

High-density sampling reveals volume growth in human tumours

Reviewed Preprint

v2 • October 3, 2024

Revised by authors

Reviewed Preprint

v1 • May 9, 2024

Arman Angaji, Michel Owusu, Christoph Velling, Nicola Dick, Donate Weghorn , Johannes Berg 

Institute for Biological Physics, University of Cologne, Cologne, Germany • Centre for Genomic Regulation, Barcelona, Spain • Universitat Pompeu Fabra (UPF), Barcelona, Spain

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

In growing cell populations such tumours, mutations can serve as markers that allow tracking the past evolution from current samples. The genomic analyses of bulk samples and samples from multiple regions have shed light on the evolutionary forces acting on tumours. However, little is known empirically on the spatio-temporal dynamics of tumour evolution. Here, we leverage published data from resected hepatocellular carcinomas, each with several hundred samples taken in two and three dimensions. Using spatial metrics of evolution, we find that tumour cells grow predominantly uniformly within the tumour volume instead of at the surface. We determine how mutations and cells are dispersed throughout the tumour and how cell death contributes to the overall tumour growth. Our methods shed light on the early evolution of tumours in vivo and can be applied to high-resolution data in the emerging field of spatial biology.

eLife Assessment

The article uses a cell-based model to investigate how mutations and cells spread throughout a tumour. The paper uses published data and the proposed model to understand how growth and death mechanisms lead to the observed data. This work provides an **important** insight into the early stages of tumour development. From the work provided here, the results are **convincing**, using a thorough analysis.

<https://doi.org/10.7554/eLife.95338.2.sa3>

The evolution of a solid tumour is governed by the division, motion, and death of cancer cells. Genetic mutations arising during cell divisions can serve as cell markers to track this dynamics. From the observed spatial distribution of mutations, it should in principle be possible to infer the spatio-temporal principles of tumour evolution: Is the growth rate uniform across the tumour, or does growth predominantly take place near the edge of the tumour? What is the interplay between the tissue dynamics of the tumour and its genetic evolution? These broad modes of tumour evolution affect for instance the signature of neutral evolution, the response to selection, or the number of low-frequency mutants which can confer therapy resistance [1, 2].

However, to answer such questions on the basis of genetic tumour data is challenging because only partial information is available: (i) The sequencing depth (average number of reads covering a nucleotide in NGS sequencing) is finite. This means that only high-frequency mutations are observed. (ii) Usually only a small number of samples are taken from different parts of a solid tumour, which limits the information on mutations present in the full tumour [3, 4]. (iii) Longitudinal data from ctDNA measurements are highly limited in the observable range of mutations and provide noisy frequency estimates.

Over the next years, some of these restrictions will be lifted by the advent of spatial genomics [5, 6, 7]. These techniques allow assaying the genomic information almost at single-cell level in intact tissue sections. Currently, the attainable sequencing depth is too low to identify point mutations across different parts of the tumour. However, it is clear that the coming-of-age of spatial genomics will bring new opportunities to understand the past evolution of a population of tumour cells from a late-stage snapshot. This implies a need for new tools to analyse spatio-temporal evolution, since standard tools of population genetics, like the site-frequency spectrum, are designed for spatially mixed populations and disregard spatial information.

One particular question concerns two different modes in which a tumour can grow; surface growth and volume growth. Under surface growth, the cancer cells divide predominantly at the border with healthy tissue. The potential reasons for this spatial dependence include higher nutrient levels near normal tissue, higher levels of metabolic waste products in the tumour bulk, or mechanical stress in the tumour centre [8, 9]. A faster growth rate at the edge of the tumour leads to a radially outward growth of the cell population. The surface growth mode is well-known from bacterial growth [10]. In tumours, some evidence for surface growth comes from histological stainings, which show an enhanced level of the Ki-67 protein (a cellular marker for proliferation) near a tumour surface [11, 12]. However, the reverse situation has been found as well [6], with elevated Ki-67 levels near the centre of a renal carcinoma. The surface growth mode also has a long history in the modelling of tumour evolution [13, 14, 15, 16, 17, 18, 19, 20, 21, 22] and has been used to analyze multi-region tumour sequencing data [19, 21, 23, 24, 25, 26].

