Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Early Colorectal Cancer Growth Is Often Slow and Prolonged

Darryl Shibata

dshibata@usc.edu

The University of Southern California https://orcid.org/0000-0002-4567-1639

Marc Ryser

University of Geneva

Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11545400/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
There is **NO** Competing Interest.

The need for consent was waived by the approving ethics committee.

Early Colorectal Cancer Growth Is Often Slow and Prolonged

Marc D. Ryser[1,2] and Darryl Shibata[3*]

[1] Translational Research Center in Oncohaematology, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2]Department of Population Health Sciences, Duke University Medical Center, Durham, NC, USA

[3] Department of Pathology, University of Southern California Keck School of Medicine, Los Angeles, CA, USA

Correspondence to Darryl Shibata (dshibata@usc.edu)

Abstract

It is uncertain how human colorectal cancers (CRCs) grow before they become clinically detectable. Here, we used approximate Bayesian computation to identify, among millions of simulated growth trajectories, those consistent with lineage-tracing fluctuating CpG (fCpG) methylation patterns in individual tumor glands. Early growth trajectories were constrained in all 12 tumors, indicating that cell loss commonly slows expansion while tumors are still small. Tumor mitotic ages depended largely on the duration of constrained early growth. Maximum-likelihood phylogenies reconstructed from independent single-cell DNA methylation data also supported slow early CRC growth. Inferred mitotic ages of TCGA CRCs (200–800 divisions) were not associated with MSI status, driver mutations or stage. Instead, older CRCs had more detectable subclonal mutations, greater immune infiltration, and stronger markers of immune exhaustion, consistent with a legacy of prolonged early immunoediting. Thus, CRCs can differ substantially in mitotic age, with these differences arising largely from early constraints that produce prolonged, slow growth while tumors remain small and clinically undetectable.

Introduction

It is uncertain how rapidly a tumor grows from a single progenitor cell to a clinically detectable mass or whether similar-sized tumors differ substantially in mitotic age. Human tumorigenesis is often inferred from genomic alterations because direct observations are impractical. Subclonal mutation frequencies (1,2) and the paucity of subclonal driver mutations (3) indicate that most CRCs grow as single, near-neutral expansions (4). It appears that the driver mutations and epigenetic configuration of the final progenitor cell are sufficient to support subsequent tumor expansion (5,6). However, neutral growth is compatible with many growth patterns (1,7,8), including prolonged, slow early growth (Fig. 1A). Reconstructing growth is useful because slower growing tumors can reveal mechanisms that naturally constrain human tumorigenesis. Tumor mitotic age, defined as the number of divisions between the start of growth and tumor removal, likely reflects both positive and negative selective forces. Mutation data can identify positive selection, but mechanisms that constrain tumor growth are difficult to detect in human tumors (9-11).

The tumor microenvironment (TME) can constrain growth (12) and co-evolve during tumorigenesis (13). Clinically detectable tumor growth slows with size and is typically Gompertzian (14), likely reflecting the increasing inability of the TME to support progressively larger tumors (15,16). Less is known about constraints during subclinical tumor growth, especially when tumors are very small ($<10^6$ cells). Smaller tumors may impose lower metabolic demands but may still be constrained by immune surveillance (17), particularly because many

targeted neoantigens are clonal (18) and are therefore present at the start of growth. A legacy of early immunoediting is potentially detectable as immune cell exhaustion from chronic immune stimulation in the TME of older tumors. Immune cell exhaustion rather than immune escape could relax this early constraint, allowing continued neutral CRC growth without subclonal selection (19).

Tumor growth can be inferred from mutations (20), but the paucity of subclonal mutations limits reconstruction to relatively sparse lineage trees. Resolving more detailed growth dynamics, especially during the earliest tumor expansion, therefore requires molecular clocks that record changes at nearly every cell division. Here, fCpG methylation (21) is used to infer CRC growth dynamics. This clock, comprising 1,164 diploid fCpG sites, records an average of 2.3 methylation changes per cell per division, with a 90.3% probability that each division generates at least one informative change. We inferred growth by comparing fCpG tumor gland data to millions of simulated growth trajectories. Maximum-likelihood phylogenies reconstructed from orthogonal single-cell methylation data (22) provide independent support for the inferred growth trajectories. We show that prolonged, slow early growth is common in CRCs.

Results

fCpG methylation is more correlated within than between tumor sides

We analyzed multiregional fCpG DNA methylation patterns in 12 tumors (3 adenomas and 9 CRCs) using EPIC arrays. In each tumor, we analyzed gland-level fCpG methylation in spatially separated glands from opposing sides of the tumor mass (4 glands from side A and 4 from side B) (Fig 1B). Comparisons within a tumor side reflect more recent growth, whereas comparisons between sides capture divergence from the original progenitor cell and therefore earlier growth (23).

The fCpG clock comprised 1,164 fCpG sites (see Methods for details). fCpG sites undergo rapid and approximately balanced methylation and demethylation (21), resulting in average methylation near 50% and enabling high-resolution tracking of cell divisions. Consistent with this mechanism, average gland fCpG methylation was about 50% (Fig 1C). A key feature of the fCpG clock is that its methylation is synchronized to the 0, 50, and 100% pattern of its progenitor cell at the start of growth, resulting in “W-shaped” distributions in younger clonal expansions. fCpG methylation rapidly oscillates and becomes desynchronized and more unimodal in older clonal expansions. Tumor glands generally had high-variance, W-shaped fCpG distributions indicating relatively younger clonal populations; older, less synchronized glands had more unimodal, lower-variance distributions (Fig 1D). Gland variances were variable within and between tumors (Fig 1E), indicating different degrees of gland synchronization.

fCpG methylation correlated poorly between glands from different tumors (Fig 1F), indicating that each tumor’s progenitor cell starts with its own unique pattern. Consistent with fCpG methylation being synchronized by the progenitor cell pattern at the start of growth, fCpG patterns were more correlated between glands within the same tumor side than between glands from opposite sides (Fig 1G). These results support fCpG methylation as a lineage tracer in which initial synchronization records shared ancestry and subsequent desynchronization records lineage divergence during tumor expansion.

ABC of tumorigenesis

We used approximate Bayesian computation (ABC) rejection sampling to fit a flexible family of stochastic tumor evolution models to the fCpG methylation data and to infer the ancestries recorded by fCpG methylation in each tumor. We first generated 9.4 million growth trajectories by sampling from non-informative prior distributions over model parameters. We

then defined a composite distance metric to compare the simulated trajectories with the experimental data, which allowed us to identify parameter combinations that best fit the data.

The guiding principle was neutral evolution (4), and cell loss was applied uniformly across all tumor cells. Cell loss during growth reflects growth constraints that can produce tumors with different growth trajectories and mitotic ages. To model many different possible growth trajectories (Fig 1A), tumorigenesis was divided into five phases: phase 1 consisted of a pure birth process to form the first tumor gland; phases 2 to 4 were modeled as birth-death processes with variable birth-death ratios; and phase 5 represented the clinical plateau in which cell birth was balanced by cell loss, resulting in a stable population size but continued mitotic aging, reflecting the slow growth of detectable cancers (14). To match the simulated growth trajectories with the data, fCpG methylation was embedded within each simulated cell, with random errors possible at every division. We assumed a molecular clock hypothesis in which fCpG error rates were identical across tumors. Using ABC, we estimated an fCpG error rate of approximately 10^{-3} flips per fCpG site per division and an effective gland stem cell number of approximately 80 cells, values that maximized agreement between simulated and observed methylation distributions (Fig. S1). This inferred effective gland stem cell number is consistent with experimental lineage tracing showing that stem cells comprise 5–10% of cells in mouse intestinal adenomas (24).

In 11 of 12 tumors, DNA sequencing data were available (23) and suggested that subclonal mutations reached clonal frequencies within individual glands (4,23,25). Incorporating both subclonal mutations and fCpG methylation into the simulations increased the stringency of ABC filtering, excluding rapidly growing trajectories because fewer total divisions reduce the likelihood of stochastic fixation of subclonal mutations (Fig 2B).

Mitotic ages inferred from ABC runs consistent with the experimental gland fCpG data and the requirement that tumor-subclonal mutations reach near-clonal frequencies within individual glands varied across tumors, with a mean of 389 divisions (range, 335–532) (Fig. 2C). Acceptance rates ranged from 10^{-6} to 10^{-4} (Table 1). ABC favored growth trajectories in which early cell loss preserved most lineages, whereas lineage extinction increased later in growth (Fig. 2D). ABC also favored longer phase 5 stationary intervals, averaging 106 divisions (Fig. 2E), and gland splitting later in growth (Fig. 2F).

