations (Ardaševa et al., 2020), we observed phenotypic convergence in our simulations, as different lineages within computational tumor were capable of finding optimal phenotypic solutions (Gerlinger et al., 2012). Although consideration of non-cell autonomous environmental engineering increased phenotypic diversity, it did not qualitatively impact convergence.

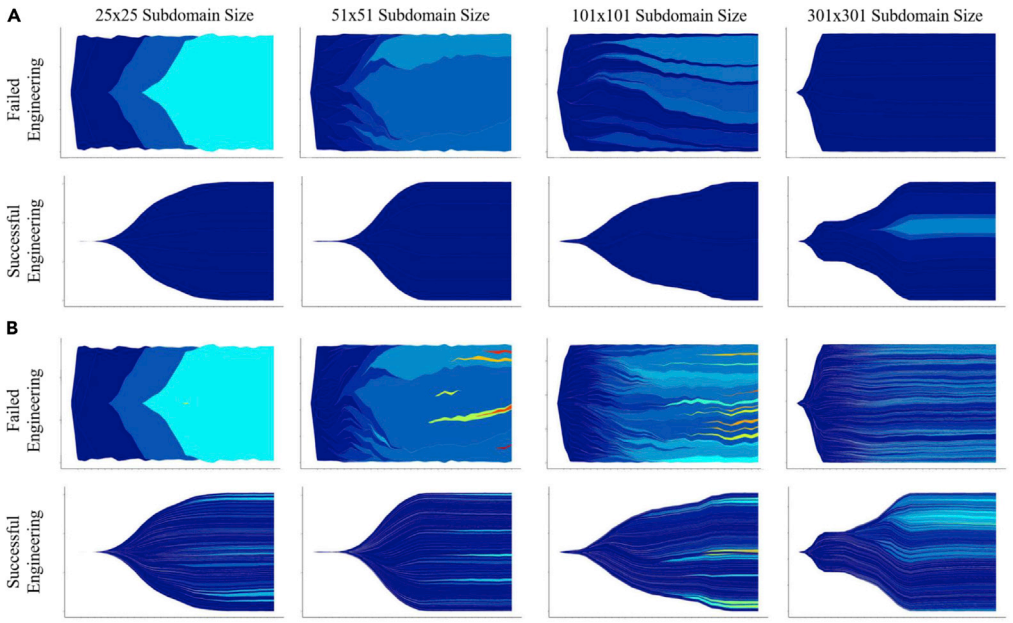


Figure 4. Muller plot visualization of the spatiotemporal clonal dynamics

Two representative simulations for each subdomain size are shown, one where engineering failed and one where engineering succeeded. The x-axis on each plot is time, and the y-axis is population size. Colors indicate time of clone's emergence, with earlier clones presented by dark blue and later clones presented by warmer colors. Only clones with above-threshold frequency are plotted.

(A) Visualization with a frequency threshold of 10%.

(B) Visualization with a frequency threshold of 1%. Visualization with a frequency threshold of 0.1% is shown in Figure S2. Plots at different resolution levels represent the same individual simulations.

The addition of public good-enabled populations of tumor cells led to higher levels of clonal heterogeneity. To a large extent, this effect likely reflected the ability of these tumors to reach a much higher population size, as it was most pronounced within smaller subdomain size. However, public good engineering action also enabled a more effective maintenance of clonal ITH (notice the lack of a dip after reaching the peak in green versus black lines in Figure 2A) – consistent with the report of maintenance of clonal heterogeneity in a mouse model of non-cell autonomous driver of tumor growth (Marusyk et al., 2014).

The inclusion of producers of public good leads to more complex ecological dynamics, involving local and global phenotypic successions. The ability to degrade ECM barriers can, in some cases, lead to strong selection for engineering phenotypes, as they were capable of getting priority access to the newly available space – making the engineering strategy more effective at the growing edge of a tumor. However, as proliferative phenotypes were more successful in utilizing space, they eventually outcompeted engineers, leading to robustly reproducible patterns of phenotypic succession, which parallels ecological succession.

Even though the dynamic behavior in our simulations was designed to be governed by strong selection for competing phenotypic strategies, it produced patterns of clonal architecture that could be interpreted as indicative of neutral evolution patterns within current approaches toward inferences of evolution mode from genomic sequencing data. Most likely, this result reflects a lack of mutational record of history of phenotypic switching. Surprisingly, inferences of neutrality based on goodness of fit between the inverse of VAF and cumulative numbers of mutations showed higher consistency of the outcomes of our selection-driven simulations compared to simulations with explicitly neutral dynamics. Although increased resolution of mutation sampling led to superior fit in neutral simulations, the effect was only observed at resolution levels that exceed those employed in typical mutational analyses studies, with the exception of analyses of single crypts in colorectal cancers. Although our results do not disprove recent reports of neutral evolution in many cancers (Williams et al, 2016, 2018), they do warrant a more cautious interpretation of evolutionary inferences based on mutational analyses.