In volume growth, on the other hand, cancer cells grow irrespective of their location in the tumour: although each cell has a physical location, and upon division its offspring is in a similar location, location does not affect cell division or death. Under volume growth, subclones can originate from any location in the tumour [6]. As a result, under volume growth mutation frequencies evolve exactly in the same way they would do in a well-mixed population, even though the resulting tumour will generally be spatially heterogeneous (see below). Mixed population models have been used extensively to model all aspects of tumor evolution, from tumorigenesis to the formation of metastases and the response to therapy [27, 28, 29, 30, 31, 32]. They also form the basis of almost all population-genetic approaches to analyzing tumour data [33, 34, 35].

To tell between these two different evolutionary dynamics, we use high-resolution data on the spatial distribution of mutations found in solid tumours. We use whole-exome mutation data obtained from hepatocellular carcinomas and published previously [36, 24]. From these snapshots of latestage tumours we infer how the tumours grew in earlier stages. To this end, we develop metrics of intra-tumour heterogeneity which leverage the information from the spatial position of all samples.

Methods

Spatially resolved data

We analyze two datasets where large numbers of samples (> 100) were taken from different hepatocellular tumours, once from a two-dimensional section [36], and once from a three-dimensional microsampling [24].

In Ling *et al.* [36], 285 samples of a planar section of a hepatocellular carcinoma of diameter 35 mm were taken and analyzed. Each sample consisted of about 20000 cells [36] determined by cell counting. 23 of these samples were subjected to whole-exome sequencing (average read depth of 74 reads per nucleotide), and loci with mutations found in these 23 samples were probed by genotyping (by Sequenom and Sanger sequencing) in all 285 samples [36]. **Figure 1A** shows the positions of the samples and the mutations found at each position. The data thus effectively consist of a set of 285 samples at high spatial resolution, but with limited genomic information (genotyping), and a subset of 23 samples at high genomic resolution from whole-exome genomic sequencing, but limited spatial resolution. We reanalyzed the whole-exome data as described in SI S4.2. The genotyped data consists of 35 mutated loci. Calling variants jointly on all 23 whole-exome samples and filtering as described in SI S4.2 returned 217 mutated loci.

In the second data set, Li *et al.* [24], samples were taken from multiple planar sections of two different hepatocellular tumours from the same patient. This gives a three-dimensional picture of the genetic heterogeneity of these tumours. 169 samples were taken from 11 sections of tumour T1, which was approximately of 20 mm diameter, and 160 samples from 6 sections of tumour T2, which was of approximately of 15 mm diameter. Each sample consisted of ca. 3200 cells. 16 (9) samples from T1 (T2) were whole-genome sequenced and the remainder genotyped using the Ampliseq method. Li *et al.* report 906 (564) mutations used for genotyping of T1 (T2), of which 563 (259) pass our filters.

Computational model of volume and surface growth

We use an off-lattice cell-based model, well-known from simulations of tissue mechanics [37, 38, 39, 40], which we endow with mutations entering the population at a constant rate per cell division.

In this model, each tumour cell is described by a sphere centered on a particular point in 3d space or 2d space, as well as a genome. Each cell division introduces a new point adjacent to its parent in a random (uniformly distributed) direction, and each cell death removes a point. At division both mother and daughter cell can acquire mutations, the number of mutations is Poisson distributed with mean μ . The spherical are not allowed to overlap and can push one another out of the way to achieve this. This off-lattice approach differs from cellular automaton models of tumour growth, where cells occupy discrete lattice sites [1, 2]. Specifically, in off-lattice models like the one we use, cells can gradually push one another out of the way during growth, see below. Cells cannot move by themselves but as they divide overlaps are introduced and resolved by pushing the neighboring cell.

We implement the resolution of overlaps introduced by a cell division using a ‘width-first’ pushing algorithm, where the immediate neighborhood of the newly placed cell is resolved first before proceeding to the new overlaps this step introduces. A fast search first identifies overlaps with the current cell in its close neighborhood. The cell then pushes its neighbor along their connecting axis only by the overlap plus a small margin which acts as a buffer and avoids many tiny recurring overlaps. The neighboring cell is then added to a queue for subsequent iterations. Neighborhoods are shuffled such that there is no particular order in which cells are pushed. After each overlapping cell in the neighborhood is pushed and added to the queue, the search for overlaps is repeated on the first cell in the queue, adding new pairs to the queue, which are in turn iteratively resolved until the queue is empty. Thereby each cell division can set off a cascade