Prolonged, slow tumor growth before reaching detectable size

Accepted runs did not show canonical Gompertzian growth, that is rapid early expansion with a continuously decaying growth rate, but instead showed slow early growth (Fig. 2G). Tumors had different mitotic ages, primarily driven by variations in cell loss during the expansion phases rather than by differences in the terminal divisions at the end of growth (Fig. 2C). These findings suggest that mitotic age differences arise from growth constraints that begin while tumors are still small.

To further test whether cell loss begins early, growth curves constrained to start with death-free, exponential expansion for the first 7 gland doublings were generated and compared with the experimental data (Fig. S2). Acceptance frequencies under this exponential constraint were consistently lower than those from simulations without this constraint, indicating that early expansion without cell loss is a poor fit to the data. Across tumors, acceptance frequencies were typically reduced by several-fold and in some cases by more than two orders of magnitude (Fig. S2). Hence, the data favor prolonged, slow growth while tumors are still small, rather than rapid early expansion.

When the phase 5 stationary interval was removed and the remaining expansion trajectories were normalized to the time at which tumors reached detectable size, the inferred trajectories tended to be N-shaped, with slow initial growth followed by further slowing and then acceleration (Fig. 3A, B). N-shaped trajectories were more frequently observed with five than with three growth phases, indicating that this pattern persisted when model flexibility was

increased (Fig. S3). Reaching the 18th gland split generally required far more than the approximately 32 divisions expected under death-free exponential growth (Fig. 3C, left). Moreover, intermediate gland splits generally occurred later than expected if successive gland doublings were uniformly distributed across the expansion, followed by relative acceleration near the end of growth (Fig. 3C, right). Overall, the number of inferred divisions required to reach detectable size differed among tumors, but most tumors showed delayed, heterogeneous expansion, producing prolonged intervals of slow growth before becoming clinically detectable. Thus, tumor growth appears constrained at both ends of its trajectory—while tumors are small and, as in Gompertzian growth, after they become large—with relative acceleration between these phases.

Phylogenetic reconstruction also indicates slow initial tumor growth

Phylogenetic analysis of DNA methylation provides an independent approach to reconstruct tumor growth. Using a standard maximum-likelihood method (IQ-TREE 2 (26)), phylogenies from ABC-accepted simulations of slow initial growth and ABC-rejected simulations with fast initial growth (Fig 4A) were compared with experimental gland fCpG phylogenies (Fig 4B). Mean root-to-tip length of the experimental gland phylogenies correlated with mean mitotic age inferred from ABC-accepted simulations across tumors (Fig. S4), consistent with branch length recording accumulated mitotic divisions. Tip-to-total tree branch-length ratios differ between simulated early rapid exponential growth (star-shaped trees with long terminal branches) and simulated slow initial growth with shorter terminal branches. Experimental tip-to-total branch-length ratios more closely matched the ABC-accepted slow-initial-growth simulations than the ABC-rejected fast-initial-growth simulations (Fig 4C). Hence, a standard phylogenetic method independently supports the slow initial growth inferred by ABC rejection.

Single-cell fCpG methylation infers slow early CRC growth

Next, growth trajectories were inferred from orthogonal single-cell methylation data from seven CRCs with 18 to 228 tumor cells (22). To mitigate the sparse coverage of single-cell whole-genome bisulfite sequencing, we analyzed fCpGs on the X chromosome (see Methods). Maximum-likelihood phylogenies reconstructed from simulated single-cell fCpG methylation data showed that rapid, death-free early exponential expansion produced star-shaped trees with long terminal branches, whereas slower initial growth produced more hierarchical trees with shorter terminal branches (Fig. 4D). Phylogenies reconstructed from the experimental single-cell fCpG data (Fig. 4E) resembled the hierarchical subclonal structures observed in published CRC mutation trees (6). More than 90% of cells clustered with cells from their sampled region, supporting the ability of fCpG methylation to recover tumor lineage structure (Table S3).

Two phylogenetic metrics generally favored slow-initial-growth simulations. Older CRC data trees had relatively long internal branches, and their tip-to-total branch-length ratios (Fig. 4F) more closely matched slow-initial-growth simulations. Cells within the same region were more closely related in older CRC data trees, and their within-region pairwise patristic distances (Fig 4G) also more closely matched slow-initial-growth simulations. The exception was the youngest tumor, CRC12, which was more consistent with early rapid exponential expansion. Thus, an independent single-cell methylation dataset supports slow initial growth as the predominant pattern in older CRCs.

The tumor microenvironment differs between younger and older CRCs

The inferred trajectories suggest that CRC growth is constrained while tumors are small, accelerates as these early constraints diminish, and slows again after tumors become large. To identify factors associated with these growth-rate differences, inferred TCGA CRC mitotic ages (200–800 divisions) were correlated with precalculated annotated metadata (27). Mitotic age did not increase across stages, indicating that more advanced tumors are not necessarily older (Fig

5A). Mitotic ages showed no strong associations with MSI status or driver mutations, although weak correlations with TP53 mutations were observed, suggesting that intrinsic genetic features do not strongly explain variation in growth rates.

Mitotically older CRCs contained significantly more detectable subclonal mutations (1) (Fig. 5B). Only mutations acquired sufficiently early during growth reach detectable frequencies in bulk samples, and their numbers reflect either increased mutation rates or slower tumor expansion (1). Hence, slower initial growth can help explain the greater number of subclonal mutations detectable in bulk samples from older MSI and MSS CRCs.

Extrinsic TME features were strongly associated with mitotic age. Older CRCs had higher immune and stromal cell content and higher lymphocyte and macrophage scores (Fig. 5C). These trends were observed in both MSI and MSS CRCs. Mitotic ages before and after correction for immune cell infiltration were correlated (R = 0.58; Fig. S5), indicating that the inferred ages were not driven solely by leukocyte content. In contrast, younger CRCs showed higher proliferation and greater numbers of altered genomic segments (Fig. 5D).

Markers of chronic immune stimulation and exhaustion (28) were higher in older CRCs. These tumors had stronger TGF-β and IFN-γ responses (Fig. 5E) and higher TCR and BCR Shannon diversity (Fig. 5F). Older CRCs also exhibited higher expression of genes associated with immune exhaustion (Fig. 5G), with stronger associations in MSS CRCs (Table S2). Although causality cannot be established from these cross-sectional data, correlations between inferred TCGA CRC mitotic ages and the extensive metadata suggest that variation in growth rates is more closely associated with extrinsic TME factors than with intrinsic genetic differences. Older CRCs exhibited greater immune infiltration and immune exhaustion, which may represent a legacy of the early growth constraints inferred from fCpG modeling.

Discussion

Tumor evolution can be inferred from somatic alterations, and the analysis of CRC subclonal mutation frequencies is broadly consistent with single near-neutral “Big Bang” expansions (4). Big Bang tumorigenesis is a flexible model that encompasses more than perfect exponential expansion, and many growth trajectories are compatible with neutral competition between subclones (1,7,8). The faster-ticking fCpG methylation clock provides substantially greater temporal resolution than mutation-based analyses, enabling reconstruction of trajectories while tumors are still very small ($<10^6$ cells). Across two independent datasets analyzed using complementary approaches, prolonged, slow early growth was common, with mitotic age differences arising primarily from variation in the subclinical growth phase rather than from differences in the terminal stationary phase. Slow early growth also provides a plausible mechanism for the greater number of detectable subclonal mutations in older near-neutral Big Bang CRCs (1). Therefore, although the clonal driver mutations and epigenetic configuration of the progenitor are sufficient for tumorigenesis, tumor expansion is often slow and prolonged even when tumors are still very small.

Characterizing growth trajectories is important because growth-rate changes can help identify mechanisms that naturally constrain tumorigenesis. The TME of a clinically detectable tumor can be analyzed directly, although its composition is complex and often difficult to interpret (29,30). The Gompertzian slowing of growth as tumors enlarge ultimately produces a terminal stationary phase (14), likely reflecting increasing metabolic and vascular constraints (15,16). Our inferred trajectories reveal a complementary pattern at the other end of tumorigenesis: net expansion was also strongly constrained while tumors were small, before accelerating as these early constraints diminished.

