Evolution under Spatially Heterogeneous Selection in Solid Tumors

Guanghao Li [iD],[†,1,2,3,4,5] Zuyu Yang,[†,2,3,6] Dafei Wu,[2,3,7] Sixue Liu,[2,3,5] Xuening Li,[2,3,5] Tao Li,[2,3,5] Yawei Li,[2,3,5] Liji Liang,[2,3] Weilong Zou,[8,9] Chung-I Wu,[2,3,10] Hurng-Yi Wang,*[,11,12] and Xuemei Lu*[,1,2,3,4,5]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
[2]China National Center for Bioinformation, Beijing, China
[3]CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
[4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China
[5]University of Chinese Academy of Sciences, Beijing, China
[6]Institute of Environmental Science and Research, Porirua, New Zealand
[7]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China
[8]Surgery of Liver Transplant, The Third Medical Center of Chinese PLA General Hospital, Beijing, China
[9]Surgery of Hepatopancreatobiliary, Peking University Shougang Hospital, Beijing, China
[10]State Key Laboratory of Biocontrol, College of Ecology and Evolution, Sun Yat-Sen University, Guangzhou, China
[11]Graduate Institute of Clinical Medicine, National Taiwan University, Taipei, Taiwan
[12]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan
[†]These authors contributed equally to this work.
*Corresponding authors: E-mails: hurngyi@g.ntu.edu.tw; xuemeilu@mail.kiz.ac.cn.
Associate editor: Bing Su

Abstract

Spatial genetic and phenotypic diversity within solid tumors has been well documented. Nevertheless, how this heterogeneity affects temporal dynamics of tumorigenesis has not been rigorously examined because solid tumors do not evolve as the standard population genetic model due to the spatial constraint. We therefore, propose a neutral spatial (NS) model whereby the mutation accumulation increases toward the periphery; the genealogical relationship is spatially determined and the selection efficacy is blunted (due to kin competition). In this model, neutral mutations are accrued and spatially distributed in manners different from those of advantageous mutations. Importantly, the distinctions could be blurred in the conventional model. To test the NS model, we performed a three-dimensional multiple microsampling of two hepatocellular carcinomas. Whole-genome sequencing (WGS) revealed a 2-fold increase in mutations going from the center to the periphery. The operation of natural selection can then be tested by examining the spatially determined clonal relationships and the clonal sizes. Due to limited migration, only the expansion of highly advantageous clones can sweep through a large part of the tumor to reveal the selective advantages. Hence, even multiregional sampling can only reveal a fraction of fitness differences in solid tumors. Our results suggest that the NS patterns are crucial for testing the influence of natural selection during tumorigenesis, especially for small solid tumors.

Key words: cancer evolution, tumor spatial growth model, intra-tumoral heterogeneity, natural selection, phenotypic diversity.

Introduction

Tumorigenesis starts from a single cell that grows into a population of more than 10^9 cells through a process of somatic evolution and spatial range expansion (Sottoriva et al. 2015; Wu et al. 2016; Sun et al. 2017). As solid tumors usually form a compact structure that obstructs tumor cells from free migration, it is reasonable to expect that intratumor heterogeneity may result from independent progression of spatially distributed segments. Indeed, spatial phenotypic heterogeneity within tumors has been well documented in genomic analyses and pathological characterization of various types of cancers (Ling, et al. 2015; Sottoriva et al. 2015; Waclaw et al. 2015; Uchi et al. 2016; Yuan 2016; Carmona-Fontaine et al. 2017; Li et al. 2017; Chan-Seng-Yue et al. 2020; Gerstung et al. 2020; Losic et al. 2020; Dentro et al. 2021). For instance, Ki-67 staining suggests that proliferating cells at the tumor

edge outnumber cells at the core (Waclaw et al. 2015; Lloyd et al. 2016; Zhao et al. 2021). In addition, cells in the central hypoxic regions are thought to be the most aggressive and therapy-resistant (Terry et al. 2017), whereas cells at the tumor margin may be replenished frequently by the continually fluctuating microenvironment, leading to more division and death (Li et al. 2020). This evidence indicates that evolutionary rates and population dynamics may vary across the range of structured tumor cell populations.

Although understanding the process of population differentiation and adaptation driven by evolutionary forces, including genetic drift, migration, and natural selection, is the central theme in evolutionary biology (Yoshida et al. 2003; Schoener 2011; Hendry 2016), most evolutionary theories attempting to explain population dynamics are based on a simplified assumption that populations are at spatial panmixia. Accordingly, many studies assume homogeneous distribution of genetic variation within tumors (Williams et al. 2016; Chowell et al. 2018; Williams et al. 2018). Since solid tumors do not evolve as the standard population genetic model due to the spatial constraint, a new "null" model for the study of intratumor genetic heterogeneity is thus necessary.

Here, we build a neutral spatial (NS) model of tumor growth. We assume tumor growth in a discrete-time birth-death process initiated from a single cell. Under neutral conditions, the tumor expanded at a constant rate as the growth rate of each subclone was similar. As a result, the shapes and relative sizes of all subclones were stable over time. Because of spatial constraint, peripheral regions exhibited higher proliferation rate than central regions which led to a higher generation numbers, the former accumulated more mutations and showed longer branch lengths than the latter. To test the NS model, we performed a three-dimensional microdissection of two tumors obtained from a single hepatocellular carcinoma (HCC) patient. The results of whole-genome sequencing (WGS) and genotyping revealed a roughly 2-fold increase in mutations going from the center to the periphery.

We also tested the operation of natural selection by examining the spatially determined clonal relationships and the clonal sizes. When the tumor grows under Darwinian evolution, random accumulation of driver mutations in different subclones resulted in changes in their shapes and proportions. Dominant clones with the highest fitness increased its proportions in the tumor rapidly, whereas other small clones that occurred early went extinct as the expansion of dominant clones. In addition, due to spatial constraints, only the expansion of highly advantageous clones can sweep through a large part of the tumor to reveal the selective advantages. Hence, even multiregional sampling can only reveal a fraction of fitness differences in solid tumors. This modeling yielded insights into the neutral and environmental constraints that drive the formation and progression of intratumor genetic structure, as well as the tempo and mode of tumor diversification and clonal expansion.

Results

Tumor Growth under Strict Spatial Constraint

We first evaluate a neutral tumor expansion model with strict spatial constraints. Suppose that the shape of tumor is an ideal sphere consisting of layers of cells. The most recent common ancestor (TMRCA) of the tumor is at layer 0. Layer n is the ring of cells with n cells distant from layer 0. Cells in layer n is generated from those in layer $n-1$. The number of cells in layer n is $4\pi n^2$ (supplementary fig. 1A, Supplementary Material online). Assume the birth and death rates of each tumor cell are b and $1-b$, respectively. The average number of divisions for cells in layer n is

$$D_n = log_{2b}\left[(4\pi+1)\prod_{i=1}^{n-1}\left(1+\left(\frac{i+1}{i}\right)^2\right)\right], \ n \geq 2$$

D_n is approximate linear related with n (supplementary fig. 1B, Supplementary Material online).

For cells in layer n and kn, $\frac{D_{kn}}{D_n}$ approximates to k as n increase (supplementary fig. 1C, Supplementary Material online). Under the assumption that mutation rate is constant, mutations accumulated in cells of layer kn are k folds than those of cells in layer n. Therefore, spatial heterogeneity of solid tumors can be simply due to limited migration within the tumor mass. Under a simple spatial growth model, mutations accumulate more rapidly in the periphery than in the interior.

Modeling Tumor Growth with NS Structure

To investigate how spatial heterogeneity is generated under NS model in solid tumors in more detail, we propose a spatial tumor growth model (Lloyd et al. 2016). For cells in a peripheral region, the birth rate was b_0 and death rate was $1-b_0$, whereas cells in the central region stay quiescent when resources are limited, and the birth, quiescence, and death rates are b_1, q, and $1-b_1-q$, respectively (fig. 1A and supplementary table 1, Supplementary Material online). We demonstrated modeled tumor growth under neutral evolution, where mutations have no effect on cell growth rate. We simulated tumor growth in a discrete-time birth-death process initiated from a single cell to more than 10^6 cells (corresponding to 10^9 in three dimensions).