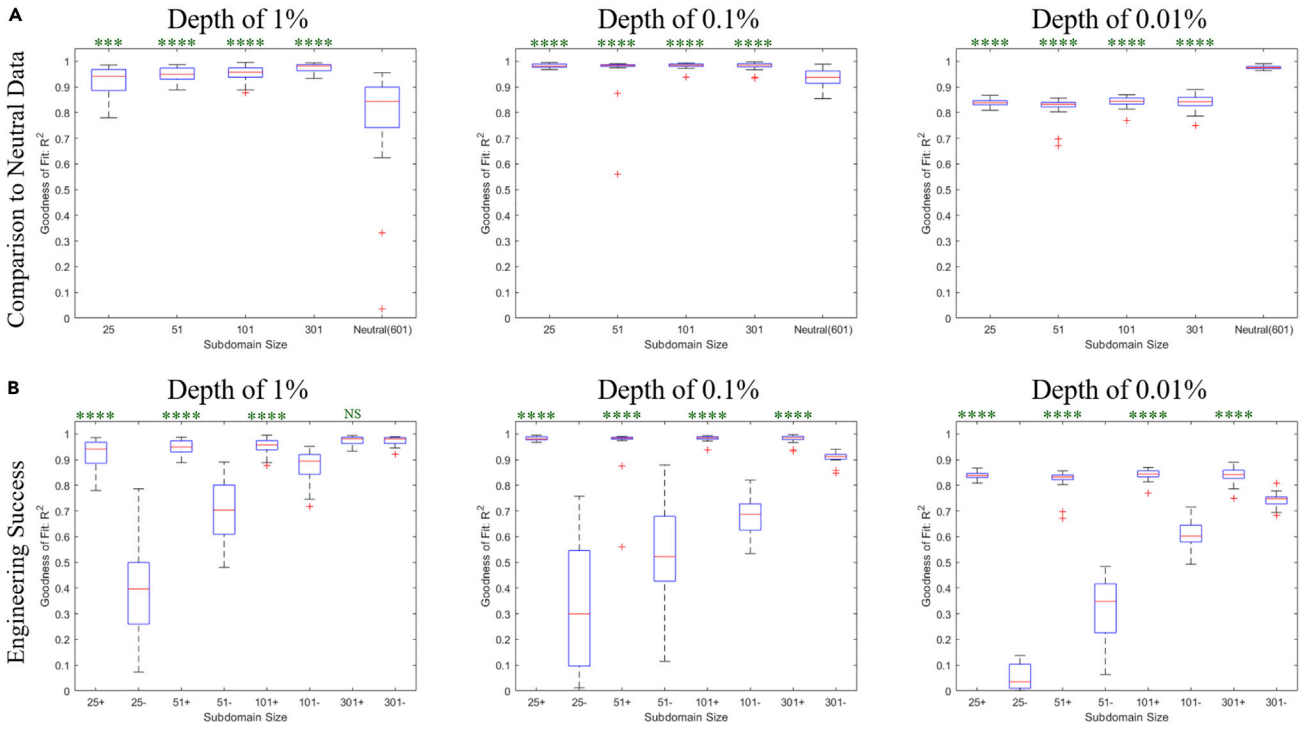


Figure 5. Consistency of clonal architecture with metrics of evolution neutrality

Goodness of fit to a linear regression between the cumulative mutation function and the inverse VAFs ($M(f) \propto 1/f$) is used as a metrics of evolution neutrality.

(A) Goodness of feet for the in-silico growth simulations from Figure 2 with the indicated domain size and simulations with neutral growth dynamics. Statistical comparisons are made between simulations with successful engineering and the neutral simulations.

(B) Impact of engineering success on the goodness of fit. Statistical comparisons are made between the success and failure (indicated by + and – respectively) of each subdomain size. NS, *, **, ***, and **** represent p values of Kolmogorov-Smirnov between two distributions: NS indicates $p > 0.05$, * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$, and **** indicates $p < 0.0001$.