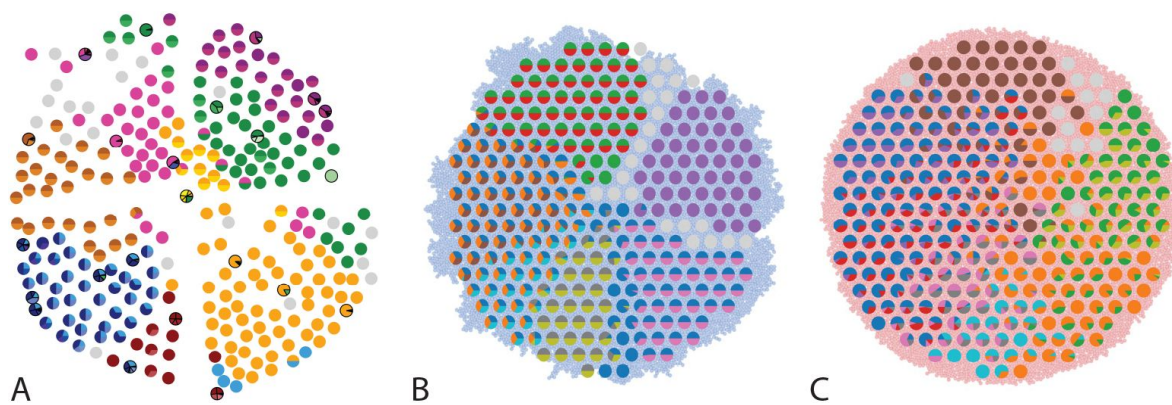


Figure 1

Multi-region sampling of a hepatocellular tumour and cell-based simulations.

A shows the spatially resolved sequencing data of 285 samples of a hepatocellular carcinoma analyzed by Ling *et al.* [36]. Each sample is indicated by a small pie chart in which colors indicate specific mutations, and slice sizes indicate the mutation frequencies within each sample. The 23 samples highlighted by a black outline were also subjected to whole-exome sequencing. The samples form a honey-comb structure, because the tumour slice had been cut into four quadrants, see Fig S1 in [36]. B and C Results of a cell-based simulation in the surface growth mode (B) and the volume growth mode (C). In each case, 280 evenly spaced samples were taken from the population of 10000 cells of a 2d simulation, see text. The most frequent mutations are shown as in (A), superimposed on the structure of the simulated tumour.

of rearrangements of cell positions. This level of microscopic detail of course comes at a computational cost. Using off-lattice cell-based modelling, population sizes of several ten thousand cells can be simulated, compared to billions of cells with a cellular automaton [1].

In regions of high cancer cell density, the division rate may be reduced (for instance due to lack of nutrients, toxic metabolic products, or mechanical stress). We effectively encode such potential spatial effects in the rate at which individual cells divide: the cell division rate $b = b(\rho)$ depends on the local density of cells $\rho = \rho(\mathbf{r})$. For simplicity, we consider a division rate $b(\rho)$ which decreases in a straight line from $b(\rho = 0) = b_0$ to $b(\rho = \rho_c) = 0$, see Fig. S1. Setting $\rho_c = \infty$ leads to a constant division rate for all cells, irrespective of their local density. This growth mode is termed volume growth. Lower values of ρ_c lead to an increased growth rate in parts of the tumour where the local density of cancer cell is low, and hence enhance the growth rate at the surface of the tumour relative to the bulk of the tumour. Surface growth is characterized by a ρ_c equal to the density defined by the minimal cell distance. By changing ρ_c one can thus tune the dynamics of this model continuously from volume to surface growth. To describe the genetic changes in the tumour, we use an infinite-sites model and a constant mutation rate per cell division. **Figure 1B** and **C** show the resulting tumor sections for the surface and volume growth mode, respectively.

For a detailed description of the model, its implementation and performance in comparison with variants of the kinetic Monte Carlo and pushing algorithms, see SI S1. The code is available as a Julia package at <https://github.com/aangaji/TumorGrowth>.