Under neutral evolution, higher birth and death rates for cells in a peripheral region than in a central region result in more generations, which further cause more mutations accumulated at the surface. To explore mutation accumulation under different birth and death rates, we performed 5,000 simulations with q sampled from the [0.1, 0.9] interval and b_0 from [0.51, 0.6]. The range of b_1 depended on the value of q and b_0, as b_1 must be bigger than central cell death rate $(1-q-b_1)$ and smaller than b_0. Therefore, the value of b_1 was sampled from $[(1-q)/2 + 0.02, \min(b_0, 1-q)-0.02]$. To define the location of a cell, we first estimated the radius of the circle-like tumor, r. For each cell, we calculated the distance between its position with the center of the tumor, d. We also set up a value m to determine the central and
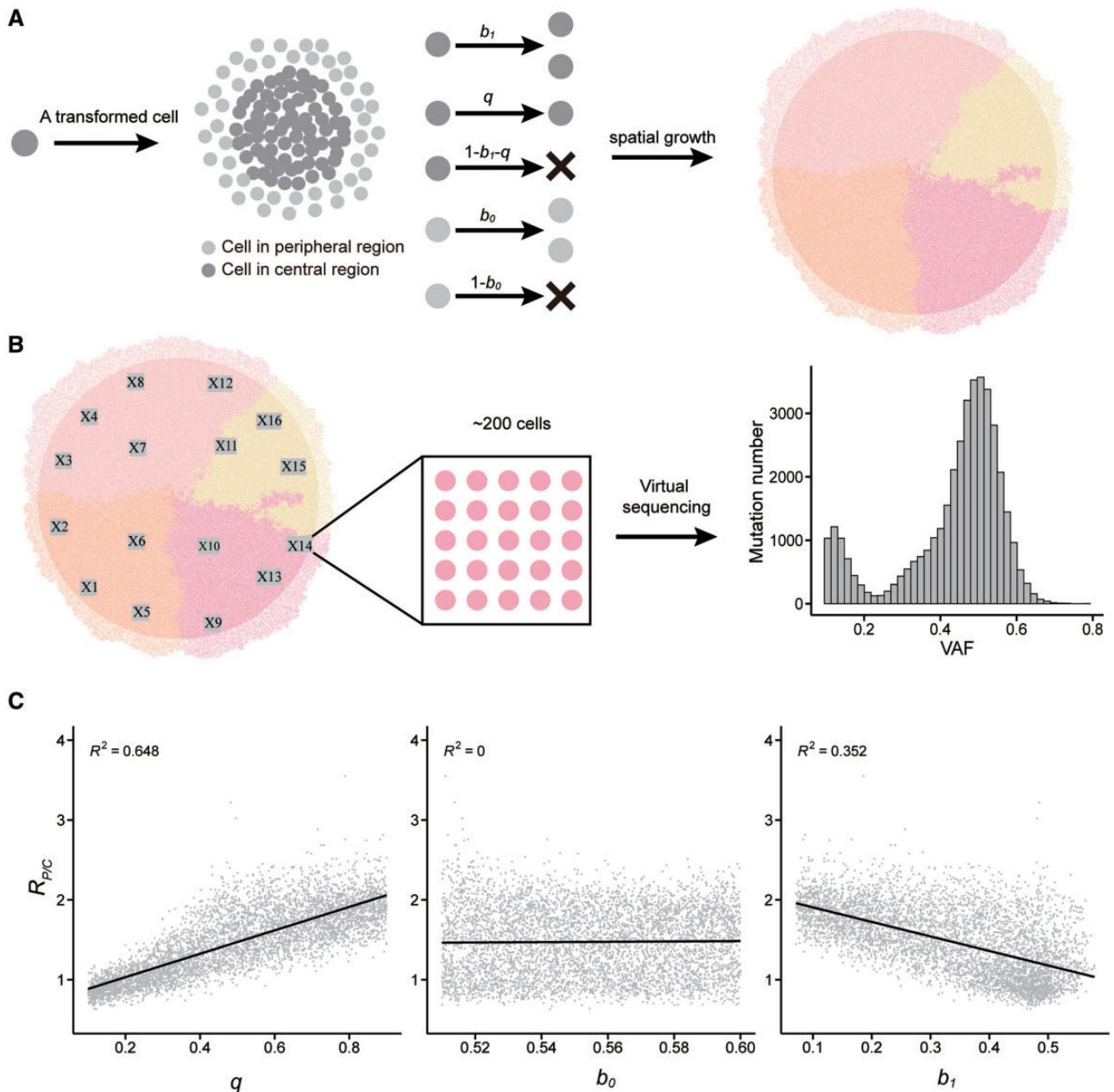
FIG. 1. Spatial tumor growth model. (*A*) Schematic of the spatial tumor growth model. Cells at the periphery and the center have different birth and death rates. (*B*) Schematic of the sampling and virtual sequencing strategy. When the tumor grows to 5×10^5, 1×10^6, 1.5×10^6, or 2×10^6 cells, 16 samples are taken uniformly to comprehensively investigate the genetic diversity of the simulated tumor. Each sample is sequenced virtually and all samples are combined to construct mutation frequency spectra and phylogenetic relationships within a whole tumor. Sample size is about 200 cells. (*C*) Relationships between average mutation number in samples from the periphery or the center ($R_{P/C}$) ($n = 5{,}000$ tumors; R square is reported) and parameters used in the model. q, b_0, and b_1 are quiescent rate, birth rate of cells at the periphery, and birth rate of cells at the center.

peripheral regions. In this study, we used $m = 0.9$. That is if $d < 0.9^*r$, the cells are in the central region, otherwise, it is in the periphery. It is worth of mentioning that different m values should only affect the proportion of peripheral region, but not mutation accumulation and clonal structure of the tumors.

Mutations accumulate randomly according to the Poisson distribution during cell proliferation, assuming infinite site model and a constant mutation rate. The neutral mutation rate is μ and the probability that i neutral mutations accumulate after each cell division is:

$$P(x = i) = \frac{\mu^i e^{-\mu}}{i!}.$$

The neutral mutation rate $\mu = 15$ per genome (corresponding to 5×10^{-9} per site per cell division for the human genome) (Sun et al. 2017). For each simulation, 16 samples were taken from the virtual tumor uniformly (fig. 1B), sequenced in silico, and single-nucleotide variants (SNVs) were called and analyzed. The ratio of average mutation number in samples from peripheral regions to the count from central regions ($R_{P/C}$) was calculated.

The ratio of average mutation number in peripheral samples to that from the center, $R_{P/C}$, was about 1.5 on average, and the 95% confidence interval was [0.79, 2.25] (supplementary fig. 2, Supplementary Material online). $R_{P/C}$ is positively correlated with q (fig. 1C), demonstrating that if more cells are quiescent in central regions, fewer mutations are generated. In contrast, if central cells divide more frequently, the difference between the central and peripheral regions is reduced. Thus, we observe a negative correlation between b_1 and $R_{P/C}$. The range of values for $R_{P/C}$ is relatively stable, regardless of the peripheral birth rate b_0.

We also explored the genetic diversity and clonal structure of the simulated tumors. In this setting, a moderate initial birth rate $b_0 = 0.53$ was used (Sun et al. 2017). The q and b_1 were set to 0.5 and 0.3, respectively, based on previous simulations (supplementary table 1, Supplementary Material online and see Materials and Methods). Sixteen virtual samples were taken (fig. 2A) when tumor size reached 5×10^5, 1×10^6, 1.5×10^6, and 2×10^6. SNVs were used to delineate clones and estimate phylogenetic relationships among subclones. Under neutral conditions, the tumor expanded at a constant rate as the growth rate of each subclone was similar. As a result, the shapes and relative sizes of all subclones were stable over time. Cancer cell fractions (CCFs) of all SNVs were low and always exponentially distributed (fig. 2B). Since peripheral regions exhibited higher birth and death rates than central regions, leading to higher generation numbers, peripheral cells accumulated significantly more mutations and had longer branches than central samples (fig. 2C). The $R_{P/C}$ values are 1.39, 1.23, 1.30, and 1.40 for the four panels in figure 2C.

Three-Dimensional Microsampling and Sequencing

To comprehensively trace the mutation accumulation process of solid tumors and describe their clonal distribution, we performed three-dimensional microdissection (fig. 3A and B) and WGS of two tumors, T1 and T2, obtained from an HCC patient who was a chronic hepatitis B virus (HBV) carrier and had undergone liver transplantation (see Materials and Methods). The three-dimensional microsampling of the 11 slices of T1 yielded 169 samples. We obtained 160 samples from six T2 slices. Samples were taken uniformly from each slice, with central and peripheral regions represented (fig. 4). Sixteen samples from three slices of T1 and nine samples from two slices of T2, each containing ~3,200 cells, were subjected to WGS. The average depth was ~78× per sample, yielding 1,264× coverage of T1 and 691× of T2. A peripheral blood sample was collected as the normal control and sequenced to ~74× coverage with estimated purity ranging from 0.70 to 0.94 (supplementary table 2, Supplementary Material online).

We identified 254,268 SNVs in T1 and 142,032 SNVs in T2. Of these, only 6,869 (2.7%) mutations from T1 and 3,721 (2.6%) from T2 were shared by all samples from the same tumor. No SNVs were found in common between T1 and T2, indicating that they originated independently. The number of SNVs varied greatly across samples from the same tumor, from 21,345 in T1L13 to 68,149 in T1F24, and from 24,370 in T2Z11 to 70,093 in T2Z1 (supplementary tables 2 and 3, Supplementary Material online). Mutations in CIC, KMT2B,

and DYRK1A were found in all T1 samples and SCAF4 and HNF1A in all T2 samples (Zhai et al. 2017; Bailey et al. 2018), whereas the PKHD1L1 mutation was detected as a subclonal event both in T1 and T2 (supplementary fig. 3 and supplementary materials, Supplementary Material online).

Diverse copy number alternations (CNA) profiles among samples within the same tumor, as well as between T1 and T2, were detected (supplementary fig. 4, Supplementary Material online). Specifically, we found significant arm-level gains of almost all chromosomes in both T1 and T2. Losses of 6q were observed in T1 and of 4p and 13q in T2. We also observed CNAs of canonical HCC driver genes at focal levels. For example, amplification of CCNE1 (Li et al. 2018) and TERT (Li et al. 2018, 2020) were found in T1 and T2, respectively, and deletion of RB1 (Wang et al. 2013) was found in T2 (supplementary fig. 4, Supplementary Material online). Genome-wide CNAs were quite different at both focal and arm levels in both T1 and T2, indicating varying evolutionary processes in tumors even in the same patient.

HBV integration into the human genome may play a crucial role in triggering oncogenic processes (Bréchot et al. 2000). We found that the HBV integrations in an intron of ATP2B2 and the intergenic region between LOC100506444 and RPL21P44 were common in T1. An HBV integration in the 5′ UTR of TERT (Sung et al. 2012; Totoki et al. 2014) was found in all T2 samples (supplementary table 4, Supplementary Material online).

We also explore mutation profiles of 96 trinucleotide changes. Five mutation signatures were identified (supplementary fig. 5A–C, Supplementary Material online) and compared with all signatures described in the Catalog of Somatic Mutations in Cancer (COSMIC) (Tate et al. 2019). The first, second, and fourth signatures were enriched for T: A > A: T mutations, which have been found in urothelial (renal pelvis) carcinomas and liver cancers with known exposure to aristolochic acid (Alexandrov et al. 2020) (COSMIC signature 22, $\rho = 0.974$, 0.972, and 0.731, respectively).