The nature of these early constraints is more difficult to determine because the TME during early tumor growth cannot be sampled directly. Although our growth curves do not reconstruct early TME composition, they infer the timing and degree of the functional constraints imposed during early growth. Many mechanisms can constrain early tumor growth, and

immunoediting provides a parsimonious explanation because immune surveillance can change during tumor development. Early growth constraints may reflect increasing immune pressure on expanding tumors (31,32), consistent with observations that most targeted neoantigens are clonal (18) and that immune evasion is often acquired very early during CRC evolution (19). Chronic immune stimulation can lead to immune cell exhaustion, providing a plausible mechanism for the inferred heterogeneous growth trajectories. Thus, the greater immune cell infiltration and exhaustion in the TMEs of older CRCs may reflect the legacy of much earlier immunoediting (33,34). Growth acceleration from immune exhaustion rather than selection for immune escape is also more compatible with the near-neutral growth dynamics of most CRCs (19). Greater immune cell infiltration is also present in mitotically older breast and endometrial cancers (35,36).

Inferred neoantigen levels were not markedly higher in older CRCs (Fig S6), which may reflect limitations of current prediction algorithms (37), preferential immunoediting of lineages carrying strongly immunogenic neoantigens, or recognition of even weak neoantigens as their density increases during tumor growth (31,32). While MSI CRCs are generally considered highly immunogenic due to their elevated mutation burdens (38), our findings indicate that MSS CRCs also exhibit evidence of immune surveillance, consistent with studies that find immune evasion is acquired early during both MSI and MSS CRC evolution (19). Further studies comparing younger and older CRCs may reveal why younger tumors have fewer early functional constraints.

A central limitation of this study is its inferential nature. Reconstruction of growth dynamics relies on model-based assumptions, including constant error rates across growth phases and the structure of the simulated growth process. Although we generated millions of growth trajectories across a large range of model assumptions, some biologically plausible scenarios may not have been explored, and the underlying histories are not uniquely identifiable. Neutrality is assumed, but weak or spatially localized selection across spatially heterogeneous microenvironments (39) cannot be excluded (2,40). Within these limitations, the consistency of prolonged slow early growth across independent model structures and sampling configurations suggests that early growth constraints are a robust feature of CRC tumorigenesis.

The temporal resolution of fCpG clocks provides unique opportunities to reconstruct growth trajectories and quantify the timing and strength of functional constraints. We show that CRCs differ substantially in mitotic age, with these differences arising largely from early constraints that produce prolonged, slow growth while tumors remain small and clinically undetectable. The associations between mitotic age and immune infiltration and exhaustion suggest that immunoediting (17) may contribute to this growth variation. These findings extend the Big Bang model (4) by showing that neutral clonal evolution can coexist with substantial constraints on net tumor expansion. More broadly, fast molecular clocks may help identify mechanisms that naturally determine whether early neoplastic populations expand, persist, or become extinct.

METHODS

Samples: The 12 tumors (Table S1) were collected as excess tissue obtained in the course of routine clinical care at the University of Southern California Keck School of Medicine at the Norris Cancer Center with local Institutional Review Board Approval. The fresh gross specimens were subdivided into small 0.5 cm^3 chunks from opposite tumor sides (Side A and Side B) and single tumor glands were obtained with an EDTA-washout method (23). Gland methylation was measured with Illumina EPIC arrays using the Restore Protocol (41). Data were processed to beta values using minfi (42).

Colorectal tumor fCpG molecular clocks: A molecular clock based on the methylation status of 1,164 fast-drifting CpG (fCpG) sites was used to resolve CRC growth trajectories. These sites accumulate methylation changes stochastically during each cell division, acting as a high-resolution mitotic recorder. The fCpG sites were selected based on the procedure described in Reference 43.

Approximate Bayesian computation (ABC) framework: Neutral tumor expansion was modeled using a neutral cell-fate branching model within an ABC framework, which constrains the space of plausible growth trajectories rather than identifying a unique solution. This ABC framework was used to identify growth trajectories consistent with the data. The simulations tracked individual cells within a gland and started with a single progenitor cell with an initialized CpG methylation barcode. For computational efficiency, 100 fCpG sites represented the full set of 1,164 sites for each simulation. Each fCpG site could randomly flip with each cell division (estimated error rate 10^{-3} (Fig S1)). The progenitor cell expanded to form a gland, which splits into two glands when its stem-cell population exceeds 80 cells. The resulting glands continued to split whenever their stem-cell populations exceeded this threshold, with a total of 18 rounds of gland doubling needed to reach a detectable tumor (~2.6 billion total cells with ~10,000 cells per gland, including tumor stem and non-stem cells). Each growth trajectory was simulated under three alternative spatial branching configurations, with each simulation producing eight terminal glands. Side B separated from side A at the first gland split. Within each side, three additional gland lineages branched sequentially from the continuing reference lineage. These branching events occurred at gland splits 3, 4, and 5 in the early configuration; splits 8, 9, and 10 in the middle configuration; or splits 12, 13, and 14 in the late configuration. All three configurations were simulated for each growth trajectory. The designations A1 and B1 identify the continuing reference lineages rather than older physical glands; all eight terminal glands were simulated to the same endpoint.

Tumor growth is subdivided into 5 phases. The first phase (P1) is fixed, with death-free, pure-birth exponential division from a single progenitor cell into the first tumor gland. The next 3 phases (P2, P3, and P4) have variable numbers of gland divisions and use two alternative cell-fate regimes, including death-free regimes that model exponential growth. (These 3 phases were expanded to 5 phases to test the robustness of the N-shaped growth curves, see below). In the fixed-net-growth regime, individual cells were randomly assigned to produce zero, one, or two surviving daughter cells, while the aggregate numbers of these outcomes were constrained to produce a specified net population increase. Because some cells produced no surviving descendants, individual lineages could become extinct stochastically. In the fractional-division regime, a specified fraction of cells produced two surviving daughters and the remaining cells produced one surviving daughter. This regime contained no cell loss and therefore did not permit lineage extinction. Death-free exponential doubling occurred when all cells produced two surviving daughters. The last phase (P5) starts after 18 gland doublings when the tumor reaches a detectable size, and continued cell turnover was modeled by assigning equal fractions of cells to produce two or zero surviving daughters, thereby maintaining a stable population size while permitting stochastic lineage extinction. This last phase mirrors the clinical observation that detectable tumors nearing carrying capacity grow slowly (14).

A total of 9.4×10^6 individual growth histories were simulated. The ABC rejection algorithm compared the simulated methylation patterns with the observed experimental data using 7 summary statistics: average CpG methylation gland variance for side A, average CpG methylation gland variance for side B, average CpG methylation gland variance across both sides, overall variance based on the average CpG methylation of all 8 glands, mean Pearson correlation among the 4 glands on side A, mean Pearson correlation among the 4 glands on side B, and mean Pearson correlation between glands on opposite sides. Simulations were accepted if all seven summary statistics were within ±10% of their observed values.

Private or subclonal somatic mutations that arise during growth and reach detectable frequencies by WES (i.e. VAF > 5%) during neutral expansion arise during the first few divisions and reach near clonal frequencies in CRC glands (20). The simulation labels the first 8 cells and tracks their frequencies in the final glands. Accepted ABC runs using the joint fCpG summary statistics and the private mutation data required each of the 8 glands to also be “clonal” or have >90% of its cells trace back to one of the first 8 cells. Multiple growth histories may satisfy these constraints, and the inferred trajectories represent a set of solutions consistent with the observed data rather than a unique reconstruction.

Synthetic tumor validation: To determine whether rejection-sampling ABC using the selected summary-statistic distance criterion could recover known growth parameters, synthetic tumors with known growth histories were generated and analyzed using the same framework applied to the experimental tumors. Fifty synthetic tumors spanning a range of total division counts and growth-phase configurations were generated. For each synthetic tumor, candidate growth trajectories were compared with the synthetic data using the same seven summary statistics and tolerance intervals used for the experimental tumors. Simulations in which all seven summary statistics fell within their respective tolerance intervals (“7-hit” acceptance) were retained, producing a distribution of accepted growth parameters for each synthetic tumor.

Parameter recovery was assessed by comparing the means and 95% simulation intervals of the accepted simulations with the known ground-truth values. Total mitotic age, overall architectural variance, regional gland variance, and Pearson correlation structure were recovered accurately across the synthetic tumors, demonstrating that these parameters can be estimated using rejection-sampling ABC with the selected distance criterion. Growth-phase interval estimates exhibited greater uncertainty, particularly in older synthetic tumors with noisier trajectories (Fig. S7). Thus, the ABC framework reliably estimates overall growth architecture and mitotic age, although precise latent phase boundaries are less identifiable.

Normalization of growth curves: To identify conserved growth patterns across the cohort, accepted growth trajectories were normalized for both time and magnitude. The terminal stationary phase (phase 5) was excluded.

