Supplementary Information

Early Colorectal Cancer Growth Is Often Slow and Prolonged

Marc D. Ryser and Darryl Shibata

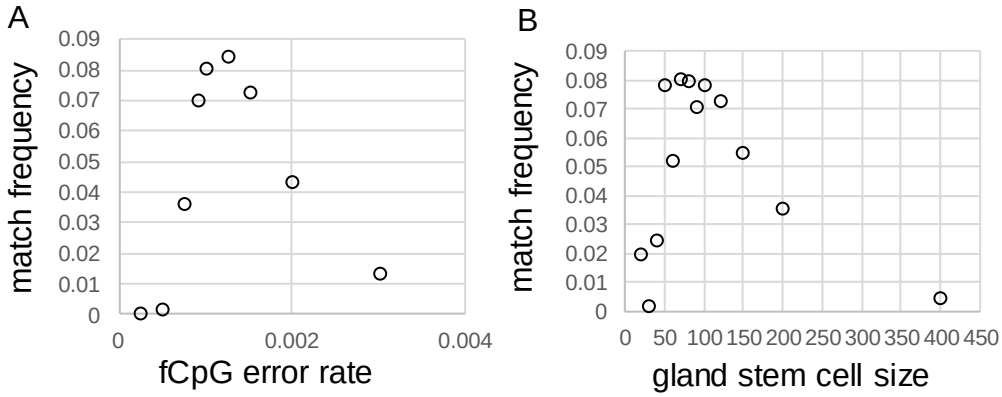


Figure S1. Calibration of fCpG simulation parameters.
A: ABC acceptance frequencies as a function of fCpG methylation error rate. Maximum agreement between simulated and observed CRC methylation distributions was obtained at approximately 10^{-3} flips per division.
B: ABC acceptance frequencies as a function of effective gland stem cell number. Maximum agreement was observed at approximately 80 stem cells per gland. These values were used in all subsequent simulations.