Mutation Accumulation and Subclone Spatial Structure

Phylogenetic trees were constructed based on SNVs from each sample using peripheral blood as the outgroup (fig. 3C and D). In the T1 phylogeny, T1Z5 branched out first (defined as clone γ), followed by five samples sharing 721 SNVs (clone β), and ten samples sharing 16,412 SNVs (clone α). In T2, six samples sharing 17,493 SNVs clustered together (clone δ). T2Z1 and T2Z13 shared 5,141 SNVs, constituting clone θ. Phylogenetic trees show that branch lengths vary greatly with the long-branched subclones tending to occur in peripheral regions (fig. 4C–F). The $R_{P/C}$ for T1 is 2.19 and T2 2.07, within the range of simulated results. Collectively, T1 and T2 underwent independent evolutionary paths but had similar spatial patterns, that is peripheral regions accumulated more mutations than central locations, both containing bigger subclones.

To further delineate the size and spatial distribution of subclones and investigate the relationship between branch length and subclone size, we conducted additional
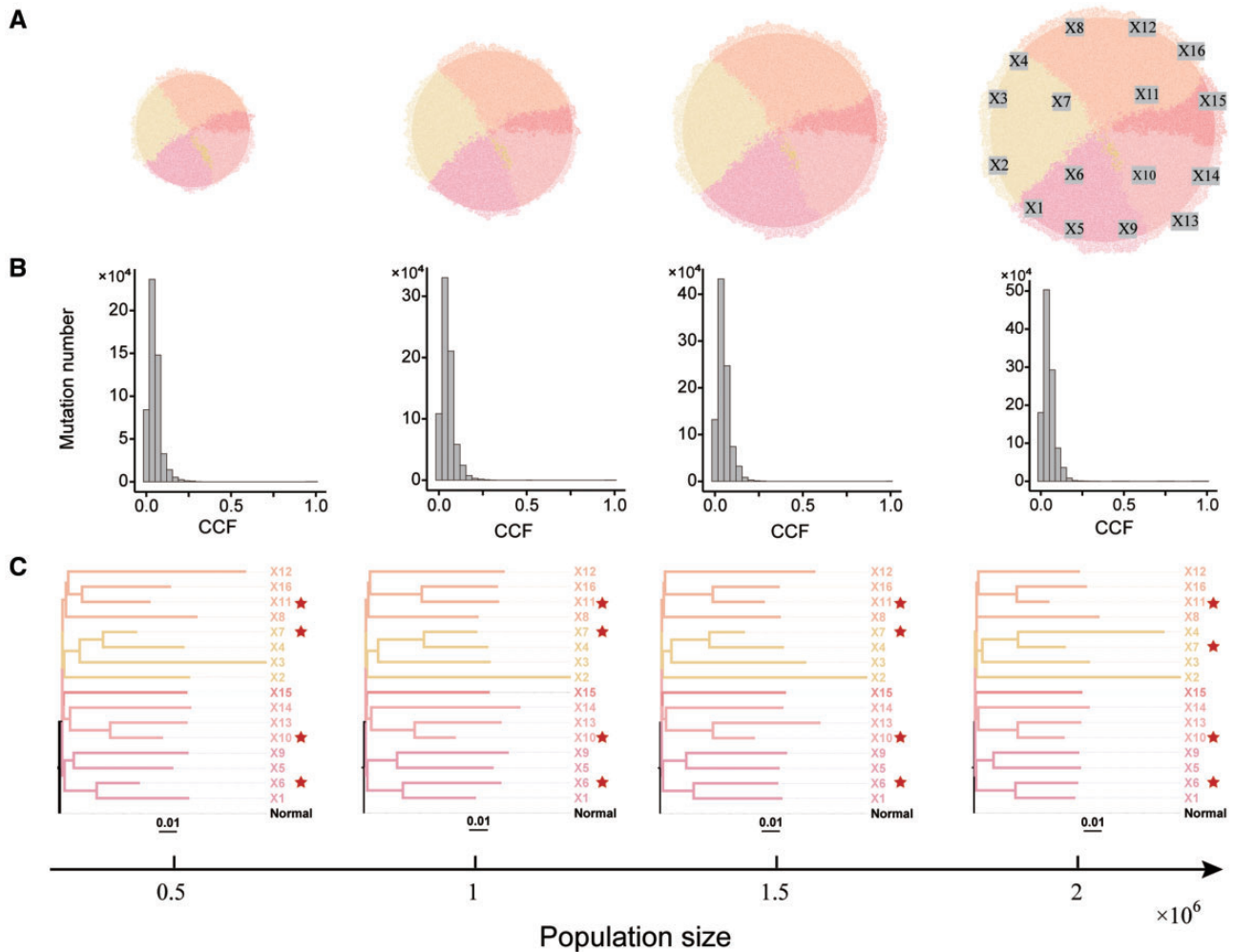
FIG. 2. Clonal structures and phylogenies of tumors under neutral evolution over time. (A) Clonal distributions of the same virtual tumor when cell population size is 5×10^5, 1×10^6, 1.5×10^6, and 2×10^6. Colors represent subclones. (B) Sixteen samples, each with about 200 cells, are taken uniformly from each tumor and sequenced virtually at each time point. Position of each sample is in the panel 4 of (A). Histograms are the distributions of CCFs for all SNVs detected. (C) Sample phylogenetic trees. Normal sample without any SNVs is used as the outgroup. Samples marked with red stars are from central regions. Internal branch lengths are significantly shorter than peripheral branches ($P < 0.001$, $= 0.043$, <0.001, and <0.001 for population size 5×10^5, 1×10^6, 1.5×10^6, and 2×10^6, respectively). The $R_{P/C}$ values are 1.39, 1.23, 1.30, and 1.40. Unit of branch length in (C) is evolutionary distance in maximum parsimony.

genotyping of selected 906 and 565 SNVs in 153 T1 and 151 T2 samples, using the Ampliseq method (supplementary tables 5 and 6, Supplementary Material online). The genotypes were used to construct expanded phylogenies including all the sequenced and validated samples to delineate clonal distribution in each slice. Consistent with the phylogenies based on WGS data, the three clones α, β, and γ from T1 and the two clones θ and δ from T2 were identified in the expanded phylogenies (fig. 4A and B). According to the phylogenies, both α and δ are of the same age as the other clones, implying that their larger sizes resulted from a growth advantage over other cells. Clonal structure of slices shows that subclones (β in T1 and θ in T2) containing long-branch samples occur in peripheral regions of the tumors (fig. 4C and E).

Most subclones defined using selected SNVs were highly consistent with those characterized by whole-genome variants. For example, 78 SNVs were chosen from the shared

SNVs of subclone α from WGS, an average 95.7% of which were identified in all validation α samples. The values for subclone β, γ, δ, and θ were 96.3%, 97.2%, 99.1%, and 59.4%, respectively (supplementary table 7, Supplementary Material online). Consistency for subclone θ is relatively low. The possible reason might be that subclone θ was located at the surface of the tumor with higher genetic diversity.

Using clone distributions across slices, we inferred changes in subclone composition in three dimensions for both tumors. Subclone α, accounting for 42.6% of T1, was further divided into α1 and α2. The subclone δ (84.4% of T2) was further divided into δ1, δ2, and δ3 (fig. 4D and F). Clone α was the biggest in slice T1F, T1H, and T1K. α, β, and γ were similar in size in slices T1O, T1Q, and T1S. β was the dominant subclone in slice T1V and γ was similar to β in size in slices T1Y and T1AB. Different subclonal distributions across slices suggests a significant spatial
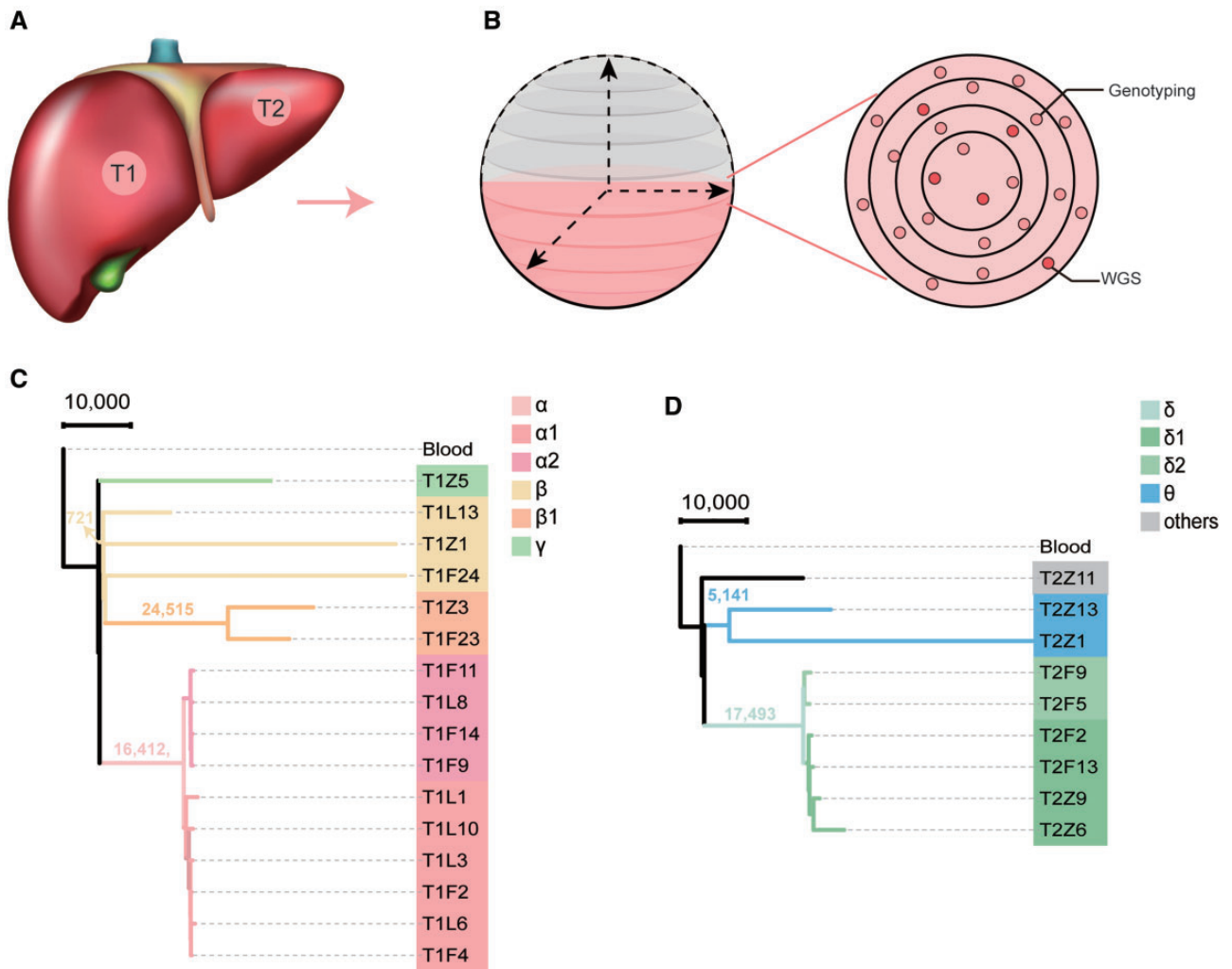
FIG. 3. Sampling strategy and phylogenetic relationships. (A) Tumor locations in the liver. (B) Sketch map of the three-dimensional sampling and sequencing strategy. Half of a tumor was cut into dozens of slices. We took samples from several slices to perform WGS (dark red) and genotyping (light red). (C) Phylogenetic tree for T1 based on WGS data. Branch colors (red, yellow, and green) represent different clones in the same tumor. (D) Phylogenetic tree for T2 based on WGS data. Branch colors (green and blue) represent different clones in the same tumor. Unit of branch length in (C) and (D) is mutation number. Label 10,000 represents 10,000 mutations.