Limitations of the study

Although mathematical and computational modeling enables quantitative analysis of evolutionary dynamics in space in time, modeling deals with abstraction and simplification of biological reality. Moreover, the use of computational modeling depends on the accuracy of model assumptions and parameters, which are specified explicitly. While tumor growth requires gaining access to new space, this access is mediated not only by the action of tumor cells but also by complex networks of microenvironmental interactions involving non-tumor cells, such as fibroblasts, macrophages, etc (Janiszewska et al., 2019; Marusyk et al., 2014). A concrete understanding of the underlying biology at this level, permitting adequate capture in a mathematical model, is still missing. Even if such knowledge were available, quantitatively specifying all of relevant interactions would not be feasible. While primary tumors might be simpler than natural ecosystems, their ecological complexity is certainly higher than what can be captured in our models (Brown, 2016). Further, two-dimensional lattices considered in our work do not faithfully recapitulate three-dimensional tissue structures within primary or experimental tumors. On the other hand, computational modeling enables the examination of outcomes when the key ecological and evolutionary drivers of population dynamics are considered – something which is not accessible not only to inferences from analyses of clinical samples but also to experimental studies. Therefore, while quantitative results of our simulations might not be directly applicable to understanding of tumor biology, qualitative finding of the impact of ecological engineering on tumor heterogeneity and phenotypic succession is likely to be relevant to real tumors. Non-cell autonomous actions, including public goods, appear to be common in tumor ecosystems, including access to vasculature nutrients, immune invasion, etc (Tabassum and Polyak, 2015). Although our simulations only considered non-cell autonomous effects in terms of accessing spatial resources, results from our simulations are likely to be generalizable to other scenarios involving other public goods. We posit that consideration of ecological aspects of cancer evolution, the presence of non-cell autonomous effects, as well as integration of mathematical modeling with experimental studies, and analyses

of clinical samples will be essential to better understand clonal evolution – which is a prerequisite to the development of successful strategies to prevent and eradicate advanced cancers.

Resource availability

Lead contact

Further information and requests for resources and data should be directed to and will be fulfilled by the Lead Contact, Jack Edwards (jackedwards1@usf.edu).

Materials availability

This study did not utilize any physical materials nor does it contain any biological data.

Data and code availability

The model is publicly available on GitHub at CE_ABM: https://github.com/jackedwards1/CE_ABM.

The accession number for the data used to produce the figures reported in this paper is available on Mendeley: https://dx.doi.org/10.17632/5wpcrnkvc7.1.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2020.101901.

ACKNOWLEDGMENTS

We would like to thank the Summer Program for the Advancement of Research Knowledge (SPARK) at the H. Lee Moffitt Cancer Center & Research Institute for sponsoring J.E. and the Integrated Mathematical Oncology Department at Moffitt for hosting this research project. We would also like to thank Dr. Joel Brown and Ryan Schenck for discussions and ideas, and Jongsong Phane for artistic contributions to Figure 1. J.E. and D.B. were partially supported by an NCI IMAG-MSM grant (U01CA202958). D.B. was also funded by U01 CA244101. A.M. and D.B. were partially supported by the State of Florida award 20B06 (30-20450-99-01).

AUTHOR CONTRIBUTIONS

J.E. performed code development, data processing, and statistical analyses. J.E., A.M., and D.B. designed the study and wrote the manuscript. D.B. provided funding and overall supervision for the work.

DECLARATION OF INTERESTS

No competing interests.

Received: April 30, 2020
Revised: October 7, 2020
Accepted: December 1, 2020
Published: January 22, 2021

REFERENCES

Altrock, P.M., Liu, L.L., and Michor, F. (2015). The mathematics of cancer: integrating quantitative models. Nat. Rev. Cancer *15*, 730–745.

Ardaševa, A., Gatenby, R.A., Anderson, A.R.A., Byrne, H.M., Maini, P.K., and Lorenzi, T. (2020). A Mathematical Dissection of the Adaptation of Cell Populations to Fluctuating Oxygen Levels. Bulletin of Mathematical Biology *82* (6), 81, https://doi.org/10.1007/s11538-020-00754-7.

Axelrod, R., Axelrod, D.E., and Pienta, K.J. (2006). Evolution of cooperation among tumor cells. Proc. Natl. Acad. Sci.U S A *103*, 13474–13479.

Basanta, D., and Anderson, A.R.A. (2013). Exploiting ecological principles to better understand cancer progression and treatment. Interf. Focus *3*, 20130020.

Bissell, M.J., and Radisky, D. (2001). Putting tumours in context. Nat. Rev. Cancer *1*, 46–54.

Bissell, M.J., Radisky, D.C., Rizki, A., Weaver, V.M., and Petersen, O.W. (2002). The organizing principle: microenvironmental influences in the normal and malignant breast. Differentiation *70*, 537–546.

Brown, J.S. (2016). Why Darwin would have loved evolutionary game theory. Proc. R. Soc. B Biol. Sci. *283*, 20160847.

Cawston, T.E., and Wilson, A.J. (2006). Understanding the role of tissue degrading

enzymes and their inhibitors in development and disease. Best Pract. Res. Clin. Rheumatol. *20*, 983–1002.