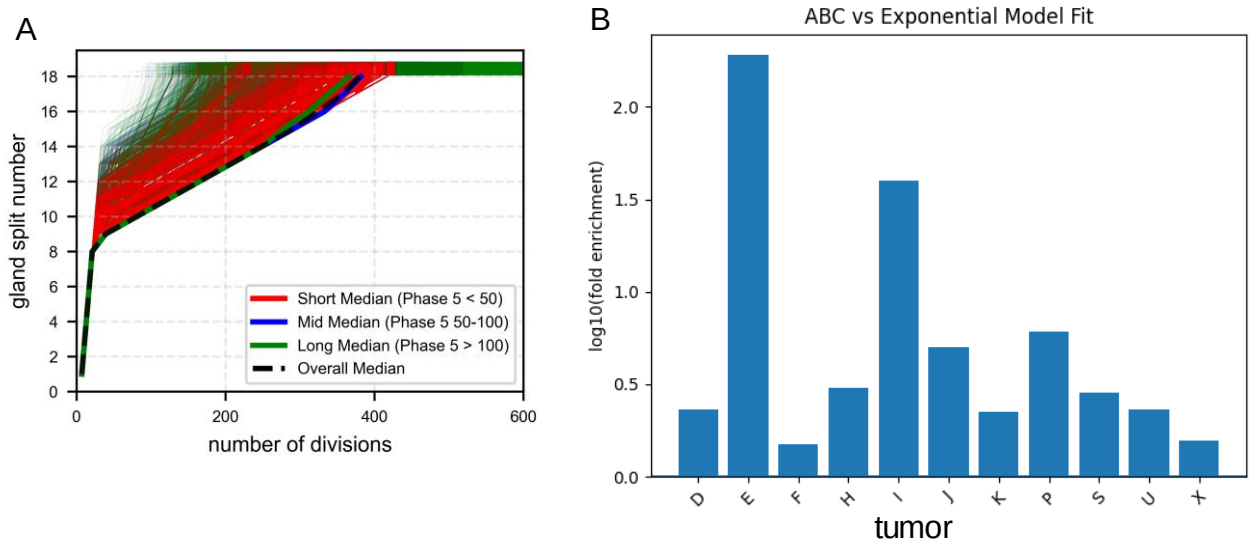


Figure S2. Early exponential growth is a poor fit to tumor methylation data.
A: Growth trajectories generated under a death-free exponential growth constraint requiring the first seven gland doublings to occur without cell loss. Individual ABC-accepted simulations are shown in gray. Simulations are grouped according to the duration of the terminal stationary phase (Phase 5), with median trajectories shown for short (<50 divisions, red), intermediate (50–100 divisions, blue), and long (>100 divisions, green) stationary intervals. The overall median trajectory is shown in black.
B: Relative fit of constrained death-free exponential growth simulations compared with unconstrained simulations for each tumor. Bars show the log10 enrichment of accepted simulations from the unconstrained model after normalization for the number of simulations performed. Positive values indicate that unconstrained growth trajectories were accepted more frequently than trajectories constrained to begin with death-free exponential expansion. Across tumors, early death-free exponential growth consistently showed lower acceptance frequencies, indicating that growth constraints and cell loss are likely present while tumors are still small.

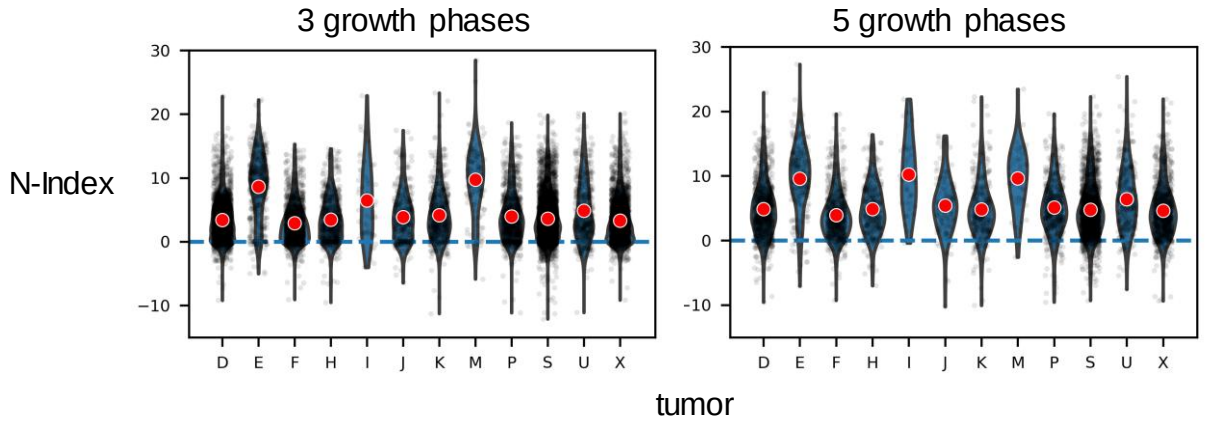


Figure S3. N-shaped growth dynamics are robust to model complexity.
Violin plots show inferred phase-specific growth rates from ABC-accepted simulations using three-phase (left) and five-phase (right) growth models. Red circles indicate median values. Both model structures show evidence of slow early growth, followed by further slowing and subsequent re-acceleration, indicating that the inferred N-shaped pattern is not dependent on the number of growth phases included in the model.

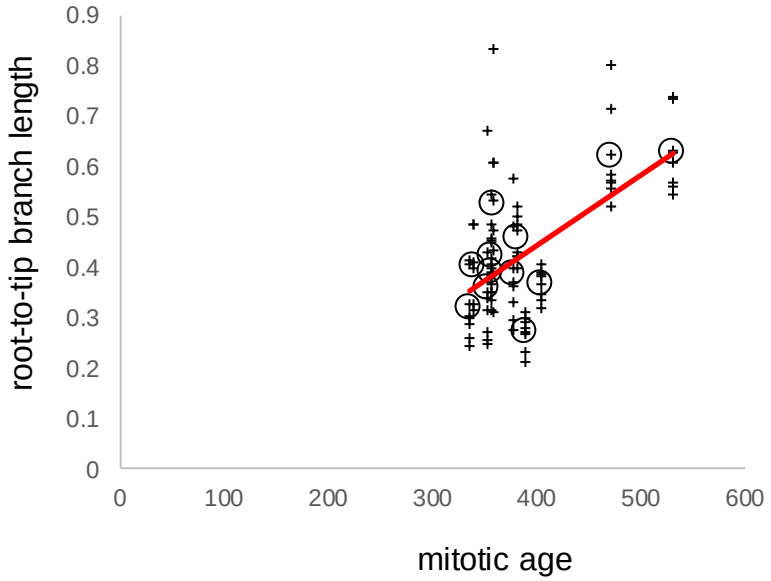


Figure S4. Mean root-to-tip length of the experimental gland phylogenies correlated with mean mitotic age inferred from ABC-accepted simulations (Pearson R = 0.725, P = 0.0077, n = 12 tumors). Crosses represent individual glands, and large circles represent tumor means. The regression line was fitted to the tumor means.