heterogeneity in T1 (fig. 4C). Clone α size decreases from the center to the surface, whereas the size of clone β and γ increases (fig. 4C). Subclones from different slices showed a similar pattern of both boundaries and sizes in T2. δ was the dominant subclone for all slices, and subclone θ was much smaller than δ (fig. 4E), indicating spatial diversity of T2 was much lower than in T1. In both T1 and T2, some samples located at the tumor peripheral regions did not belong to any subclones detected by WGS, implying higher genetic diversity on the surface of the tumor.

We noticed that peripheral regions not only accumulated more mutations, but also contained more changes in genes related to cell proliferation and cell cycle function (supplementary table 8, Supplementary Material online). For example, only three and six mutations in these genes were identified in the central region samples from T1 and T2. In contrast, 56 and 30 SNVs were found in the peripheral samples from T1 and T2, respectively. In particular, the two longest branches, T1Z1 had 19 and T2Z1 had 17 of

these mutated genes. These mutations may have accelerated cell division rates in the peripheral samples, in turn coinciding with a higher number of mutations found in these regions.

Tracing the Temporal and Spatial Tumor Growth under Neutrality and Natural Selection

Although modeling neutrality, we observed that diversity and the subclonal structure emerged rapidly, and higher birth and death rates at the periphery compared with the center led to more mutations and longer branches, consistent with our empirical observations. It suggests that the spatial pattern of mutation accumulation is an intrinsic feature of solid tumors and can be generated in a short period of time.

However, the rapid clonal expansion in central regions of tumors has not appeared in our simulations. Furthermore, whereas α and δ are large, both have low intra-clonal diversity compared with other clones, implying that they might have growth advantages and experienced recent expansion.
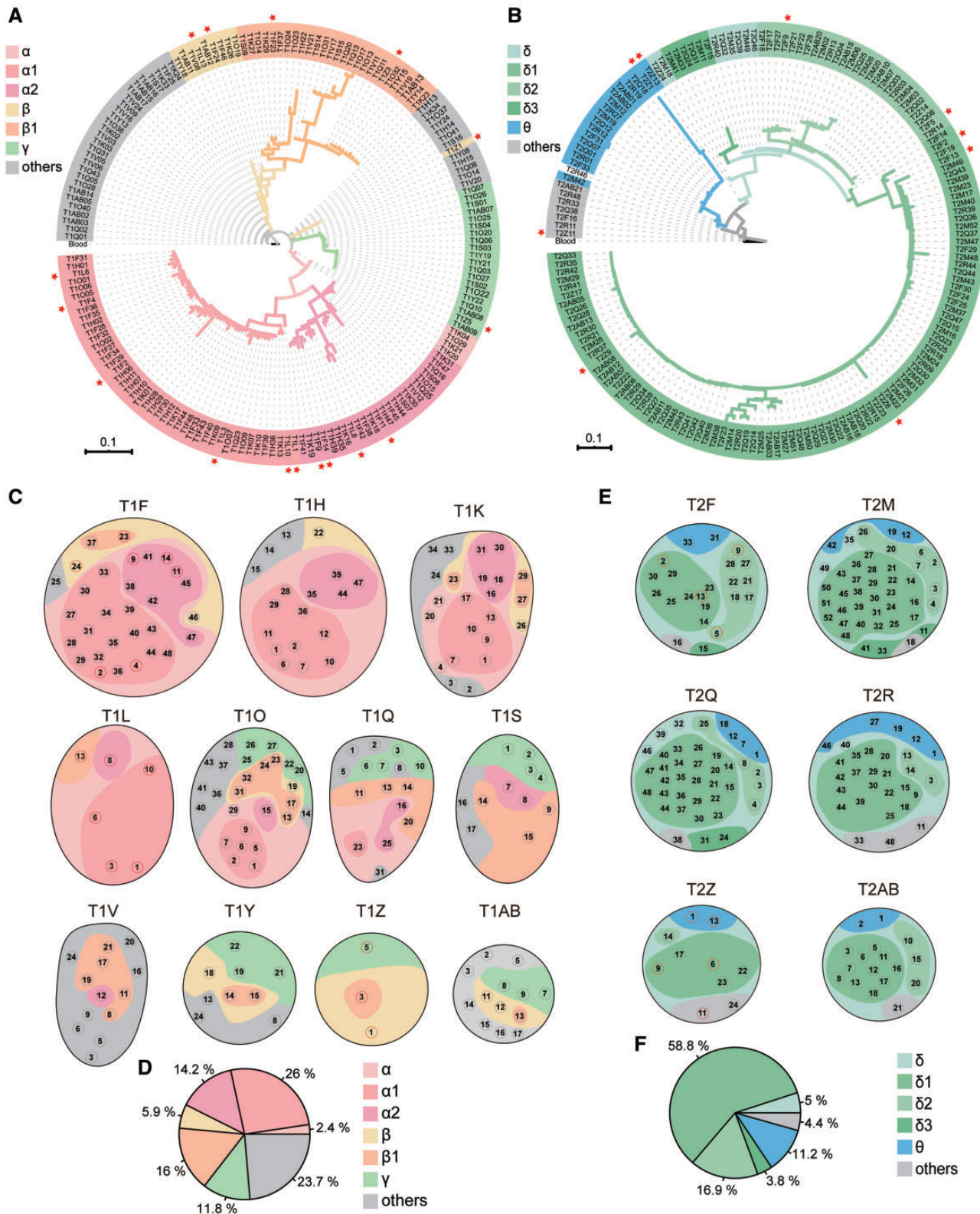
Fig. 4. Extended sample phylogenetic relationships. (A) Extended phylogenetic tree of T1 samples based on WGS and genotyping data. Colors are consistent with figure 1C. (B) Extended phylogenetic tree of T2 samples based on WGS and genotyping data. Colors are consistent with figure 1D. Unit of branch length in (A) and (B) is evolutionary distance in maximum parsimony method. (C) Clonal structure of 11 T1 slices (T1F, T1H, T1K, T1L, T1O, T1Q, T1S, T1V, T1Y, T1Z, and T1AB). These slices were taken from a half of the spherical tumor sequentially from the center to the periphery. T1F is the biggest and T1AB the smallest slice. (D) Proportion of each subclone within T1. (E) Clonal structure of six T2 slices (T2F, T2M, T2Q, T2R, T2Z, and T2AB). The order of slices is also from top to the bottom of the sampled half tumor. T2Q is the largest and T2AB is the smallest slice. (F) Proportion of each subclone within T2. Samples marked with red stars were used for WGS in (A) and (B). Samples in red circles are WGS samples and in gray circles are genotyped samples in (C) and (E). Colors represent different subclones. Tumor regions marked with gray do not belong to defined clones.

We tested whether the growth of the tumors was within neutral expectation by comparing the expected site frequency spectrum (SFS) of exponentially growing cancer cell populations (Durrett 2013; Ohtsuki and Innan 2017) with the observed data (see Materials and Methods). The observed SFSs in both T1 and T2 were significantly different from the expectation under the neutral model (fig. 5A and supplementary table 9, Supplementary Material online) ($\chi^2 = 363,125$, $p < 2.2e^{-16}$ For T1, $\chi^2 = 247,540$, $p < 2.2e^{-16}$ For T2). We then investigated the distribution of adjusted variant allele frequencies (VAFs) (supplementary figs. 6–8 and supplementary material, Supplementary Material online) and CCFs for all SNVs (fig. 5B). If tumor growth is neutral, polymorphic snvs only occur in subclones that are at low frequencies. In contrast, Darwinian selection increases the growth rate of fitter cells, elevating frequencies of snvs they accumulate (Fay and Wu 2000). There was a peak between 0.3–0.4 for VAF and 0.5–0.8 for CCF distributions in both tumors, suggesting selection shaped their growth (Williams et al. 2018). Therefore, Darwinian selection may have driven the evolution of these tumors. We next modeled tumor growth under Darwinian evolution (supplementary table 1, Supplementary Material online). Instead of maintaining similar shapes and proportions of all subclones during tumor progression under the neutral model, random accumulation of driver mutations in different subclones changed these parameters (figs. 2A and 6A). Dominant clones with the highest fitness rapidly increased their proportions in the tumor, whereas other small clones occurring early went extinct as the dominant clone expanded. The distribution of cCFs of SNVs under natural selection had an extra peak in addition to the low-frequency peak after the appearance and expansion of the dominant clone. The peak gradually shifted to the right and finally became fixed when the selected clone accounted for all tumor cells (fig. 6B).