Gatenbee, C., West, J., Baker, A.M., Guljar, N., Jones, L., Graham, T.A., Robertson-Tessi, M., and Anderson, A.R.A. (2019a). Macrophage-mediated immunoediting drives ductal carcinoma evolution: space is the game changer (preprint). Cancer Biol.

Chkhaidze, K., Heide, T., Werner, B., Williams, M.J., Huang, W., Caravagna, G., Graham, T.A., and Sottoriva, A. (2019). Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. PLOS Computational Biology *15* (7), e1007243, https://doi.org/10.1371/journal.pcbi.1007243.

Gatenbee, C.D., Schenck, R.O., Bravo, R.R., and Anderson, A.R.A. (2019b). EvoFreq: visualization of the Evolutionary Frequencies of sequence and model data. BMC Bioinformatics *20*, 710.

Gatenby, R.A., and Gillies, R.J. (2008). A microenvironmental model of carcinogenesis. Nat. Rev. Cancer *8*, 56–61.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. *366*, 883–892.

Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306–313.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57–70.

Janiszewska, M., Tabassum, D.P., Castaño, Z., Cristea, S., Yamamoto, K.N., Kingston, N.L., Murphy, K.C., Shu, S., Harper, N.W., Del Alcazar, C.G., et al. (2019). Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. Nat. Cell Biol. *21*, 879–888.

Kagel, J.H., and Roth, A.E. (1995). The Handbook of Experimental Economics (Princeton University Press).

Korolev, K.S., Xavier, J.B., and Gore, J. (2014). Turning ecology and evolution against cancer. Nat. Rev. Cancer *14*, 371–380.

Lloyd, M.C., Cunningham, J.J., Bui, M.M., Gillies, R.J., Brown, J.S., and Gatenby, R.A. (2016). Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces. Cancer Res. *76*, 3136–3144.

Margolus, N., and Toffoli, T. (1987). Cellular Automata Machines. Complex Systems *27*, 967–993.

Marusyk, A., Tabassum, D.P., Altrock, P.M., Almendro, V., Michor, F., and Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature *514*, 54–58.

Myers, K.V., Pienta, K.J., and Amend, S.R. (2020). Cancer cells and M2 macrophages: cooperative invasive ecosystem engineers. Cancer Control *27*, 107327482091105.

Noble, R., Burri, D., Kather, J.N., and Beerenwinkel, N. (2019). Spatial Structure Governs the Mode of Tumour Evolution (Preprint). Cancer Biol., In preparation.

Nowell, P. (1976). The clonal evolution of tumor cell populations. Science *194*, 23–28.

Poleszczuk, J., and Enderling, H. (2014). A high-performance cellular Automaton model of tumor growth with dynamically growing domains. Appl. Math. *05*, 144–152.

Sottoriva, A., Kang, H., Ma, Z., Graham, T.A., Salomon, M.P., Zhao, J., Marjoram, P., Siegmund, K., Press, M.F., Shibata, D., and Curtis, C. (2015). A Big Bang model of human colorectal tumor growth. Nat. Genet. *47*, 209–216.

Swanton, C. (2012). Intratumor heterogeneity: evolution through space and time. Cancer Res. *72*, 4875–4882.

Tabassum, D.P., and Polyak, K. (2015). Tumorigenesis: it takes a village. Nat. Rev. Cancer *15*, 473–483.

West, J., Schenck, R.O., Gatenbee, C., Robertson-Tessi, M., and Anderson, A.R.A. (2019). Tissue structure accelerates evolution: premalignant sweeps precede neutral expansion (preprint). Cancer Biol., In preparation.

Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A., and Sottoriva, A. (2016). Identification of neutral tumor evolution across cancer types. Nat. Genet. *48*, 238–244.

Williams, M.J., Werner, B., Heide, T., Curtis, C., Barnes, C.P., Sottoriva, A., and Graham, T.A. (2018). Quantification of subclonal selection in cancer from bulk sequencing data. Nat. Genet. *50*, 895–903.