1. Temporal Normalization: The absolute mitotic age (total divisions) was rescaled to a normalized timeline from 0 (tumor initiation) to 1 (detectable tumor size at the 18th gland split).
2. Magnitude Normalization: Tumor size was represented as log2 doublings, with the expansion phase analyzed up to 18 gland doublings (approximately 2.6×10^9 cells).

The median trajectory and 95% simulation intervals were calculated from the posterior distribution of accepted runs for each of the 12 tumors to visualize the 'N-shaped' growth dynamics.

Testing the early death-free expansion scenario: To test whether growth remained death-free after formation of the first tumor gland, we performed a constrained ABC analysis in which the first seven gland-doubling intervals were restricted to pure-birth expansion (d = 0). The acceptance frequencies (the ratio of simulations meeting the distance threshold to the total number of simulations) were compared between the unconstrained N-shaped model and the constrained early death-free exponential model. A reduction in acceptance frequency under the death-free exponential constraint was used as evidence of early slow growth.

Quantification of growth constraints: The strength of growth constraint was inferred from the cell-loss fraction (d/b) across the reconstructed timeline, where (d) represents the cell-loss rate and (b) represents the birth rate. The N-shaped growth pattern was characterized by three distinct phases: (1) an initial expansion with low cell loss, (2) a period of greater growth

suppression with high cell loss, and (3) subsequent growth acceleration associated with lower cell loss.

N-index calculation: To evaluate N-shaped growth dynamics characterized by an initial growth constraint, stronger constraint, and subsequent relaxation, the simulations were expanded from three to five variable growth phases. We calculated an N-index from the cumulative growth trajectories. Each accepted trajectory was evaluated at evenly spaced points along normalized mitotic time, spanning 0 to 1. Local growth rates were estimated by calculating first differences between adjacent cumulative gland-doubling values using np.diff, generating a slope profile across the trajectory.

To reduce stochastic noise from individual accepted simulations, slope profiles were lightly smoothed using a centered moving-average filter (default width = 3). The position of the minimum smoothed slope was identified as the trough corresponding to the slowest-growth phase. A local window surrounding this trough (20% of slope positions) was defined as the middle phase. Mean slopes before and after this trough window were respectively defined as the early and late growth phases. The N-index was then calculated as: N-index = ((Early slope+ Later slope)/2) – Middle slope.

Higher positive N-index values indicate a stronger transient slowing relative to the preceding and subsequent growth intervals. N-index values near zero indicate relatively steady growth, whereas negative values indicate that middle growth rates exceed the early and/or late growth rates.

Tumor gland phylogenetic reconstruction with IQ-TREE 2

To reconstruct Maximum Likelihood (ML) phylogenies, gland fCpG methylation was transformed into three discrete states—State 1: beta < 0.40; State 2: beta between 0.40 and 0.60; State 3: beta > 0.60. Ancestral root states were estimated by averaging fCpG methylation across all eight glands and transforming the resulting values into the same three-state representation.

For each of the 12 tumors, three 8-gland datasets were analyzed: (1) the experimental gland methylation data; (2) synthetic gland methylation datasets generated from accepted ABC simulations; and (3) synthetic gland methylation datasets generated from rejected ABC simulations constrained to undergo rapid death-free initial exponential growth.

Each 8-gland dataset was converted into a discrete multistate alignment and analyzed using IQ-TREE 2 (26). Maximum Likelihood phylogenetic trees were constructed using the MORPH data type with an ordered multistate substitution model (-st MORPH -m ORDERED -o Root -bb 1000). To compare between trees, topological metrics between the experimental and simulated trees were computed from the resulting NEWICK tree files (.treefile):

1) Mean Root-to-Tip Distance: Computed as the average path length from the ancestral root node to all terminal leaf nodes (glands), representing overall accumulated mitotic depth.
2) Tip-to-Total Branch Length Ratio: For each terminal tip, the length of its terminal branch was divided by the total branch length of the corresponding tree. Thus, each gland or cell (see below) contributed one tip-to-total branch-length ratio.

Single-cell CRC phylogenetic reconstruction via IQ-TREE 2: To mitigate the sparse coverage of single-cell whole-genome bisulfite sequencing (scWGBS) data (22), only X-chromosome CpG methylation from male tumors was analyzed (N = 8 tumors; Table S3). For each CpG site, methylation values <0.4 were converted to 0 (unmethylated), values >0.6 were converted to 1 (methylated), and intermediate values (0.4–0.6) were treated as missing data and discarded. X-chromosome fluctuating CpG (fCpG) sites (N = 165,674) were defined as CpG sites with average methylation between 0.4 and 0.6 across all analyzed tumor cells. Because of

sparse single-cell coverage, each cell contained data for an average of ~35,000 fCpG sites. Trees were reconstructed with IQ-TREE 2 (v2.x) using the binary data type (-st BIN), the GTR2+FO+R3 substitution model (-m GTR2+FO+R3), an inferred ancestral root (-o ROOT), and 1000 ultrafast bootstrap replicates (-bb 1000). For each tumor, the ancestral root sequence was estimated from the average methylation of all its tumor cells. Average methylation values <0.4 were assigned 0, values >0.6 were assigned 1, and intermediate values (0.4–0.6) were randomly assigned either 0 or 1 to produce a binary root sequence required by IQ-TREE. One male tumor (C01) was not analyzed because its apparent mitotic age was much higher than the other seven male tumors. The inferred trees successfully recovered the spatial sampling for more than 90% of the single cells (Table S3).

The ABC simulation was modified for single-cell analysis by randomly selecting one cell early during growth to represent each sampled region, using only one allele to model X-chromosome methylation in male tumors. Three growth scenarios were simulated: (1) rapid, death-free exponential growth; (2) slow early growth; and (3) slow linear growth followed by exponential growth (N-shaped growth). Each simulation included 100 terminal steady-state divisions to model the prolonged plateau phase of clinically detectable tumors.

Four simulated regions, each containing 40 cells, were generated for every simulation and analyzed with IQ-TREE 2 as described above. The methylation profile of the original starting cell was provided as the ancestral root sequence. Because experimental tumors differed in their total branch lengths (mean root-to-tip distances), simulations with different total numbers of divisions were selected to match the experimental branch lengths before comparing tree topology.

The experimental and simulated trees were compared using the tip-to-total branch length ratio, as described for the gland trees. Intra-region patristic distances, which quantify the relatedness of cells within a sampled region, were also calculated from the reconstructed Newick phylogenies using Biopython (Bio.Phylo).

Estimation of mitotic age from TCGA DNA methylation data: Mitotic ages of TCGA CRCs (27) were estimated from their downloaded methylation arrays using a fCpG molecular clock calibration curve. Briefly, stochastic simulations were used to generate the expected relationship between fCpG methylation variance and cumulative cell divisions (mitotic age). This simulation-derived reference curve served as a “golden ruler” that maps methylation variance to the corresponding number of cell divisions.

For each TCGA sample, fCpG variance was calculated. Mitotic age was then estimated by linear interpolation of the observed variance onto the simulation-derived calibration curve. Specifically, each observed variance was mapped to the corresponding position on the reference curve, and the number of cell divisions was estimated by interpolation between adjacent simulated values. The resulting interpolated value was recorded as the estimated mitotic age of the tumor. This approach assumes that the relationship between fCpG variance and cumulative cell divisions established by the simulations is applicable to TCGA CRCs, allowing methylation variance to be converted into an estimate of cumulative mitotic age.

Statistical software and visualization: Simulations and ABC inferences were performed using custom Python scripts. Statistical analyses were performed in R and Python. Normalized expansion plots were generated using Matplotlib and Seaborn, displaying median trajectories and shaded 95% simulation intervals across all 12 tumors.

Data availability: The tumor gland EPIC methylation array data will be deposited into GEO. The single cell CRC data are in GSE97693.

Code availability: Python code used to simulate CRC growth trajectories and perform fCpG methylation-based approximate Bayesian computation is publicly available at https://github.com/darrylshibata/fCpG_Mitotic_Clock.

Acknowledgements: The authors thank Calum Gabbutt for deriving the fCpG sites used in the analysis. This work was supported by the National Institutes of Health (P01CA196569 and R01CA271237).

Author contributions: The authors conceived and performed the research.

Competing interests: None

References

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

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

3. Reiter, J. G. et al. Minimal functional driver gene heterogeneity among untreated metastases. Science 361, 1033–1037 (2018).

4. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. Nat. Genet. 47, 209–216 (2015).