Tree shapes in adaptively evolving tumors also varied according to stage (fig. 6C). Before the emergence of the dominant clone, the tree structure was similar to that of neutral tumors. Once formed, the dominant clone rapidly increased in size and contained an increasing number of samples. Nevertheless, regardless of stage, the subclones that accumulated more mutations were from peripheral regions. This pattern is evident even when the selected clone arose from the central region.

We also explored the relationships between Fst and the physical distance of paired samples within each tumor. Assuming the tumor evolved neutrally, a positive correlation between Fst and physical sample distance is expected. That is because as tumor growth, cells that are farther apart shall have more mutation accumulated between them which, in turn, leads to higher Fst. Nevertheless, tumors under positive selection will increase their size shortly which leaves no time for mutation to accumulate. Thus, a positive correlation between Fst and the physical distance of samples within tumor would not be expected. As shown in supplementary figure 9, Supplementary Material online, there is no correlation between Fst and physical distance of samples within tumor.

Nevertheless, a strong positive correlation between them is revealed in simulated neutral tumors (supplementary fig. 10A, Supplementary Material online). Furthermore, strong positive correlation between Fst and physical distance of samples is found before the emergence of the dominant clone (supplementary fig. 10B, Supplementary Material online). This correlation vanishes as the emergence of dominant clone and reappears after the selected clone accounted for all tumor mass.

The distribution of CCFs of SNVs and clonal structure of T1 and T2 are consistent with simulation results of tumor growth under selection. To estimate the mutation rate (μ) and the number of driver mutations (N) needed to invade of each tumor, we developed a framework incorporating our spatial growth model into approximate Bayesian computation (ABC) (see Materials and Methods) (Beaumont et al. 2002). Mutation rates of T1 and T2 are similar, about 15 per genome per cell division (corresponding to 5×10^{-9} per site per cell division) (table 1). N is four in T1 and three in T2 (table 1). T2 has a smaller N, consistent with the smaller T2 population, and a bigger dominant subclone.

Discussion

Spatial genetic and phenotypic diversity within solid tumors has been well documented (Ling, et al. 2015; Sottoriva et al. 2015; Waclaw et al. 2015; Uchi et al. 2016; Li et al. 2017; Maley et al. 2017; Losic et al. 2020). Nevertheless, how this heterogeneity affects temporal dynamics of tumorigenesis has not been rigorously examined. Here, we developed a spatial tumor growth model where molecular heterogeneity among cancer cells is governed by the rigidity of solid tumors. Relatively high turnover rates of cells at the periphery leads them to accumulate more mutations than at the center. By examining two independently originated tumors, we find that samples from the edge have twice the number of SNVs compared with samples from the core, consistent with our model prediction. As the spatial heterogeneity is generated by the constraint of solid tumors, we propose the model should be applied to at least all small tumors with limited cell migration. For large tumors, angiogenesis can generate a network of blood vessels that sustains optimal growth in the central part of a tumor mass. Therefore, the distinction between peripheral and central is blurred.

A recent study of clear cell renal cell carcinoma (ccRCC) found that cells in the tumor interior exhibited higher Ki67 staining and somatic copy number alteration burden (Zhao et al. 2021). Although the mechanism of higher cellular proliferation at the center of ccRCC requires further investigation, the results actually support the notion that elevated cellular division increases mutation burden and population dynamics actually vary across the range of structured tumor cell populations. It has to point out that, in Zhao et al. (2021), the median tumor size was nine cm and the tumor margin was one cm from the nearest boundary. As the tumors in this study were only 1.5 and 2 cm in size, the definition of central and peripheral are quite different from Zhao et al. (2021). The
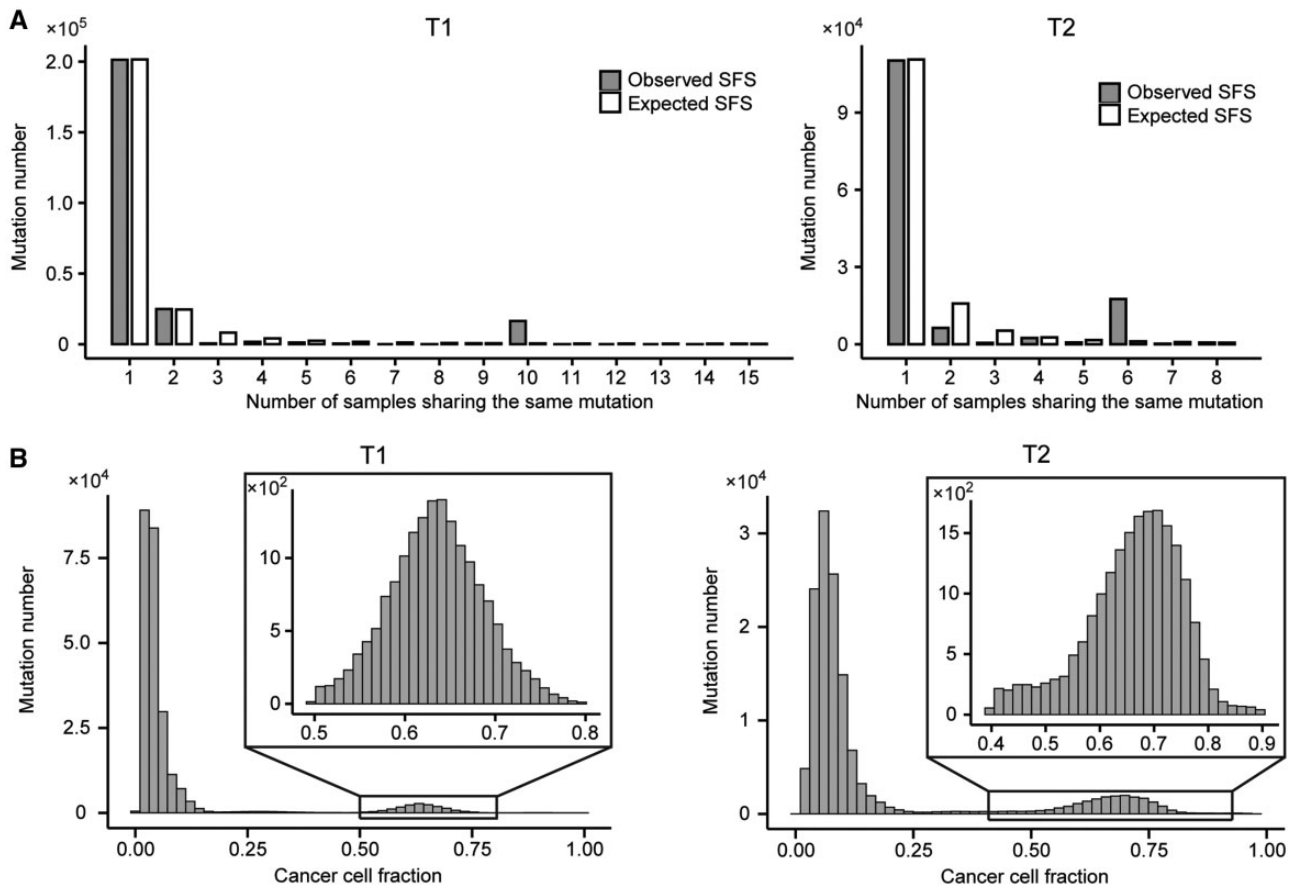
FIG. 5. Distributions of SNV frequencies in tumors. (A) Comparison of the observed site frequency spectrum (SFS) (gray) and the SFS calculated using equation (1) (white). T1 has 16 and T2 has nine samples. (B) Histograms of CCFs of private and shared SNVs for each tumor. VAF of each SNV is adjusted and CCF is calculated (see Materials and Methods) to investigate the prevalence of each SNV in the tumor. Insets are enlarged pictures of the peaks between 0.5–0.8 for T1 and 0.4–0.9 for T2.

results derived from the two studies should not be directly compared.

Long dendrogram branches of peripheral samples have not been previously reported in the literature. This happens due to the high proliferation rate of the cells in these samples. We can detect this pattern because we collected multiple samples that are dense enough within the tumor to capture the spatial heterogeneity. However, as we show in figure 6, natural selection can blur this phenomenon. During and right after selection, the dominant clone can grow quickly making the distinction between the interior and peripheral regions less clear.

Our observation that peripheral cells grow faster than interior in solid tumors may be important to consider in devising cancer therapy. Conventional chemotherapy aims at quickly proliferating cells (Trédan et al. 2007), which may effectively kill peripheral cells but leave interior cells intact. This explains why initial chemotherapy can lead to tumor shrinkage but rarely to complete elimination. That is because interior tumor cells are not dividing much, and are thus relatively insensitive to drugs that interfere with the cell cycle. After eliminating peripheral cells, interior cells get access to more resources and proliferate rapidly. Combination therapy considering diverse phenotypes and tumor microenvironments may be required (Hausser and Alon 2020).

In addition to cancer therapy, our results that tumor cells in different regions have different survival rates resemble the r and K strategies of life history in natural populations (MacArthur 1962; Pianka 1970). Individuals with high reproductive rates are favored in an r-selective environment, whereas those with high survival rates are favored in K-selective conditions. This is a trade-off between proliferation and survival (Aktipis et al. 2013; Li et al. 2020). As population density decreases from the tumor core to edge, we expect that peripheral cells may exhibit decreased survival and increased proliferation. Intriguingly, we find increased cell proliferation and cell cycle gene mutations in peripheral samples, consistent with life-history theory in natural populations.