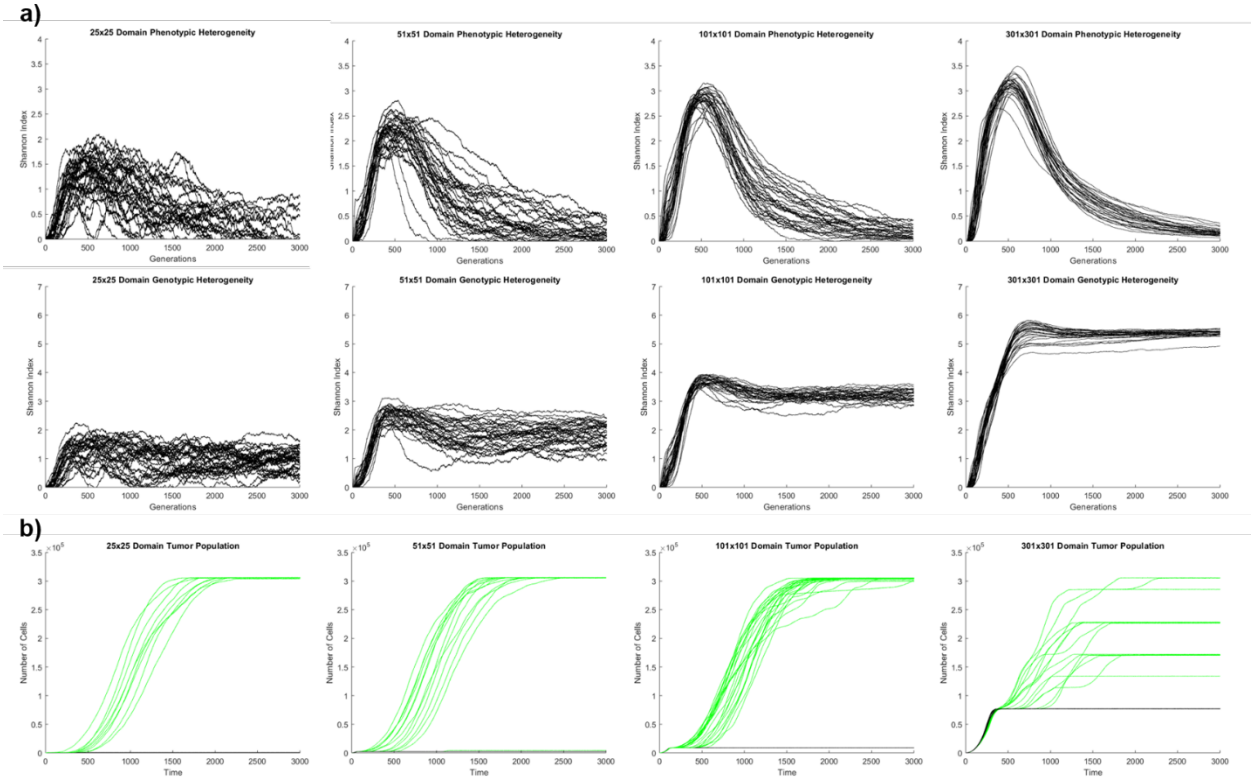
iScience, Volume *24*

Supplemental Information

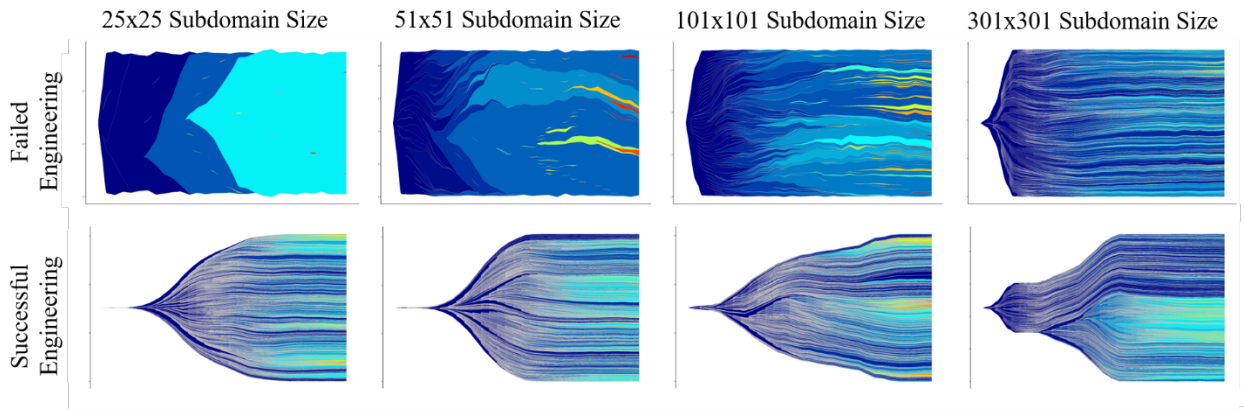
Selection-driven tumor evolution with public goods leads to patterns of clonal expansion consistent with neutral growth