Results

Volume versus surface growth

Under surface growth, cells divide predominantly near the edge of a tumour. Under volume growth on the other hand, cells divide uniformly across the entire tumour mass. To distinguish these two modes in empirical data, we look at the angle between a line from the tumour centre to a parent clone and from the parent to their offspring, see **Fig. 2A** and SI S2. Under surface growth, offspring tend to lie radially outward from their parents, leading to a distribution of these direction angles centered around zero. Under volume growth, cells divide isotropically, leading to a uniform distribution of angles. The distribution of direction angles is a model-independent metric, it does not rely on a particular model of cell dynamics, population dynamics or sampling. It is also robust against the presence of selected subclones: if a subclone grows isotropically from some point, but at an elevated rate, it will contribute to a flat distribution of angles like all other clones.

In **Fig. 2B** we show the distribution of direction angles found in the spatially resolved data of Ling *et al.* [36]. The empirical data clearly show a uniform distribution of the direction angle, and thus no evidence for radially outward growth caused by faster growth near the tumour's edge. For comparison, we look at the distribution of direction angles in simulations of different growth modes. For surface growth, the histogram of direction angles shows a pronounced maximum at $\theta = 0$ produced by the radial outgrowth of clones (**Fig. 2C**), in a simulation of volume growth, we find a flat distribution (**Fig. 2D**). These simulations were done with samples placed in the same positions as in the real tumour to eliminate effects from uneven sampling.

This result is corroborated by the average number of mutations of a clone as a function of its distance from the tumour centre. Under surface growth and neutral evolution, the number of cell divisions since tumorigenesis, and hence the number of mutations found in a cell, increases with distance from the tumour centre. However, we do not find this signature of surface growth in the tumour data. Instead, we find that the number of mutations at different distances from the centre is compatible with volume growth (see Fig. S11).

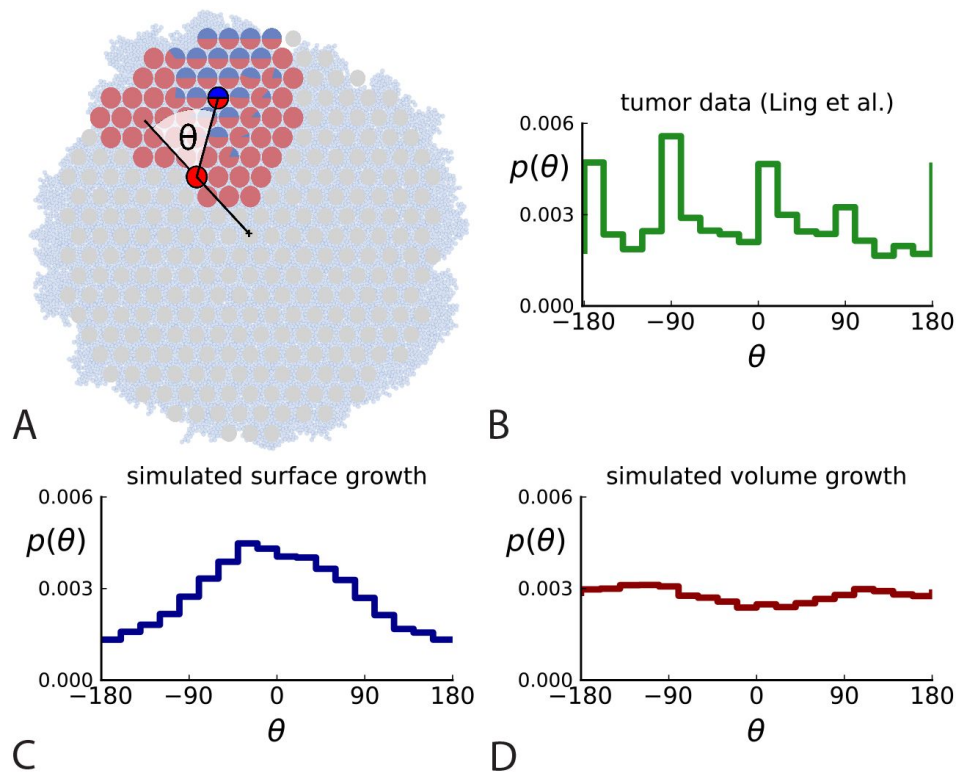


Figure 2

Relative position of mutants under different growth modes.