5. Heide, T. et al. The co-evolution of the genome and epigenome in colorectal cancer. Nature 611, 733–743 (2022).

6. Househam, J. et al. Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature 611, 744–753 (2022).

7. Chkhaidze K, Heide T, Werner B, Williams MJ, Huang W, Caravagna G, Graham TA, Sottoriva A. Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. PLoS Comput Biol. 2019 Jul 29;15(7):e1007243. doi: 10.1371/journal.pcbi.1007243. PMID: 31356595; PMCID: PMC6687187.

8. Noble R, Burri D, Le Sueur C, Lemant J, Viossat Y, Kather JN, Beerenwinkel N. Spatial structure governs the mode of tumour evolution. Nat Ecol Evol. 2022 Feb;6(2):207-217. doi: 10.1038/s41559-021-01615-9. Epub 2021 Dec 23. PMID: 34949822; PMCID: PMC8825284

9. Zapata L, Pich O, Serrano L, Kondrashov FA, Ossowski S. *Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome.* Genome Biology. 2018;19:67. doi:10.1186/s13059-018-1434-0.

10. Weghorn D, Sunyaev S. *Bayesian inference of negative and positive selection in human cancers.* Nature Genetics. 2017;49(12):1785–1788. doi:10.1038/ng.3987.

11. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ. *Universal Patterns of Selection in Cancer and Somatic Tissues.* Cell. 2017;171(5):1029–1041.e21. doi:10.1016/j.cell.2017.09.042.

12. Maley, C. C. et al. Classifying the evolutionary and ecological features of neoplasms. Nat. Rev. Cancer 17, 605–619 (2017).

13. Jones MG, Sun D, Min KHJ, Colgan WN, Wang H, Török T, Ribeiro J, Xue J, Cardoso EC, Rong Y, Tian L, Weir JA, Chen VZ, Koblan LW, Yost KE, Mathey-Andrews N, D'Souza E, Russell AJC, Stickels RR, Balderrama KS, Rideout WM 3rd, Dai M, Marrero G, Kumar V, Saqi A, Herzberg B, Izar B, Chang HY, Lee JH, Jacks T, Chen F, Weissman JS, Yosef N, Yang D. Spatiotemporal lineage tracing reveals the dynamic spatial architecture of tumor growth and metastasis. Nat Genet. 2026 Sep 2. doi: 10.1038/s41588-026-02739-z. Epub ahead of print. PMID: 42687033.

14. Norton, L. A Gompertzian model of human breast cancer growth. Cancer Res. 48, 7067–7071 (1988).

15. Castorina, P. & Carco', D. Nutrient supply, cell spatial correlation and Gompertzian tumor growth. Theory Biosci. 140, 197–203 (2021).

16. Laird, A. K. Dynamics of tumor growth. Br. J. Cancer 18, 490–502 (1964).

17. Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. Annu. Rev. Immunol. 22, 329–360 (2004).

18. McGranahan, N. et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science 351, 1463–1469 (2016).

19. Lakatos E, Williams MJ, Schenck RO, Cross WCH, Househam J, Zapata L, Werner B, Gatenbee C, Robertson-Tessi M, Barnes CP, Anderson ARA, Sottoriva A, Graham TA. Evolutionary dynamics of neoantigens in growing tumors. Nat Genet. 2020 Oct;52(10):1057-1066. doi: 10.1038/s41588-020-0687-1. Epub 2020 Sep 14. PMID: 32929288; PMCID: PMC7610467.

20. Tarabichi, M. et al. A practical guide to cancer subclonal reconstruction from DNA sequencing. Nat. Methods 18, 144–155 (2021).

21. Gabbutt, C. et al. Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues. Nat. Biotechnol. 40, 720–730 (2022).

22. Bian S, Hou Y, Zhou X, Li X, Yong J, Wang Y, Wang W, Yan J, Hu B, Guo H, Wang J, Gao S, Mao Y, Dong J, Zhu P, Xiu D, Yan L, Wen L, Qiao J, Tang F, Fu W. Single-cell multiomics sequencing and analyses of human colorectal cancer. Science. 2018 Nov 30;362(6418):1060-1063. doi: 10.1126/science.aao3791. PMID: 30498128.

23. Ryser, M. D., Min, B. H., Siegmund, K. D. & Shibata, D. Spatial mutation patterns as markers of early colorectal tumor cell mobility. Proc. Natl Acad. Sci. USA 115, 5774–5779 (2018).

24. Schepers AG, Snippert HJ, Stange DE, van den Born M, van Es JH, van de Wetering M, Clevers H. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science. 2012 Aug 10;337(6095):730-5. doi: 10.1126/science.1224676. Epub 2012 Aug 1. PMID: 22855427

25. Van Egeren, D. et al. Polyclonal origins of human premalignant colorectal lesions. Nature 650, 1017–1024 (2026).

26. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution, 37(5), 1530–1534.

27. Thorsson, V. et al. The immune landscape of cancer. Immunity 48, 812–830.e14 (2018).

28. Batlle, E. & Massagué, J. Transforming growth factor-β signaling in immunity and cancer. Immunity 50, 924–940 (2019).

29. Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., ... & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nature Medicine, 24(5), 541–550.

30. de Visser, K. E., & Joyce, J. A. (2023). The evolving tumor microenvironment: From tumorigenesis to metastasis. Cancer Cell, 41(3), 374–403.

31. Ohashi, P. S. et al. Ablation of tolerance and induction of diabetes by virus infection in viral antigen transgenic mice. Cell 65, 305–317 (1991).

32. Spiotto, M. T. et al. Increasing tumor antigen expression overcomes ignorance to solid tumors via cross-presentation by bone marrow-derived stromal cells. Immunity 17, 737–747 (2002).

33. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. Nat. Rev. Immunol. 15, 486–499 (2015).

34. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. 3, 991–998 (2002).

35. Monyak, D.L., Holloway, S.T., Gumbert, G.J. et al. Mapping the temporal landscape of breast cancer using epigenetic entropy. Commun Biol 8, 1477 (2025).

36. Kim, K. and Shibata, D. (2026), Differences between mitotically old and young endometrial tumors. J. Pathol. https://doi.org/10.1002/path.70090

37. Wells, D. K. et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. Cell 183, 818–834.e13 (2020).

38. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. N. Engl. J. Med. 372, 2509–2520 (2015).

39. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. Cell 182, 1341–1359.e19 (2020).

40. Heide, T., Zapata, L., Williams, M. J., Werner, B., Caravagna, G., Barnes, C. P., Graham, T. A. & Sottoriva, A. Reply to “Neutral tumor evolution?”. Nat. Genet. 50, 1633–1637 (2018).

41. Illumina. Infinium HD FFPE Restore Protocol (Illumina, 2011).

42. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369 (2014).

43. Manojlovic V, Gabbutt C, Shibata D, Noble R. Fluctuating DNA methylation sites encode colorectal tumour growth history. *bioRxiv*. 2026. doi:10.64898/2026.06.04.730217.