Jack Edwards, Andriy Marusyk, and David Basanta

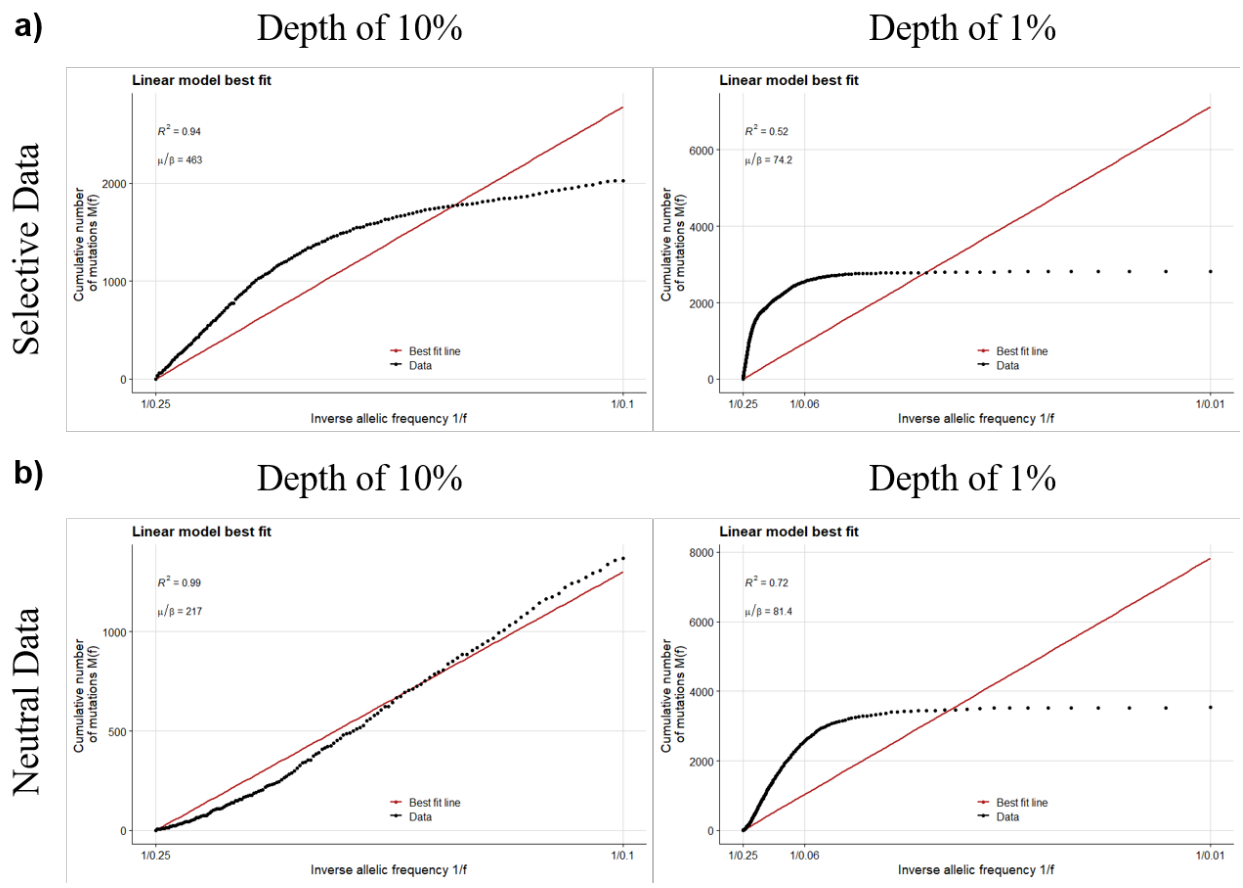
Supplemental Figures



Supplemental Figure 1: **Control data and population data.** Relates to Figure 2. **a)** Phenotypic heterogeneity (top) and clonal heterogeneity (bottom) for simulations over time when the engineering phenotype was disabled. There are 30 simulations per plot, each run for 3000 timesteps. **b)** Total tumor population over time in simulations with the same parameters as Figure 2. Black lines represent failed engineering, green lines represent successful engineering. The failed engineering tumor carrying capacity is dependent on what the subdomain size is.



Supplemental Figure 2: **Muller plot visualization of the spatiotemporal clonal dynamics at 0.1% resolution.** Relates to Figure 4. Two representative simulations for each subdomain size are shown (same simulations as in Figure 4), one where engineering failed, and one where engineering succeeded. The x-axis on each plot is time, and the y-axis is population size. Colors indicate time of clone's emergence, with earlier clones presented by dark blue, and later clones presented by warmer colors. Only clones with above-threshold frequency are plotted.



Supplemental Figure 3: Examining the effect of frequency on the inverse allelic frequency metric for neutrality. Relates to Figure 5. Artificial data that comes with the R package neutralitytestr (Williams et al., 2018, 2016). The plots show the cumulative number of mutations as a function of the inverse allelic frequency, and the goodness of fit to a linear distribution, as a metrics for a neutral mode of evolution. **a)** The selective data plotted from frequencies of 25% to 10%. **b)** The neutral data plotted from frequencies of 25% to 1%.