A The direction angle θ quantifies the direction of a new mutant clone relative to its parent clone. In this illustration, a new mutant clone indicated in blue appears and grows radially outward on a red parental background, resulting in an angle θ near zero. Each pair of cells indicated in red and blue contributes to the distribution of θ , with the statistical weight of each mutant clone adding to one (see SI S2). B The distribution of angles θ for different mutant clones found in the spatially-resolved data of Ling *et al.* [36]. The peaks at $\theta \approx -180^\circ$, -90° , 0° come from individual clones. The contributions from different clones to the distribution of angles is shown in Fig. SI S2. C and D show the corresponding distribution of angles for numerical simulations. Subfigure C shows simulations of surface growth, resulting in a distribution of direction angles with a pronounced maximum near zero. Under volume growth (D) a nearly flat distribution is seen. For C and D, simulations were run in three dimensions with a maximum population size of 40000 cells grown at division rate $b = 1$, a rate of cell death $d = 0.4$ and $d = 0.8$ for surface and volume growth respectively, and a whole-exome mutation rate $\mu = 0.3$ before taking a two-dimensional cross-section of 280 samples mimicking the sampling procedure in [36]. (The different death rates were chosen to make the extinction probabilities of the populations comparable for the two cases. Changing these rates did not affect the distributions of angles.)

Similarly, the dataset of Li *et al.* [24] with three-dimensional sampling also shows the signature of volume growth, both in the distribution of direction angles (SI S2.2) and in the number of mutations as a function of distance from the tumour centre, see SI S3.3. This result is at variance with a recent spatio-phylogenetic analysis performed by Lewinsohn *et al.* [41], which finds a signal of surface growth on the same data. In SI S3.4, we show that the findings of [41] are compatible with volume growth; on artificially generated volume-growth data sampled like in Li *et al.*, the algorithm used in [41] also returns a spurious signal of surface growth.

Spatial dispersion of cells

Another aspect where surface and volume growth lead to qualitatively different behaviour is the spatial dispersion of mutations. Cells carrying a particular mutation can in principle form a tightly spaced colony or be dispersed throughout the tumour mass, see Fig. 3A. It turns out this dispersion of mutants differs between surface and volume growth.

To quantify the spatial dispersion of a mutation, we define the dispersion parameter σ : For a given mutation, we compute the average distance of all pairs of samples carrying that mutation and divide by the average distance these cells would have if they were arranged next to one another in a spherical shape (the tightest configuration possible). Values of σ much larger than one can arise when a mutation is scattered throughout the tumour, with many cells that do not carry that mutation located between cells that do. The dispersion σ can be estimated from mutation frequencies in the whole-exome sequencing data of [36], see SI S4 for details.

Fig. 3B shows a histogram of the dispersion parameter σ for 217 mutations in the whole-exome sequencing data of [36]. We find an average value of the dispersion of approximately 3.5, but the values for specific mutations can be much higher than that (up to 9). Fig. 3C and D show the corresponding distributions for simulations of volume and surface growth respectively. The distribution of dispersion parameters found under volume growth agrees with that in the empirical data (Kolmogorov-Smirnov statistics 1.13), whereas surface growth generates lower values of the dispersion parameter (Kolmogorov-Smirnov statistics 5.74 indicating a poorer match with the empirical data, see SI Section 4.3 for details). Also in the three-dimensional dataset of Li *et al.* [24] we find higher values of the dispersion parameter than expected under surface growth (SI S4.4). However, the larger distances between samples in 3 dimensions, compared to the samples in a single plane, turn out to limit the accuracy with which the dispersion parameters can be computed. A higher number of samples would be needed to make the distances between samples comparable between the two and three-dimensionally sampled data.

Volume growth offers a simple mechanism causing the dispersion of mutations: Tumour cells displace one another when new cells are born and grow. In this way, two cells which are initially close to one another and share a mutation can be pushed apart by the birth and growth of adjacent cells that do not share the mutation. The further apart a pair of cells becomes, the higher the probability that they will be moved even further apart due to the motion of the increasing number of cells between them. This amplification of initially small distances leads to an instability that can generate the large values of the dispersion σ found both in the data of [36] and in numerical simulations of the volume growth model (Fig. 3B and D, respectively). On the other hand, under surface growth, cells do not push each other apart, leading to lower values of the dispersion parameter. Also, high values of the dispersion parameter cannot arise spuriously due to positive selection, as selected clones would enter the population later than neutral ones (given the same final frequency), and thus do not have as much time to be moved within the tumour.