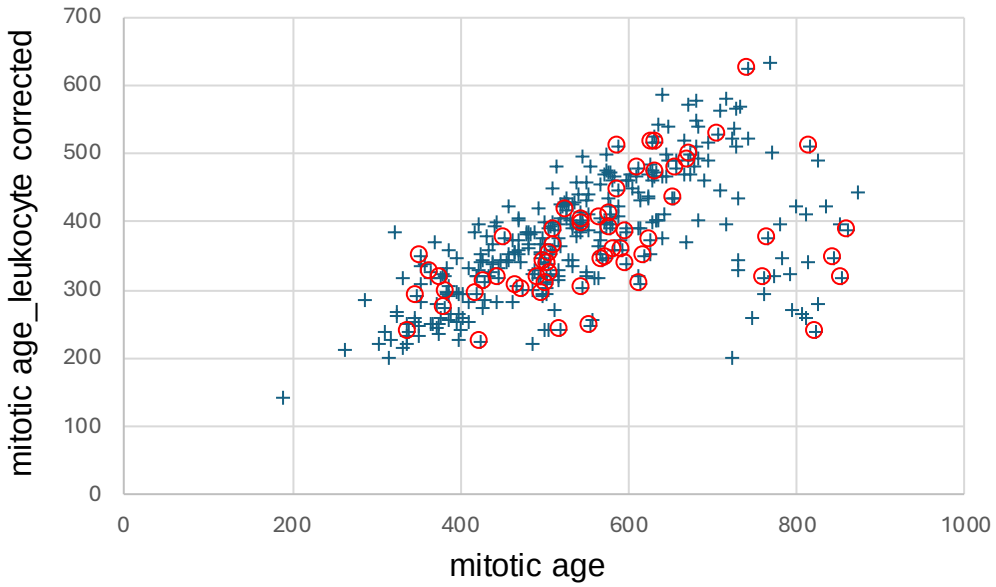


Figure S5. Leukocyte correction preserves inferred CRC mitotic ages.
Scatterplot comparing mitotic age estimates before and after correction for leukocyte infiltration. Immune-cell fractions estimated by CIBERSORT and reported by Thorsson et al. (Ref. 27) were used to adjust methylation-derived mitotic age estimates. Blue crosses represent individual CRCs and red circles indicate MSI CRCs. Mitotic age estimates remained significantly correlated following correction (R = 0.58), indicating that associations between mitotic age and tumor microenvironment features are not driven solely by the extent of leukocyte infiltration.