Transparent Methods

The model uses a 2D on-lattice agent-based model, where each grid point is either unoccupied, blocked by ECM, or is occupied by a living cell. There are three cell phenotypes: basic, proliferative, and engineering. Cells with the basic phenotype (basic) are capable of proliferating while ignoring homeostatic signaling constraints, as long as space is available, but are incapable of accessing new space through degradation of matrix. Cells with the engineering phenotype (engineers) proliferate at the same rates as basic cells but can provide access to new space by degrading ECM barriers through secretion of matrix degrading enzymes. No single engineer can produce enough of the enzyme on its own to degrade the ECM; a substantial population of engineers local to the ECM must secrete enzymes at the same time to access new space. Cells with a proliferative phenotype (proliferators) can divide at higher rates than basic and engineer cells, thus they are more effective at claiming available space, but they cannot gain access to new space on their own.

The CA uses a Moore neighborhood to define the cells in direct contact with each other (Margolus, n.d.). Cells are iterated over stochastically as to avoid artifacts or giving preference to cells in specific spatial locations. Each cell determines if it has the ability (fitness) and resources (space) to divide during that timestep. Cells have a potential proliferation rate which defines its ability to proliferate if space is available. For instance, a cell with a potential proliferation rate of 75% has a 75% probability to divide at each time step, provided there is space. Potential proliferation starts at 50% within the basic phenotype and can be increased by mutations that optimize proliferative phenotype – up to 100% (dividing at each time step given space availability). At any given timestep, a cell has a 15% probability of dying, which leaves an empty space in the next timestep, enabling cell turnover within the core of the tumor.

At each cell division, one of the following two types of *driver* mutations might occur at a total probability of 0.07%. First, a proliferation-driving mutation (0.07% probability), can increase cell proliferation with a randomly chosen increment of 0.01% between 10 and 20%. Second, engineering-driving mutation, occurring at 0.07% probability, bestows a cell with an engineering phenotype, enabling the mutated cell to produce a matrix-degrading enzyme at 50% probability per time step. The enzyme-production ability of engineering phenotypes can be further increased by additional engineering mutations, through increments of 0.01% in the 10 - 20% range. In addition to driver mutations, *passenger* mutations, occurring with 0.5% per time step probability, could be either entirely neutral (no impact on cell fitness), or slightly deleterious by decreasing probability of proliferation in the range 0 to 0.5%, with 0.01% increment. If no mutations occur, a cell retains its current phenotype, which is inherited upon cell division.

When a basic cell first acquires a driver mutation, it will randomly become either a proliferator cell or an engineer. Further driver mutations will be chosen dependent on the cell's current phenotype; proliferative cells will always acquire proliferative drivers, and engineering cells will always acquire engineering drivers. Based on the principle of evolutionary tradeoff, we consider engineer (ECM-degrading) or proliferative phenotypes to be mutually exclusive (Townsend et al., 2009). However, a cell can switch its phenotype from engineer to proliferator and vice versa at a probability of 0.2% at the end of each timestep. This type of stochastic phenotypic switching among cancer cell has been described in the literature (Gupta et al., 2011; Zhou et al., 2014). When a phenotypic switch occurs, we assume that the cell loses the expression of improved abilities related to the pre-switch phenotype. If this cell or its clonal progeny undergo a reverse switch, it will recover these abilities at the levels that preceded the first switch. As an example, an engineer cell that undergoes phenotypic switch toward proliferator will stop degrading ECM and will proliferate at a rate that would correspond to be basal one if it never had a proliferative phenotype before. If it subsequently undergoes a mutation that transforms it into a cell with an engineer phenotype, the ability to degrade ECM will be restored to the level it had previously as an engineer.

In order to interrogate the accumulation of *genetic* ITH, we record mutational history for each cell. Each driver mutation or any successive four passenger mutations were used in the analyses as a mutational branching point, marking a new sub-clone. *Phenotypic* strategies and mutational history were tracked for all the lineages over time. To define phenotypic ITH, we binned cells based on their proliferation and enzyme-producing phenotypes, with increments of 1%. Thus, for example, two proliferator cells representing different genetic subclonal branches, with proliferation probabilities of 59.1% and 59.7% would be counted as belonging to the same *phenotypic* subpopulation. Based on this definition of phenotypic

subpopulations, there were 103 possible distinct phenotypic groups: 51 proliferative, 51 engineering, and 1 basic. To visualize outcomes of the simulations, we map a color to a cell's phenotype and fitness as described by the aforementioned rules. This map is visible in **Figure 1a**.