Our model also demonstrates that detecting positive selection in tumors is a challenging task. Completely different conclusions may have been drawn if samples would have been taken at a different stage of tumor progression. For example, if we would have tested for selection at the four time points from figure 6A, the results would have been neutral, neutral, selection, and neutral, depending on the stage. The main driving forces, that is natural selection and genetic drift, may alternatively dominate during tumor progression (fig. 7). Selection may not be noticed unless samples are taken at the right time and sample size is sufficient to represent the clonal structure of the whole tumor. That is probably why
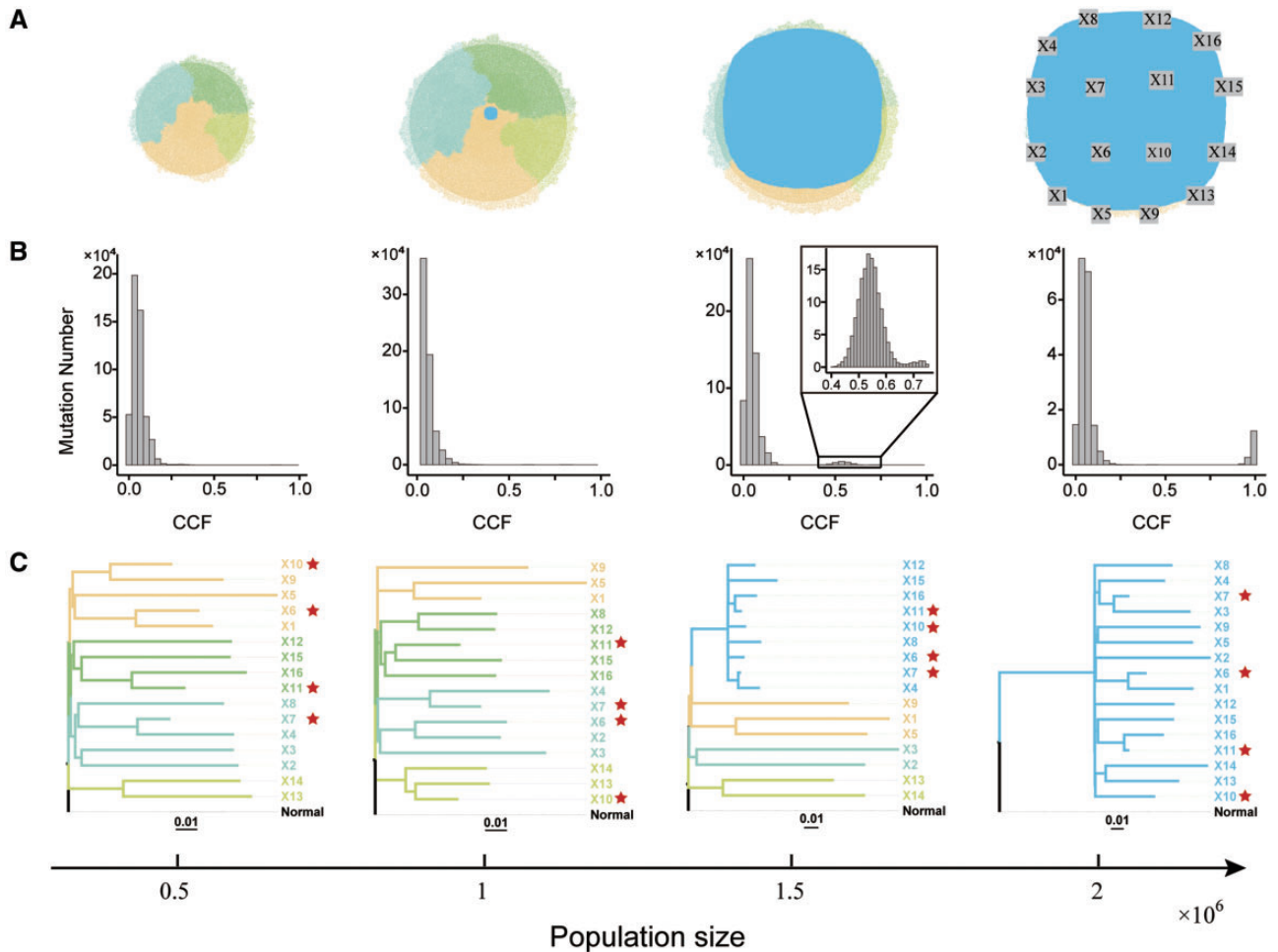
FIG. 6. Clonal structures and phylogenies of tumors under natural selection over time. (*A*) Clonal distributions in the same virtual tumor when cell population size is 5×10^5, 1×10^6, 1.5×10^6, and 2×10^6. Colors represent subclones. (*B*) Sixteen samples, each with about 200 cells, are taken uniformly from the tumor and sequenced virtually at each time point. Position of each sample is in the panel 4 of (*A*). Histograms are distributions of CCFs for all SNVs detected. (*C*) Sample phylogenetic trees. Normal sample without any SNVs is used as the outgroup. Samples marked with red stars are from central regions. Unit of branch length in (*C*) is evolutionary distance in maximum parsimony method.

Table 1. Mutation Rate and Driver Mutation Number Estimation.

Parameter	Value	Posterior Probability	
		T1	T2
u^a	6	0.04	0
	15	0.95	0.97
	30	0.01	0.03
N^b	2	0	0.007
	3	0.34	0.937
	4	0.64	0.056
	5	0.02	0

NOTE—To estimate parameters using in the spatial growth model, we used ABC, generating 30,000 virtual tumors to compare with the real tumor. Details are described in Results and Materials and Methods.
[a]Mutation rate.
[b]Driver mutation number.

several seminal studies suggested that no dominant subclones can be found, indicating non-Darwinian evolution within tumors (Ling et al. 2015; Sottoriva et al. 2015; Uchi et al. 2016). For example, it has been proposed that tumor evolution is neutral based on a linear relationship between the expected number of mutations (*M*) and inverse allele frequency (1/*f*) from one sample per tumor (Williams et al. 2016). Using this model, one-third of the tumors tested were evolving neutrally. However, this method has been widely doubted (Wang et al. 2018; McDonald et al. 2018; Tarabichi et al. 2018). Employing this approach to our data, we find that T1 and T2 are both evolving neutrally (supplementary fig. 11, Supplementary Material online), a conclusion that conflicts with the clonal structure of the two tumors. In conclusion, extremely high spatial intratumor diversity precludes robust interference of evolution modes with only one sequenced sample, necessitating multiregion sampling.

Materials and Methods

NS Model Construction

We constructed an agent-based spatial lattice model to simulate the process of tumor growth and mutation accumulation integrating a heterogeneous tumor microenvironment. Expansion of tumor cells is constrained by the spatial structure (Moore neighborhood, eight neighbors), giving rise to distinct subclones in separate tumor regions.
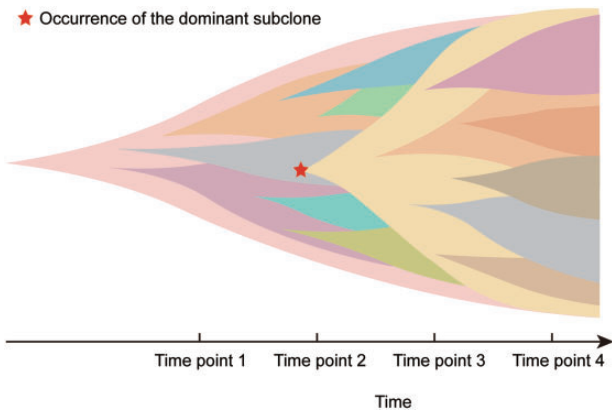
★ Occurrence of the dominant subclone

Time point 1   Time point 2   Time point 3   Time point 4

Time

Fig. 7. Tumor evolutionary process. Dominant driving forces of tumor evolution change according to stage. Genetic drift and natural selection alternate to drive this process. For example, the dominant forces at time points 1, 2, 3, and 4 are genetic drift, genetic drift, natural selection, and genetic drift.

Under neutral evolution, the clonal structure is determined by both spatial constraints and random effects of cell division. Specifically, all simulations were initiated with a single cell and one generation was used as the time unit for a discrete birth-death process. At each time step, each cell either divides, stays quiescent, or dies. When tumors grow to a certain size, spatial structure limits the ranges of cell activities, leading to microenvironment differentiation for cells from distinct regions. Peripheral cells grow faster than inner cells, as there are more resources in their surroundings (Lloyd et al. 2016). The density of immune cells at tumor margins is also higher than that at the cores (Kather et al. 2018), potentially causing higher death rates. Thus, we set different growth rates of inner and outer regions of tumors in our model.

The initial birth rate is b_0 and death rate is $d_0 = 1 - b_0$. Every cell has a unique spatial position in the two-dimensional coordinate system. Each cell's destiny is decided by a random value generated from a uniform distribution Uniform (0, 1). When the value is smaller than or equal to b_0 and there is an adjacent vacancy in the Moore neighborhood, the cell proliferates. If all neighborhood positions are occupied, a cell stays quiescent. If the random value is bigger than b_0, the cell dies. We used a moderate initial birth rate and relatively high death rate: $b_0 = 0.53$, $d_0 = 0.47$, based on the previous estimates in solid tumors (Sun et al. 2017). When spatial structure influences tumor cell growth, cells in central and peripheral regions differ in birth and death rates. The time point where the spatial structure starts to influence cell phenotypes is at ~80 generations, when the tumor population size is approximately 200 cells. The quiescent rate q in central regions is 0.5. The birth rate decreases to $b_1 = 0.3$ and the death rate is $1 - b_1 - q$. The birth rate and death rate of cells at the tumor margin remain b_0 and $1 - b_0$, respectively. To define the tumor regions, we first estimate the radius of the circle-like tumor, r. For each cell, we calculate its distance to the tumor center, d. If $d < 0.9*r$, the cell is assigned to the central region. Otherwise, it is considered peripheral.