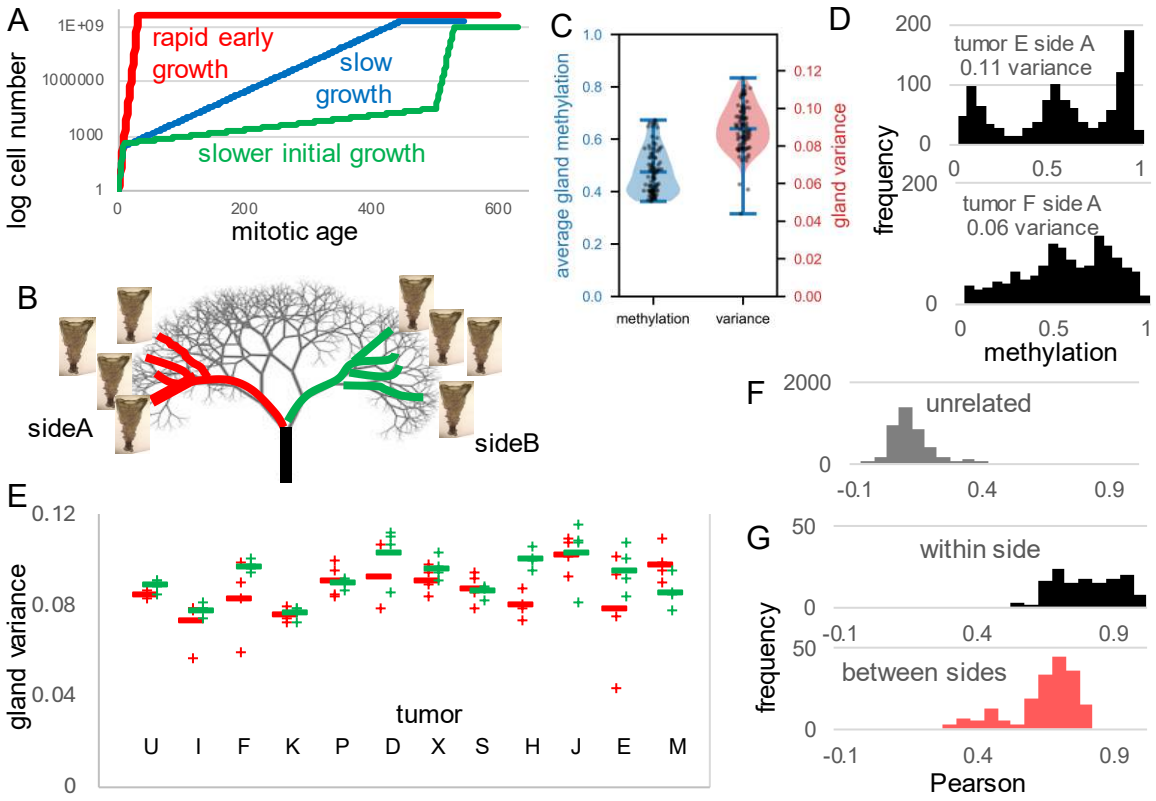
Table 1) ABC results

Tumor	Mitotic age*	P5 divisions*	acceptance count	acceptance rate
D	340 (294-405)	97 (10-200)	2517	2.8×10^{-4}
E	472 (397-574)	80 (10-200)	565	6.0×10^{-5}
F	357 (314-414)	158 (50-200)	1634	1.7×10^{-4}
H	382 (335-443)	148 (50-200)	786	8.3×10^{-5}
I	390 (335-440)	43 (10-200)	55	5.8×10^{-6}
J	359 (302-433)	54 (10-100)	403	4.3×10^{-5}
K	405 (361-458)	176 (50-200)	665	7.1×10^{-5}
M	532 (411-632)	39 (10-100)	144	1.5×10^{-5}
P	354 (307-422)	124 (50-200)	1145	1.2×10^{-4}
S	378 (329-441)	157 (50-200)	4375	4.6×10^{-4}
U	335 (290-390)	72 (10-200)	541	5.7×10^{-5}
X	357 (314-411)	130 (10-200)	2721	2.9×10^{-4}
Average	388.9	106.5		