A concentration of ~45% enzyme per unit lattice is required local to a point of ECM to remove it in the model. The enzyme produced by engineers is considered to be a diffusible molecule that decays over time. In order for an ECM barrier to be degraded successfully, there needs to be at least 5 engineers producing enzyme immediately next to the barrier, and the minimum number of engineers increases with distance away from the ECM. The maximum diffusion range on the enzyme is a radius of 3 cells out from the grid point where the engineer sits, and the effectiveness follows a simple linear decay in effectiveness (as opposed to an inverse square law). Successful destruction of ECM unlocks the space within the empty subdomain. The newly accessed space resource can be exploited by all cell types.

The parameters chosen for these simulations were selected with the goal that they would lead to realistic looking simulations. A wide variety of parameters were tested, both during and after model development and yielded similar outcomes, convincing us that our results are not just an artifact of the chose parameters. We varied mutation rate [0.005% - 1%], impact of driver mutations [2.5% - 25%], death rate [5% - 50%], and enzyme strength [~20% - ~80%]. The parameters used in the simulations we show are optimal for ensuring all of the relevant eco-evolutionary dynamics play out within a reasonable timeframe, as well as consistently getting a roughly even split of engineering fails and successes in order to make comparisons between the two simulations. The specific parameters used in the reported simulations are: a total simulation length of 2500 timesteps, a mutation bias of 0 (engineering and proliferation mutations are equally likely), a passenger mutation rate of 0.5%, a driver mutation rate of 0.07%, a switching frequency of 0.2%, a driver impact range of 10.5% to 20%, a death rate of 15%, an enzyme radius of 2, and an integer enzyme threshold for degradation of 10 (a strength of 10 enzymes must be present local to a piece of ECM in order for it to be removed).

When computing heterogeneity for both the phenotypic groups and the clonal groups, the Shannon index was used (Shannon, n.d.). The model takes the proportion of each of the phenotypic populations, p_i, and calculates the Shannon Index as follows:

$$H = -\sum_{i=1}^{R} p_i * \ln(p_i)$$

H was calculated for the phenotypic groups and the clones to measure the differences in clonal heterogeneity and the phenotypic diversity of the tumor. A Kolmogorov-Smirnov test was performed to examine the quantitative differences in the heterogeneity distributions between simulations characterized by successful and failed engineering. During the simulation, a list of all clones and their ancestry is recorded and output. From this master list, muller plots can be constructed and most non-spatial neutrality statistics can be computed. In the data processing, we used the master list for each simulation to create a long-format output of the clonal architecture, which is readable by EvoFreq, an R package designed for visualizing clonal dynamics (Gatenbee et al., 2019). We also used the master list to calculate the variant allelic frequencies (VAFs). For our neutrality metric, we compared the cumulative mutations as function of the inverse frequency to a linear model, which is theoretically a neutral mode of growth (Williams et al., 2018, 2016). It is shown that the cumulative amount of distributions, *M(f)*, is described by the following equation in a neutral model of evolution:

$$M(f) = \frac{\mu}{\beta}\left(\frac{1}{f} - \frac{1}{f_{max}}\right)$$

where *f* is the variant allelic frequency of a given mutation, f_{max} is the maximum frequency of a mutation, μ is the mutation rate, and β is the cell division rate in which both lineages survive (Williams et al., 2016). The parameters and derivation of this model are detailed explicitly by Williams et al. in 2016. We examined the goodness of fit of our data to a linear (neutral) model, and also compared our data to both an explicitly neutral simulation in our model as well as the artificial data provided in an R package, neutralitytestr, that is designed to test VAF data using the linear fit method described (Williams et al., 2018, 2016). Again, a

Kolmogorov-Smirnov test was performed to quantify the differences between the distributions of our model data and explicitly neutral simulations.

Supplemental References

Gatenbee, C.D., Schenck, R.O., Bravo, R.R., Anderson, A.R.A., 2019. EvoFreq: visualization of the Evolutionary Frequencies of sequence and model data. BMC Bioinformatics 20, 710.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., Lander, E.S., 2011. Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. Cell 146, 633–644.

Margolus, N., n.d. Cellular Automata Machines 27.

Shannon, C.E., n.d. A Mathematical Theory of Communication 55.

Townsend, C.R., Begon, M., Harper, J.L., 2009. Essentials of Ecology. [electronic resource]., 3rd ed. ed. John Wiley & Sons, Ltd.

Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A., Sottoriva, A., 2016. Identification of neutral tumor evolution across cancer types. Nat. Genet. 48, 238–244.

Williams, M.J., Werner, B., Heide, T., Curtis, C., Barnes, C.P., Sottoriva, A., Graham, T.A., 2018. Quantification of subclonal selection in cancer from bulk sequencing data. Nat. Genet. 50, 895–903.

Zhou, J.X., Pisco, A.O., Qian, H., Huang, S., 2014. Nonequilibrium Population Dynamics of Phenotype Conversion of Cancer Cells. PLoS ONE 9, e110714.