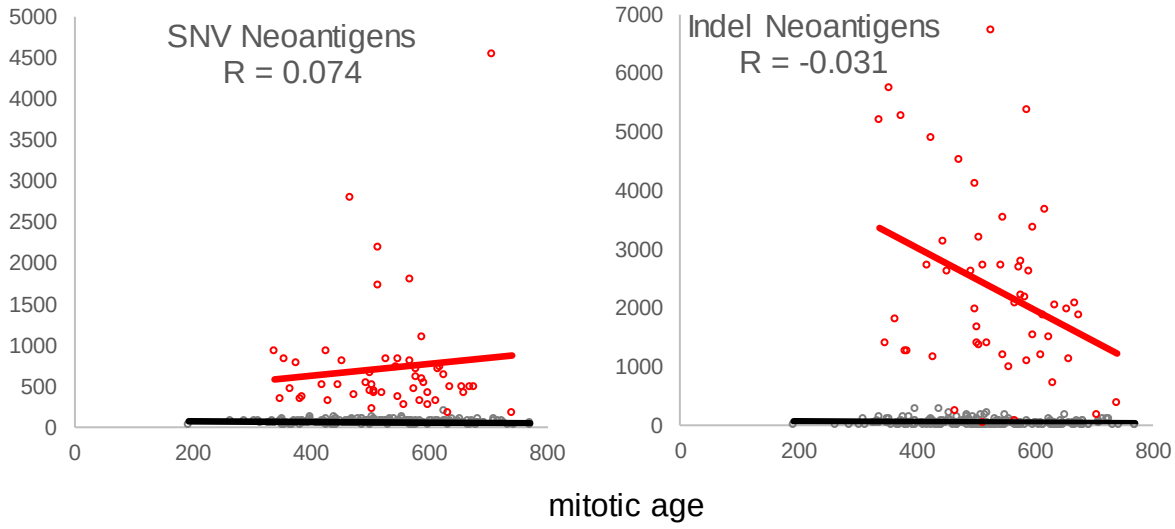


Figure S6. Neoantigen burden does not increase with inferred CRC mitotic age.
Scatterplots showing the relationship between inferred mitotic age and predicted neoantigen burden for single-nucleotide variant (SNV)-derived neoantigens (left) and insertion/deletion (indel)-derived neoantigens (right). Predicted neoantigen data were obtained from Reference 27. MSI CRCs are shown in red and MSS CRCs in gray. Correlations between mitotic age and neoantigen burden were weak for both SNV neoantigens (R = 0.074) and indel neoantigens (R = −0.031), indicating that older CRCs do not exhibit substantially higher predicted neoantigen loads. These findings indicate that the increased immune infiltration and immune exhaustion observed in older CRCs are not accompanied by correspondingly higher predicted neoantigen burdens.

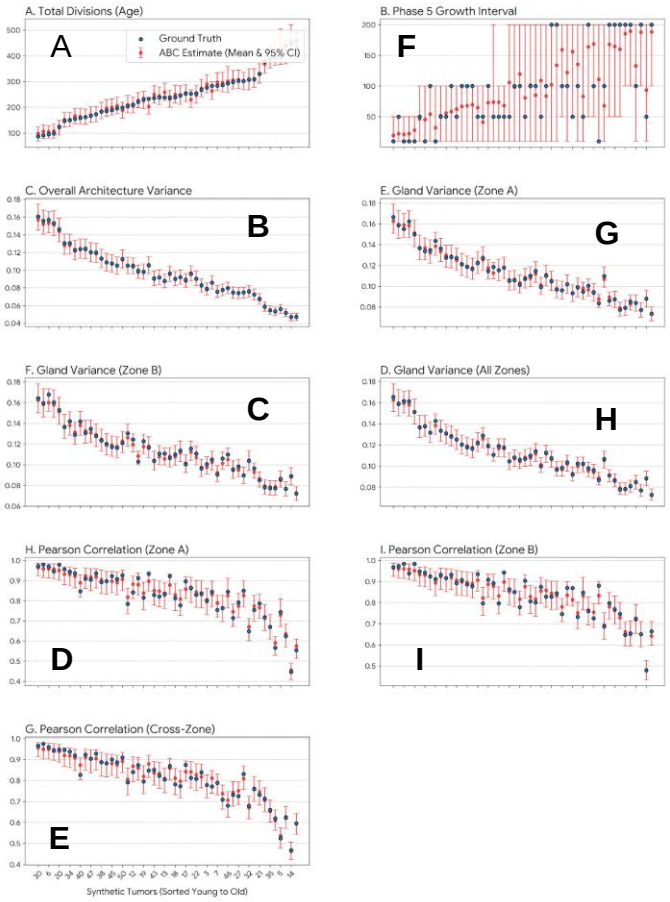


Figure S7. Validation of ABC parameter recovery using synthetic tumors.
Synthetic tumors with known underlying growth parameters were generated and analyzed using the same ABC framework applied to experimental CRCs. Blue circles indicate ground-truth values used to generate the synthetic tumors, and red triangles indicate ABC-derived parameter estimates (mean and 95% simulation interval) obtained from accepted simulations. Synthetic tumors are ordered by increasing total mitotic age (young to old).
(A) Recovery of total divisions (mitotic age).
(B) Recovery of overall architectural variance.
(C) Recovery of gland variance within Side B.
(D) Recovery of Pearson correlations among glands within Side A.
(E) Recovery of Pearson correlations between glands from opposite tumor sides (cross-side).
(F) Recovery of Phase 5 growth intervals.
(G) Recovery of gland variance within Side A.
(H) Recovery of gland variance across all spatial regions.
(I) Recovery of Pearson correlations among glands within Side B.
Overall architectural variance, gland variance, and Pearson correlation metrics were recovered with high accuracy across the full range of synthetic tumors. Recovery of total mitotic age was also robust. In contrast, estimates of Phase 5 growth intervals exhibited substantially greater uncertainty, particularly in older synthetic tumors, indicating that the ABC framework more reliably recovers overall growth architecture and correlation structure than precise latent phase boundaries.