* Mean and 95% simulation interval

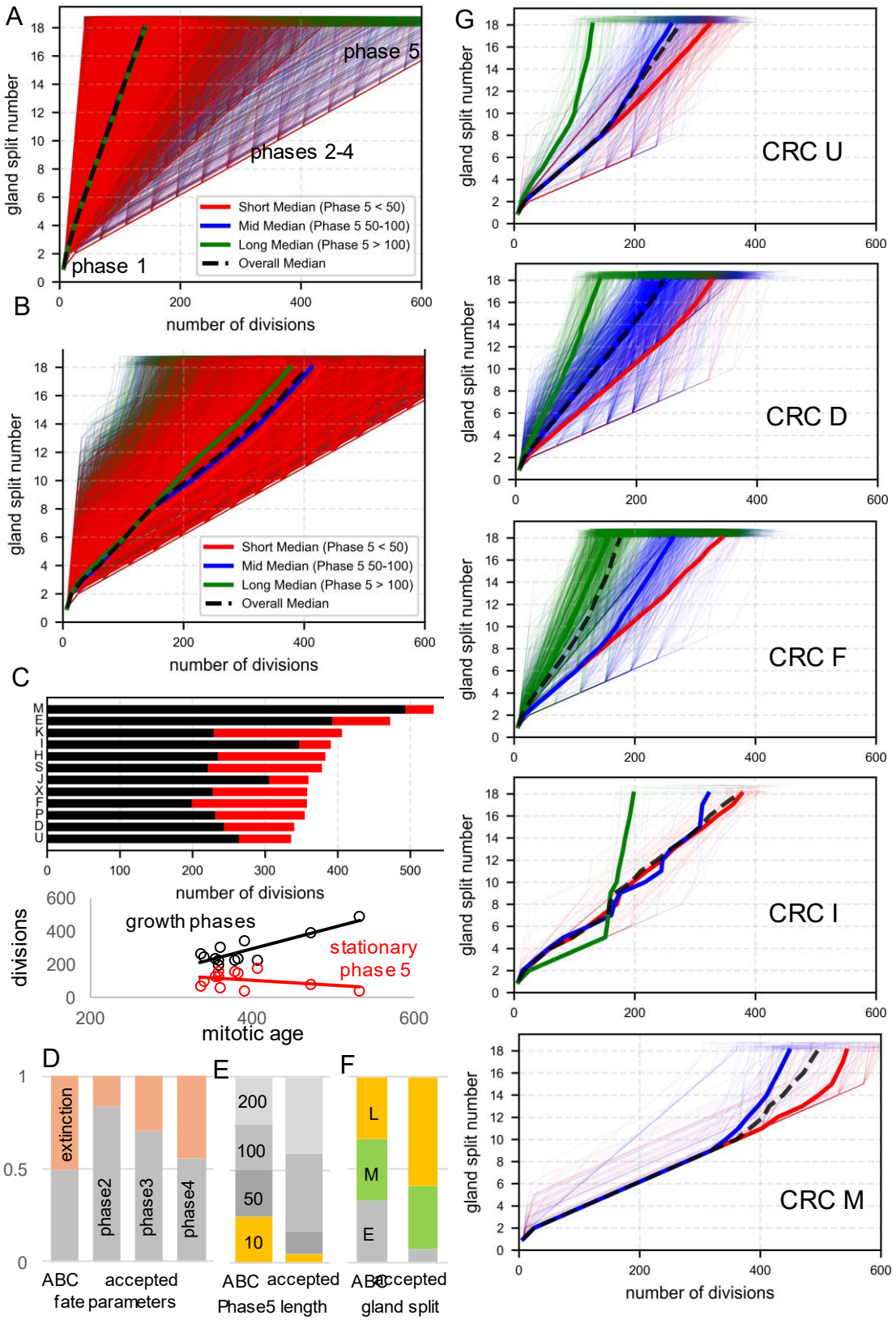
Figure Legends

Figure 1) Tumor gland fCpG methylation



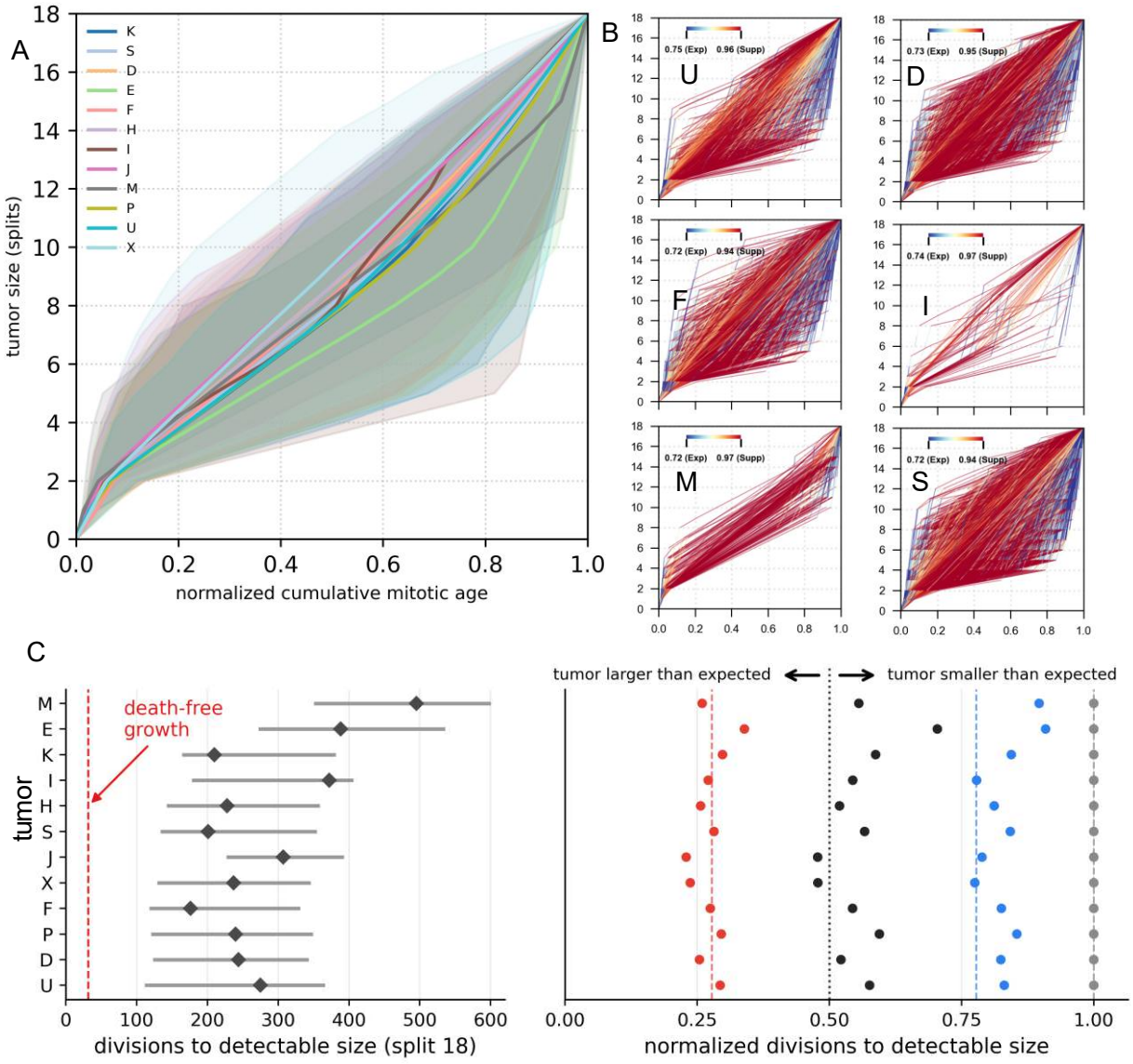
A: Tumor mitotic ages may vary when cell loss occurs during growth. Illustrated are tumor growth trajectories with death-free exponential growth and slower growing tumors. All tumors may have static terminal growth phases where cell division is balanced by cell loss. Millions of different growth trajectories generated by ABC were compared with the gland fCpG experimental data.
B: fCpG methylation was measured in four glands from each of two opposite tumor sides. Within-side comparisons interrogate more recent regional growth, whereas between-side comparisons reflect earlier divergence from the tumor progenitor.
C: Average fCpG gland methylation is about 50%, consistent with fluctuating but balanced methylation and demethylation errors. fCpG gland variances were variable.
D: Individual fCpG methylation values showed high variance, W-shaped distributions in younger gland expansions and lower variance, unimodal distributions in older gland expansions.
E: Gland variances for the 12 tumor samples were variable within a tumor but generally similar between sides (side A green and side B red, bars are means).
F: fCpG methylation patterns correlated poorly between glands from unrelated tumors, indicating that progenitor cells from each tumor start with different fCpG patterns.
G: fCpG methylation patterns were more correlated between glands within the same tumor side than between opposite sides, consistent with initial synchronization to the progenitor cell pattern followed by progressive desynchronization during tumor growth.

Figure 2) ABC of tumor growth



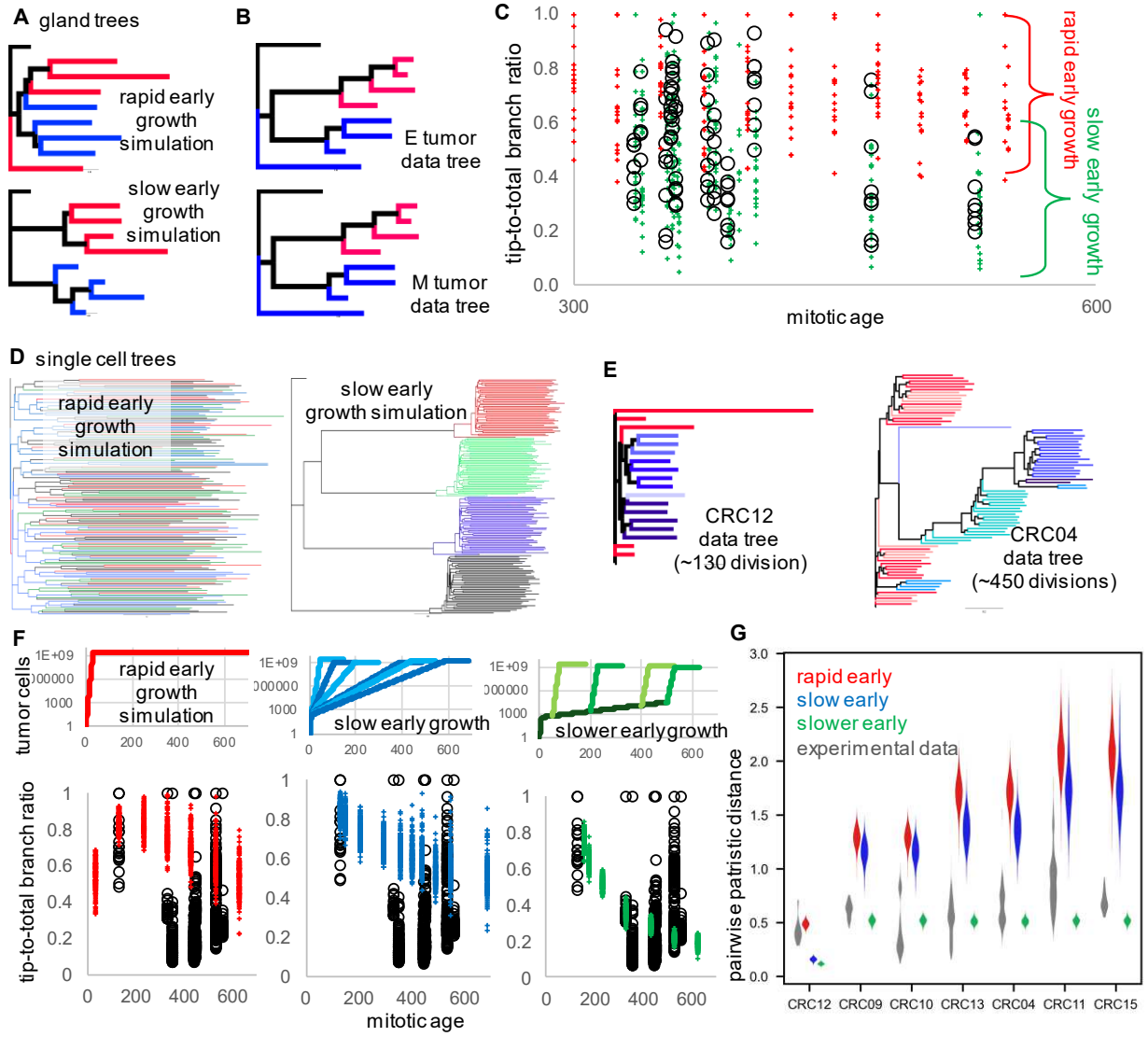
A: Approximately 9.4 million random growth trajectories were generated, of which 40,000 are shown. Trajectories begin with exponential growth without cell loss to form the first tumor gland (phase 1), followed by variable growth and cell loss (phases 2 to 4), and a terminal stationary interval in which cell division is balanced by cell loss (phase 5). The x axis shows divisions since the progenitor cell and the y axis shows the gland-split number, equivalent to the $\log_2$ number of tumor glands. The leftmost trajectory represents death-free exponential growth. Colors classify trajectories by their short (<50 divisions, red), intermediate (50-100 divisions, blue), or long (>100 divisions, green) terminal phase 5 intervals. Colored bold lines show the median trajectories within these groups; the dashed black line shows the overall median.
B: Adding the requirement that tumor-subclonal mutations have near-clonal frequencies within individual glands preferentially excluded very rapid growth trajectories.
C: ABC-accepted simulations had mean mitotic ages between 335 and 532 divisions across the 12 tumors. Black indicates divisions during the expansion phases and red indicates the stationary phase 5 interval. Differences in mitotic age were driven primarily by the expansion phases rather than phase 5.
D: Parameter values were balanced among the millions of random growth trajectories. ABC-accepted simulations showed progressively greater enrichment for lineage extinction during later growth phases.
E: Longer stationary phase 5 intervals were enriched among ABC-accepted simulations.
F: Later gland splitting was enriched among ABC-accepted simulations.
G: Examples of ABC-accepted growth trajectories for five tumors with different mitotic ages. Older tumors had longer and slower early growth phases. Trajectories are colored for short, intermediate, or long phase 5 stationary intervals as in panel A.