Mutations accumulate randomly according to the Poisson distribution during cell proliferation, assuming infinite site model and a constant mutation rate. The neutral mutation rate is μ and the probability that i neutral mutations accumulate after each cell division is:

$$P(x = i) = \frac{\mu^i e^{-\mu}}{i!}.$$

The mutation accumulation parameters are set according to previous studies. The neutral mutation rate $\mu = 15$ per genome (corresponding to 5×10^{-9} per site per cell division for the human genome) (Sun et al. 2017). Parameters used in the spatial tumor growth model are listed in supplementary table 1, Supplementary Material online.

To investigate the dynamics of clonal structure and cell lineages, we took 16 samples uniformly when tumor size reached 5×10^5, 1×10^6, 1.5×10^6, and 2×10^6 (fig. 2). To simulate the sequencing process for each bulk sample, we used a negative binomial distribution to generate read depth of each virtual point mutation and a binomial distribution to generate the mutant allele read counts. The parameters of the negative binomial distribution were estimated by fitting the depth distribution from real sequencing data, and are NB (15.8, 0.155). The binomial distribution parameters are the generated depth and the true VAF. We applied the same filtering condition to these simulated SNVs as we did to real data to make them comparable to the observed VAFs. CCFs of SNVs were two times the VAFs. CCFs and VAFs of all SNVs from whole tumors were obtained by integrating virtual sequencing results from all samples.

Clinical Information

The patient was a 57-year-old man with a chronic Hepatitis B Virus (HBV) infection and liver cirrhosis. He had undergone liver transplantation in the General Hospital of Chinese People's Armed Police Forces. Two well-encapsulated tumors: T1, ~ 2 cm in diameter, and T2, ~1.5 cm in diameter, were on the right lobe and left lobe of the liver, respectively. Both tumors were diagnosed as moderately differentiated Hepatocellular carcinoma (HCC). This study was approved by the Ethics Review Committee of the General Hospital of Chinese People's Armed Police Forces. Informed consent was signed according to the regulations of the institutional ethics review board.

Three-Dimensional Microsampling and WGS

We performed a three-dimensional microsampling of both tumors (fig. 3). Half of each tumor was embedded in the optimal cutting temperature (OCT) compound and sliced into 200-μm-thick pieces using the Leica CM1950 cryostat platform. We chose three slices of T1 (T1F, T1L, and T1Z) and two slices of T2 (T2F and T2Z) to conduct microdissection sampling for sequencing. To identify subclonal alterations at a high resolution, we performed a sampling of the frozen slices using syringes with blunt tip needles of 210 μm inner diameter. The 20 μl of cell lysis buffer (20 mM Tris, 2 mM EDTA, 20 mM KCl, 1 μg/μl QIAGEN protease, 1 μg/μl RNAse A) was

piped into the syringes before sampling to maximize a sampling success rate. We obtained 16 samples from T1 and nine samples from T2. Given that tumor cell radius on the hematoxylin-eosin-stained slice is $\sim$8 μm and the volume of a cylinder-like microsample is $\sim$7 $\times$ 10^6 μm^3, there are $\sim$3,200 cells in a sample.

Cells were directly lysed to reduce DNA loss and increase the efficiency of library preparation for WGS. DNA from blood cells collected from the patient before liver transplantation was used as the normal control. Libraries were constructed using Ovation Ultralow System Kit (NuGEN, USA). Genomic DNA was fragmented into about 500 bp segments using a Covaris S220 instrument. DNA fragments were end-repaired, ligated to adapters, and amplified following the Ovation Ultralow System Kit protocol. Purified libraries were paired-end sequenced (2 $\times$ 150 bp) using Illumina HiSeq X Ten sequencers at Novogene.

Paired-end reads were mapped to the human reference genome (UCSC hg19) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin 2010). All aligned reads were further processed using PICARD (http://broadinstitute.github.io/picard) and the Genome Analysis Toolkit (GATK) (DePristo et al. 2011), including de-duplication, base quality recalibration, and multiple-sequence realignment prior to mutation detection. The alignments were assessed using Qualimap (Okonechnikov et al. 2016) Sequence data summaries for each sample are listed in Supplementary table 2, Supplementary Material online.

Sample Purity Estimation

All samples from a single tumor originated from the same ancestor with thousands of SNVs in common. Theoretically, the frequency of these ubiquitous SNVs located at regions without CNAs and LOHs in a single tumor should be 0.5 if the sample purity is 100%. Next-generation sequencing for each SNV is a binomial sampling process and therefore ubiquitous SNV frequencies are normally distributed. Using two-copy SNVs with minor and major allele both one in each sample, we fit a normal distribution to obtain the mean value, μ. The purity of a sample is then 2μ (supplementary table 2, Supplementary Material online).

Phylogenetic Tree Reconstruction of Samples Subject to WGS

We reconstructed the sample phylogenies using maximum parsimony methods in PHYLIP (Felsenstein 1989) based on SNVs detected by WGS in each tumor. We only used SNVs outside regions exhibiting loss of heterozygosity (LOH).

Test of Selection

The SFS is the allele frequency distribution of a given set of loci in a population. It is often used to test for the presence of natural selection in population genetics. In this study, the loci are all SNVs obtained from the WGS and the populations are the 16 T1, nine T2 samples. The observed SFS for each tumor was obtained by counting the number of derived allele occurrences in all samples (supplementary table 9, Supplementary

Material online). The expected SFS for each tumor is calculated using the formula (Durrett 2013):

$$
\mathrm{E}\varepsilon_i = \begin{cases} \dfrac{\mu}{r}\displaystyle\sum_{k=1}^{Nr}\dfrac{n}{n+k}\dfrac{k}{n+k-1} & i=1 \\[2mm] \dfrac{\mu}{r}\dfrac{n}{i(i-1)} & 2 \le i < n \end{cases}.
$$

Since $\displaystyle\sum_{i=2}^{n}\frac{n u}{r}\frac{1}{i(i-1)} = \sum_{i=2}^{n}$ Observed SFS, we obtained $\frac{u}{r}$ for each tumor and calculated all SFS values.

Spatial Tumor Growth Model under Natural Selection

We next modeled tumor growth under Darwinian evolution to further investigate the effects of different forces driving cancer progression. We assumed driver mutations occur at the rate of 10^{-6} per cell division (Bozic et al. 2010). The probability of j driver mutations after each cell division is:

$$
P(y=j) = \frac{u_s^j e^{-\mu_s}}{j!}.
$$

Driver mutations elevate the birth rate in models with selection. When a cell gains a beneficial mutation, the birth rate b, increases by the selection coefficient (s) and the death rate decreases. We calculated birth and death rates for each cell at the tumor core as

$$
b = b_1(1+s)^{n_a}, \quad d = 1 - b - q.
$$

In this equation, b_1 is the initial central cell birth rate, q is the quiescence rate, and n_a is the number of accumulated driver mutations. We modeled tumor growth with a strong selective advantage ($s = 0.1$) (Uchi et al. 2016). The birth and death rates for every cell at the tumor margin were b_0 and $1 - b_0$, respectively (fig. 6A), with b_0 higher than b_1. Given that the number of drivers accumulated is limited in tumor progression (Tomasetti et al. 2015; The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium 2020), we assumed that cells gained ability to expand quickly when they have N driver mutations. In our simulations, N is chosen to be four, according to previous estimates from solid tumors (Martincorena et al. 2017).

Estimation of Tumor Growth Model Parameters

We combined our spatial tumor growth model with the approximate Bayesian computation (ABC) (Beaumont et al. 2002) framework to estimate mutation rate (μ) and the number of driver mutations (N) needed for tumors to expand quickly. Mutation rate (μ) was randomly sampled from a discrete uniform distribution with five values from one to 30 (1, 3, 6, 15, 30, per genome per cell division). The prior distribution of N was also a discrete uniform distribution with six values from one to six. We simulated 1,000 tumors for each combination of the two parameters, generating 30,000 virtual tumors. We estimated parameters of T1 and T2 independently considering their size and evolutionary process. Cell numbers for a two-dimensional slice of T1 and T2 were estimated at about 1.5×10^6 and 10^6, respectively.

When virtual tumors grew to about 10^6, we took nine samples uniformly and simulated the sequencing process described above to compare the simulated to T2 sequence data. We performed the same procedure for 1.5×10^6 T1 cells. The ABC procedure is as follows:

(1) Sample parameter $\theta = [\mu, \ N]$ from the prior distribution $f(\theta)$.
(2) Simulate tumor growth with the sampled parameters θ and calculate summary statistics S'.
(3) Compare S' with the observed S using a distance function d and tolerance rate ε. If $d(S', \ S) < \varepsilon$, accept θ.
(4) Go to step (1).

We accepted parameters which generated sequencing results similar to those we observed in T1 and T2 using a distance function d, a set of summary statistics S, and an error threshold ε. The distance function d was Euclidean distance and error threshold ε was 0.01. We used 15 summary statistics to incorporate as much information as possible to compare observed and simulated data. The summary statistics are:

(1) The number of mutations present in four intervals: [0.02, 0.25], [0.25, 0.4], [0.4, 0.9], and [0.9, 1], designated as S_1, S_2, S_3, and S_4, respectively. These four values reflect the distribution of SNV frequencies.
(2) The median, mean, and standard deviation of Fst between samples, designated as S_{F1}, S_{F2}, and S_{F3}, respectively. These three values reflect the genetic distance between samples in a tumor.
(3) The median, mean, and standard deviation of Kolmogorov-Smirnov distance between samples, designated as S_{K1}, S_{K2}, and S_{K3}, respectively. These three values reflect the similarity of mutation frequency spectra among samples within a tumor.
(4) The maximum, minimum, mean, median, and standard deviation of mutation numbers detected for each tumor, designated as S_{N1}, S_{N2}, S_{N3}, S_{N4}, and S_{N5}, respectively. These five values reflect the difference of mutation numbers among samples in a tumor.