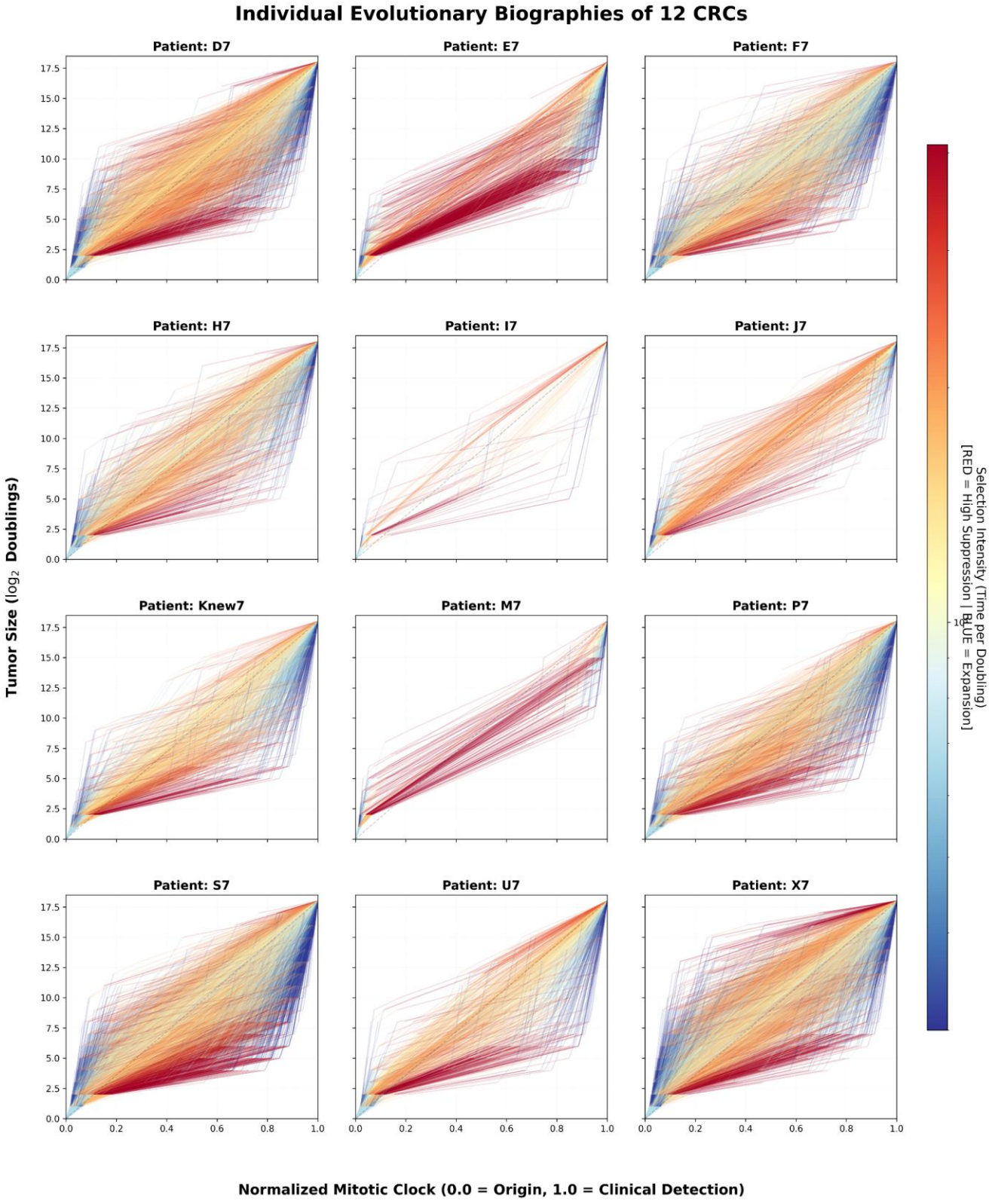

Individual Evolutionary Biographies of 12 CRCs
Patient: D7
Patient: E7
Patient: F7
Patient: H7
Patient: I7
Patient: J7
Patient: Knew7
Patient: M7
Patient: P7
Patient: S7
Patient: U7
Patient: X7
Tumor Size (log2 Doublings)
Normalized Mitotic Clock (0.0 = Origin, 1.0 = Clinical Detection)
Selection Intensity (Time per Doubling)
[RED = High Suppression | BLUE = Expansion]