Figure 3) Prolonged slow growth before tumors reach detectable size



A: ABC-accepted growth curves for the 12 tumors, normalized to the timing of the 18th gland split (i.e., detectable size), showed slow initial growth followed by further slowing and then acceleration, or "N-shaped" growth. The lines represent median growth, with 95% simulation intervals.
B: Examples of six tumors with their ABC-accepted growth curves showing N-shaped growth. The phases are color-coded by inferred cell loss fraction, with blue indicating stronger and red indicating weaker constraints during growth. ABC-accepted growth curves for all 12 tumors are presented in Fig. S8.
C: Mitotic timing of tumor expansion. Left, diamonds indicate the median number of divisions required to reach the 18th gland split (detectable tumor size), with bars showing 95% simulation intervals. The red dashed line indicates the approximately 32 divisions required under death-free exponential growth. Right, median timings of gland splits 5 (red), 9 (black), and 14 (blue), normalized to the timing of split 18 (grey). Dashed vertical lines indicate the expected timings if successive gland doublings were uniformly distributed across the expansion. Most tumors were smaller than expected at intermediate normalized division times, consistent with prolonged constrained growth followed by acceleration near the end of expansion.

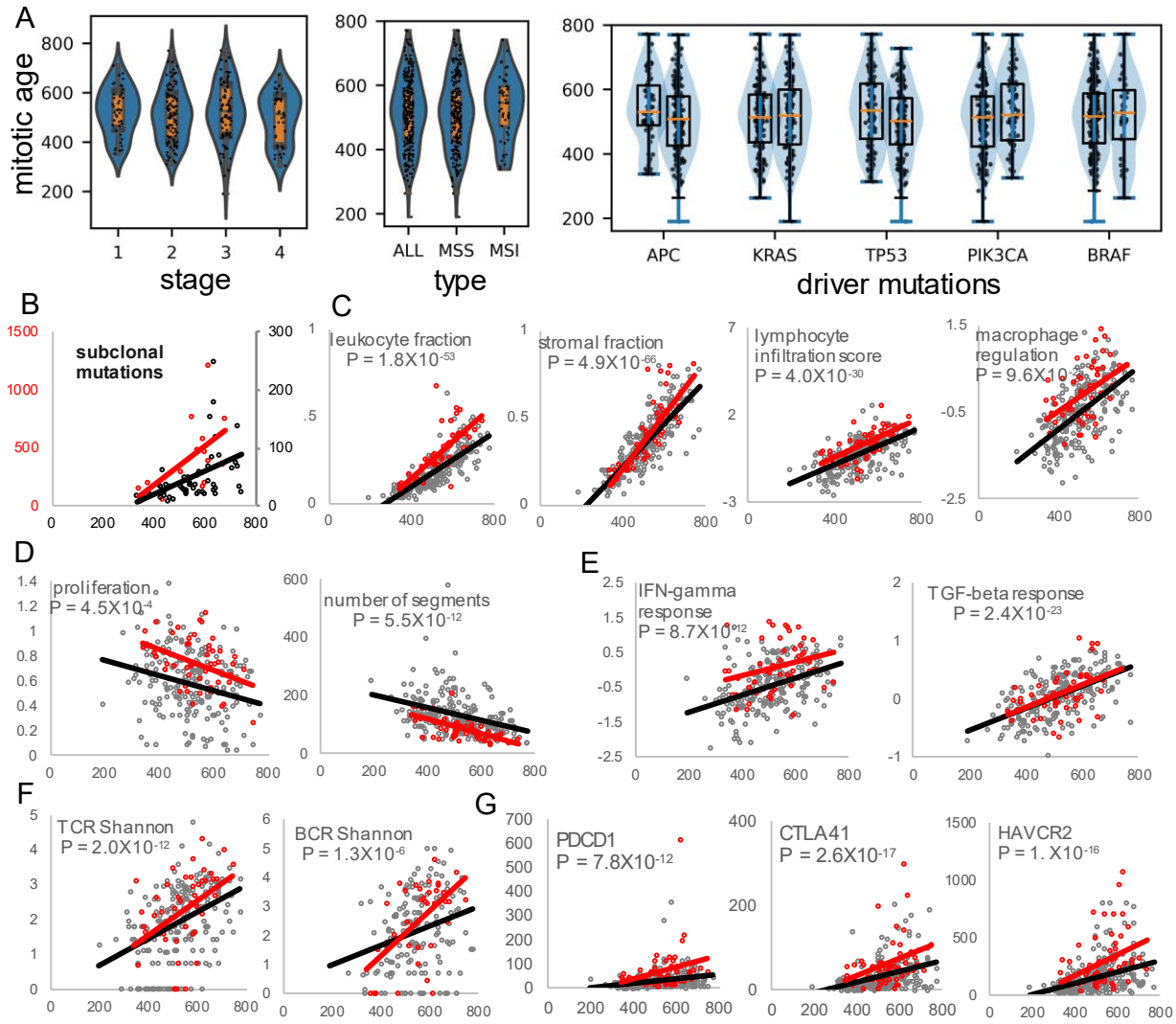
Figure 4) Tumor phylogenies



A: Maximum-likelihood (IQ-TREE) tumor gland trees. Simulations of early death-free exponential growth yield star-like maximum-likelihood gland trees with long terminal branches, whereas simulations of slow early growth yield longer internal branches and shorter terminal branches.
B: Maximum-likelihood gland trees inferred from fCpG data (Tumors E and M are illustrated) more closely resemble simulations with slow early growth, with longer internal branches and shorter terminal branches. Red and blue terminal branches denote side A and side B glands, respectively.
C: Tree topologies were summarized by the ratio of terminal branch length to total tree length across mitotic ages. Each cross represents a simulated gland under death-free exponential growth (red) and slow early growth (green) scenarios. Tumor gland data distributions (circles) overall differed significantly from death-free exponential simulations ($D = 0.4875$, $p < 0.001$), but were statistically indistinguishable from slow initial growth models ($D = 0.1497$, $p = 0.0736$; two-sample Kolmogorov-Smirnov test). The KS test favored slow initial growth for all but two of the 12 tumors (K and X).
D: Maximum-likelihood (IQ-TREE) trees of single-cell simulations illustrate the star-like phylogeny (long terminal branches) of death-free exponential growth versus the longer internal and shorter terminal branches of slow initial growth. Terminal branch colors represent cells from four different regions, with poorer spatial resolution for the simulated death-free exponential tree.

E: Representative maximum-likelihood single-cell CRC trees (data from Ref. (22)). CRC12 (18 cells) is mitotically young and has long terminal branches and is more consistent with death-free exponential early growth. CRC04 (80 cells) is mitotically older and has longer internal branches and relatively shorter terminal branches and is more consistent with slow early growth. Branch colors indicate different spatial locations of the sampled cells. Other tumor trees are presented in Fig. S9.
F: Growth curves for simulated death-free exponential growth (red), slow initial growth (blue), and slower N-shaped initial growth (green). An additional 100 terminal stationary divisions were added to the slow-initial and N-shaped growth simulations. Simulated tip-to-total branch ratios varied with mitotic ages. Each value represents a single cell. The maximum-likelihood tumor trees more closely resembled the slow or slower N-shaped initial growth simulations except for the youngest tumor (CRC12), where death-free exponential growth was favored (KS test).
G: Pairwise patristic distances between cells sampled from the same spatial region were small in the experimental tumor trees and were more consistent with simulations of slow initial growth (blue) or N-shaped growth (green), except for the youngest tumor (CRC12), for which death-free exponential growth was favored (Supplementary File; KS test).

Figure 5) Intrinsic and extrinsic differences between young and older TCGA CRCs



A: Mitotic ages were similar for CRCs with different intrinsic features. For stage, P = 0.482 (all CRCs), P = 0.426 (MSS), and P = 0.542 (MSI). For subtype, Kruskal P = 0.360. For driver mutations (right violin is mutation positive), P = 0.0679 (APC), P = 0.852 (KRAS), P = 0.0134 (TP53), P = 0.106 (PIK3CA), and P = 0.829 (BRAF).
B: Subclonal mutations detectable by bulk sequencing (data from reference 1) were significantly higher in older MSI (red, P = 0.0058) and MSS (black, P = 0.0016) CRCs.
C: CRC mitotic ages correlated with extrinsic TME features with greater lymphocytic and macrophage infiltration in older CRCs. Shown are MSI (red) and MSS (grey) tumors; Pearson P values are reported for all CRCs (MSI and MSS).
D: Younger CRCs had more proliferation and greater numbers of altered chromosomal segments.
E: Older CRCs had higher IFN-gamma (immunostimulatory) and higher TGF-beta (immunoinhibitory), typical of immune cell exhaustion.
F: Greater sequence evidence of chronic immune stimulation in older CRCs with more TCR and BCR diversity (Shannon).
G: Higher expression of genes associated with immune cell exhaustion (PDCD1, CTLA4, and HAVCR2) was observed in older CRCs. P values are from Spearman rank correlations across all CRCs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- EarlyCRCgrowthNGsupplementaryinformation.pdf