We performed 30,000 spatial simulations described above and calculated the summary statistics. abc (Csilléry et al. 2012) package was used to calculate distances and generate posterior distributions of the two parameters.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Author Contributions

Data Availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Chen et al. 2021) in National Genomics Data Center (Members and Partners 2021), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA000188) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human.

Code Availability

Tumor spatial growth model code is available at https://github.com/guanghaoli/Tumor-spatial-growth-model. The code of ABC model is available at https://github.com/guanghaoli/Tumor_growth_ABC.

References

Aktipis CA, Boddy AM, Gatenby RA, Brown JS, Maley CC. 2013. Life history trade-offs in cancer evolution. *Nat Rev Cancer*. 13(12):883–892.

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, et al.; PCAWG Consortium. 2020. The repertoire of mutational signatures in human cancer. *Nature* 578(7793):94–101.

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, et al.; Cancer Genome Atlas Research Network. 2018. Comprehensive characterization of cancer driver genes and mutations. *Cell* 173(2):371–385.e18.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.

Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B, Nowak MA. 2010. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A*. 107(43):18545–18550.

Bréchot C, Gozuacik D, Murakami Y, Paterlini-Bréchot P. 2000. Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). *Semin Cancer Biol*. 10(3):211–231.

Carmona-Fontaine C, Deforet M, Akkari L, Thompson CB, Joyce JA, Xavier JB. 2017. Metabolic origins of spatial organization in the tumor microenvironment. *Proc Natl Acad Sci U S A*. 114(11):2934–2939.

Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O'Kane GM, Connor AA, Denroche RE, Grant RC, McLeod J, et al. 2020. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat Genet*. 52(2):231–240.

Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, Dong L, Zhang Z, Yu C, Sun Y, et al. 2021. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics*. Advance Access published Aug 13, 2021, doi: 10.1016/j.gpb.2021.08.001.

Chowell D, Napier J, Gupta R, Anderson KS, Maley CC, Sayres MAW. 2018. Modeling the subclonal evolution of cancer cell populations. *Cancer Res*. 78(3):830–839.

Csilléry K, François O, Blum MGB. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol*. 3(3):475–479.

Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, Yu K, Rubanova Y, Macintyre G, Demeulemeester J, et al.; PCAWG Evolution and Heterogeneity Working Group and the PCAWG Consortium. 2021. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* 184(8):2239–2254.e39.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43(5):491–498.

Durrett R. 2013. Population genetics of neutral mutations in exponentially growing cancer cell populations. *Ann Appl Probab*. 23(1):230–250.

Fay JC, Wu C-I. 2000. Hitchhiking under positive darwinian selection. *Genetics* 155(3):1405–1413.

Felsenstein J. 1989. Phylogeny inference package. *Cladistics* 5:164–166.

Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, Mitchell TJ, Rubanova Y, Anur P, Yu K, et al.; PCAWG Consortium. 2020. The evolutionary history of 2,658 cancers. *Nature* 578(7793):122–128.

Hausser J, Alon U. 2020. Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nat Rev Cancer*. 20(4):247–257.

Hendry AP. 2016. Eco-evolutionary dynamics. Princeton (NJ): Princeton University Press.

Kather JN, Suarez-Carmona M, Charoentong P, Weis CA, Hirsch D, Bankhead P, Horning M, Ferber D, Kel I, Herpel E, et al. 2018. Topography of cancer-associated immune cells in human solid tumors. *Elife* 7:e36967.

Li C, Hou Y, Xu J, Zhang A, Liu Z, Qi F, Yang Z, Chen K, Liu S, Huang H, et al. 2017. A direct test of selection in cell populations using the diversity in gene expression within tumors. *Mol Biol Evol*. 34(7):1730–1742.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.

Li T, Liu J, Feng J, Liu Z, Liu S, Zhang M, Zhang Y, Hou Y, Wu D, Li C, et al. 2021. Variation in the life history strategy underlies functional diversity of tumors. *Natl Sci Rev*. 8(2):nwaa124.

Li X, Xu W, Kang W, Wong SH, Wang M, Zhou Y, Fang X, Zhang X, Yang H, Wong CH, et al. 2018. Genomic analysis of liver cancer unveils novel driver genes and distinct prognostic features. *Theranostics* 8(6):1740–1751.

Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbel JO, Haber JE, et al.; PCAWG Consortium. 2020. Patterns of somatic structural variation in human cancer genomes. *Nature* 578(7793):112–121.

Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, Chen K, Dong L, Cao L, Tao Y, et al. 2015. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci U S A*. 112(47):E6496–E6505.

Lloyd MC, Cunningham JJ, Bui MM, Gillies RJ, Brown JS, Gatenby RA. 2016. Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces. *Cancer Res*. 76(11):3136–3144.

Losic B, Craig AJ, Villacorta-Martin C, Martins-Filho SN, Akers N, Chen X, Ahsen ME, von Felden J, Labgaa I, D'Avola D, et al. 2020. Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat Commun*. 11(1):291.

MacArthur RH. 1962. Some generalized theorems of natural selection. *Proc Natl Acad Sci U S A*. 48(11):1893–1897.

Maley CC, Aktipis A, Graham TA, Sottoriva A, Boddy AM, Janiszewska M, Silva AS, Gerlinger M, Yuan Y, Pienta KJ, et al. 2017. Classifying the evolutionary and ecological features of neoplasms. *Nat Rev Cancer*. 17(10):605–619.

Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ. 2017. Universal patterns of selection in cancer and somatic tissues. *Cell* 171(5):1029–1041.e21.

McDonald TO, Chakrabarti S, Michor F. 2018. Currently available bulk sequencing data do not necessarily support a model of neutral tumor evolution. *Nat Genet*. 50(12):1620–1623.

Members C-N, Partners. 2021. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2021. *Nucleic Acids Res*. 49:D18–D28.

Ohtsuki H, Innan H. 2017. Forward and backward evolutionary processes and allele frequency spectrum in a cancer cell population. *Theor Popul Biol*. 117:43–50.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32(2):292–294.

Pianka ER. 1970. On r- and K-selection. *Am Nat*. 104(940):592–597.

Schoener TW. 2011. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* 331(6016):426–429.

Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, Marjoram P, Siegmund K, Press MF, Shibata D, et al. 2015. A Big Bang model of human colorectal tumor growth. *Nat Genet*. 47(3):209–216.

Sun R, Hu Z, Sottoriva A, Graham TA, Harpak A, Ma Z, Fischer JM, Shibata D, Curtis C. 2017. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat Genet*. 49(7):1015–1024.

Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, Lee WH, Ariyaratne PN, Tennakoon C, et al. 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet*. 44(7):765–769.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. 2019. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 47(D1):D941–D947.

Tarabichi M, Martincorena I, Gerstung M, Leroi AM, Markowetz F, Spellman PT, Morris QD, Lingjærde OC, Wedge DC, Van Loo P; PCAWG Evolution and Heterogeneity Working Group. 2018. Neutral tumor evolution? *Nat Genet*. 50(12):1630–1633.

Terry S, Buart S, Chouaib S. 2017. Hypoxic stress-induced tumor and immune plasticity, suppression, and impact on tumor heterogeneity. *Front Immunol*. 8:1625.

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. 2020. Pan-cancer analysis of whole genomes. *Nature* 578:82–93.

Tomasetti C, Marchionni L, Nowak MA, Parmigiani G, Vogelstein B. 2015. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc Natl Acad Sci U S A*. 112(1):118–123.

Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, Tsuji S, Donehower LA, Slagle BL, Nakamura H, et al. 2014. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet*. 46(12):1267–1273.

Trédan O, Galmarini CM, Patel K, Tannock IF. 2007. Drug resistance and the solid tumor microenvironment. *J Natl Cancer Inst*. 99(19):1441–1454.

Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, Sawada G, Iwaya T, Kurashige J, Shinden Y, et al. 2016. Integrated multiregional analysis proposing a new model of colorectal cancer evolution. *PLoS Genet*. 12(2):e1005778.

Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. 2015. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* 525(7568):261–264.

Wang H-Y, Tong D, Chen Y, Ling S, Wu C-I, Lu X, Hu Z, Tao Y. 2018. Is the evolution in tumors Darwinian or non-Darwinian? *Natl Sci Rev*. 5(1):15–17.

Wang K, Lim HY, Shi S, Lee J, Deng S, Xie T, Zhu Z, Wang Y, Pocalyko D, Yang WJ, et al. 2013. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* 58(2):706–717.

Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. 2016. Identification of neutral tumor evolution across cancer types. *Nat Genet.* 48(3):238–244.

Williams MJ, Werner B, Heide T, Curtis C, Barnes CP, Sottoriva A, Graham TA. 2018. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat Genet.* 50(6):895–903.

Wu CI, Wang HY, Ling S, Lu X. 2016. The ecology and evolution of cancer: the ultra-microevolutionary process. *Annu Rev Genet.* 50:347–369.

Yoshida T, Jones LE, Ellner SP, Fussmann GF, Hairston NG Jr. 2003. Rapid evolution drives ecological dynamics in a predator-prey system. *Nature* 424(6946):303–306.

Yuan Y. 2016. Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harb Perspect Med.* 6(8):a026583.

Zhai W, Lim TK, Zhang T, Phang ST, Tiang Z, Guan P, Ng MH, Lim JQ, Yao F, Li Z, et al. 2017. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. *Nat Commun.* 8:4565.

Zhao Y, Fu X, Lopez JI, Rowan A, Au L, Fendler A, Hazell S, Xu H, Horswell S, Shepherd STC, et al.; TRACERx Renal Consortium. 2021. Selection of metastasis competent subclones in the tumour interior. *Nat Ecol Evol.* 5(7):1033–1045.