Figure S8. Individual ABC-accepted growth trajectories for all 12 tumors.
Panels show all ABC-accepted growth trajectories for each tumor after normalization of mitotic age. The x-axis represents normalized mitotic time from tumor origin (0) to clinical detection (1), and the y-axis represents tumor size in cumulative log2 gland doublings. Each line corresponds to an ABC-accepted simulation consistent with the observed gland methylation and mutation data for that tumor. Colors indicate inferred cell loss fraction during growth, with blue representing higher cell loss (greater growth suppression) and red representing lower cell loss (greater net expansion). Most tumors exhibit an initial period of growth constraint, followed by stronger constraint and then relaxation, producing characteristic N-shaped growth trajectories. These individual tumor histories illustrate the heterogeneity of growth constraints underlying the aggregate patterns shown in Fig. 3.

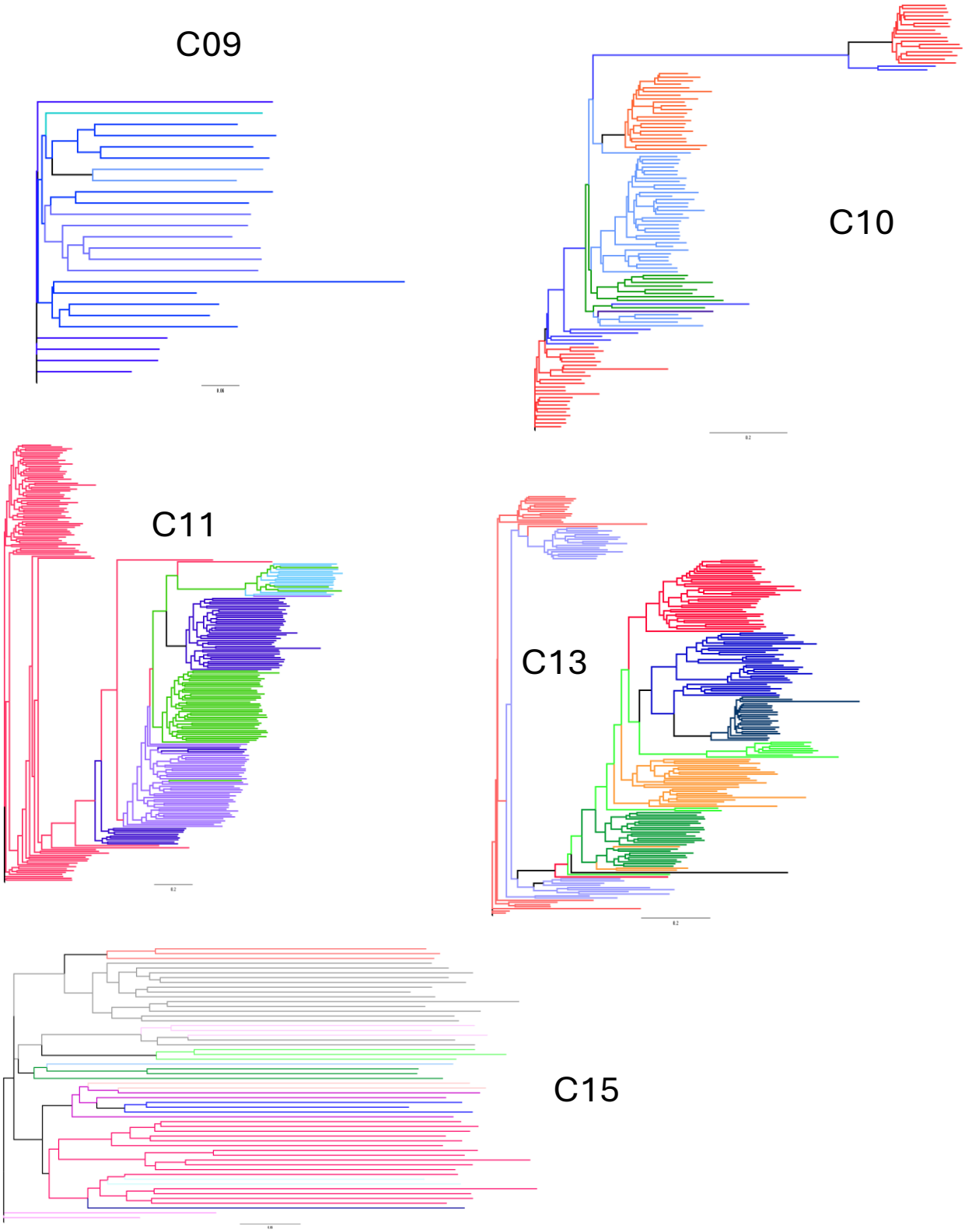


Figure S9. Additional single-cell trees.
Trees for the five other CRCs. Terminal branch colors indicate spatial locations of each sampled cell. Blue, green or violet shades represent sampling from the primary tumor, red shades represent sampling from lymph node or liver metastases, and grey represents sampling from omental metastases. Spatial samplings of the omental or liver metastases were combined for clarity for CRC15.

Table S1) CRC Gland Samples

Sample	Type	Stage	Size (cm)	Analyzed In Ref 23
D	Cancer	1	2	Yes
E	Cancer	1	6	Yes
F	Cancer	1	2	Yes
H	Cancer	4	4	Yes
I	Cancer	4	8	No
J	Cancer	3	5	Yes
K	Cancer	1	3.5	Yes
M	Cancer	2	3	Yes
P	Adenoma		3.5	Yes
S	Adenoma		6	Yes
U	Cancer	2	4	Yes
X	Adenoma		2.5	Yes

Table S2) RNA expression with respect to mitotic age

	ALL CRCs		MSI		MSS	
Gene	Pearson	p	Pearson	p	Pearson	p
PDCD1	0.247	8.57E-06	0.278	0.036	0.256	8.60E-05
CTLA4	0.372	7.95E-12	0.358	0.006	0.407	1.45E-10
LAG3	0.064	0.25	-0.131	0.33	0.243	2.02E-04
TIGIT	0.362	3.15E-11	0.477	0	0.333	2.34E-07
HAVCR2	0.358	5.81E-11	0.374	0.004	0.38	2.65E-09
GNLY	0.101	0.072	0	0.998	0.207	1.59E-03
GZMA	0.181	0.001	0.104	0.44	0.238	2.76E-04
CXCL13	0.294	9.95E-08	0.447	4.96E-04	0.252	1.11E-04
B2M	0.266	1.68E-06	0.254	0.056	0.263	5.26E-05
TNFRSF9	0.407	4.71E-14	0.389	0.003	0.449	8.05E-13
TNFRSF18	0.27	1.10E-06	0.195	0.145	0.367	9.61E-09
CD68	0.277	5.48E-07	0.257	0.053	0.363	1.47E-08
CD163	0.23	3.72E-05	0.251	0.06	0.276	2.14E-05
MRC1	0.27	1.10E-06	0.231	0.084	0.311	1.46E-06
MSR1	0.302	4.15E-08	0.357	0.006	0.297	4.70E-06
	N=316 CRCs		MSI=57		MSS=230	

Table S3) Match rates for neighboring cells in IQTREE phylogenies for the same sampled spatial region.

Tumor	Type	Cells Per Region	Same Tree Region	Match Rate (%)
C04	Primary	44	43	97.7%
	Metastases	35	27	77.1%
C09	Primary	24	22	91.7%
	Metastases	0	—	—
C10	Primary	55	52	94.5%
	Metastases	62	62	100.0%
C11	Primary	147	136	92.5%
	Metastases	40	40	100.0%
C12	Primary	14	11	78.6%
	Metastases	4	4	100.0%
C13	Primary	113	113	100.0%
	Metastases	83	81	97.6%
C15	Primary	12	11	91.7%
	Metastases	43	27	62.8%
Overall	Primary	409	388	94.9%
	Metastases	267	241	90.3%