The dispersion of cells in a growing population via this pushing effect is well known in tissue mechanics [42, 39]. In the context of cancer, it provides a mechanism for the clonal mixing found by ultra-deep sequencing of multiple samples from a solid tumour [19, 43, 21], which leads to mutations that occur at high frequencies in some samples being present at low (but

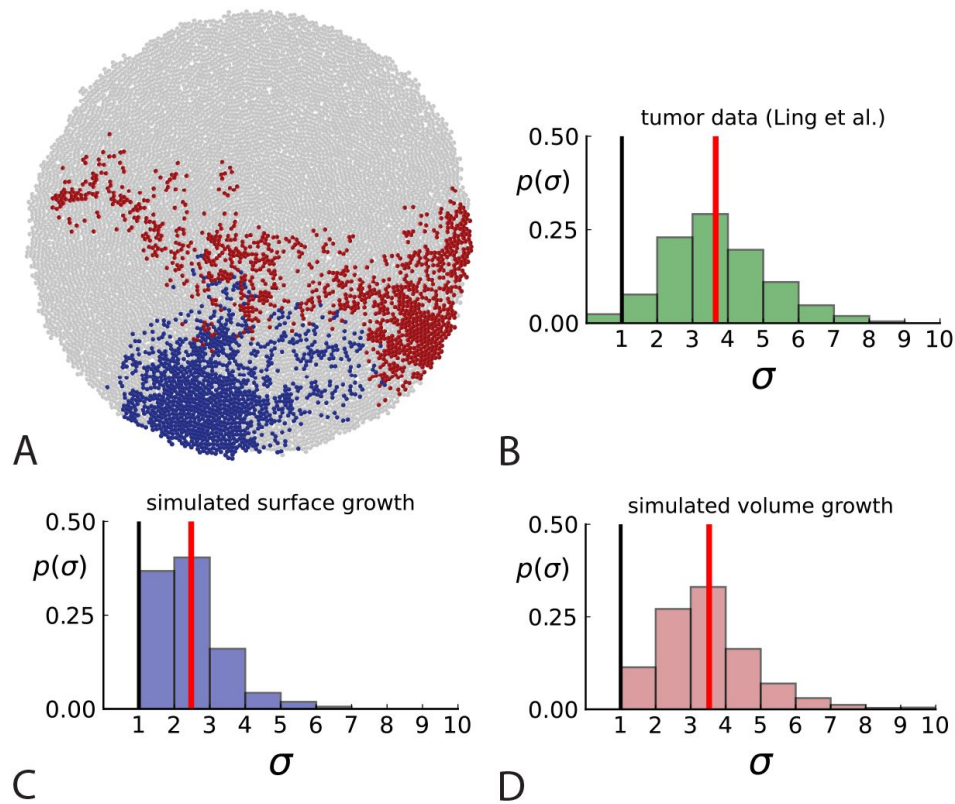


Figure 3

Dispersion of mutations within the tumour.

(A) Cells with a particular mutation can form tight spatial clusters within a tumour (simulated example here: mutant shown in blue), or they can be more widely dispersed (mutant shown in red). We quantify the dispersion of a mutation using the dispersion parameter σ , see text and SI S4. In this illustrative example, the blue mutation has a small dispersion parameter $\sigma = 1.3$, the red one has $\sigma = 2.5$. (B) Histogram of the dispersion parameters σ across 217 mutations in the whole-exome data of [36]. The red line indicates the mean of the histogram. (C) and (D) show the corresponding histograms for simulations of surface growth and volume growth, respectively. The simulations were run in 3D with populations grown up to 40000 cells before taking 23 evenly spaced samples from a 2D cross-section. Only mutations with a whole-tumour frequency larger than 1/40 were considered, mimicking the limited sequencing resolution in the Ling *et al.* data. Simulation parameters were division rate $b = 1$, mutation rate $\mu = 0.3$, and death rates $d = 0.4$ and $d = 0.8$ for surface growth and bulk growth, respectively.

non-zero) frequencies in other samples. Dispersion and clonal mixing thus arise as a by-product of volume growth. This is in contrast to surface growth, where a specific dispersal mechanism has been postulated to account for the observed spatial distribution of mutations [1].

To probe how this migration of cells between different regions of a tumour affects multi-region sampling, we look at the number of mutations detectable in a limited number of samples. Specifically, we consider the fraction of mutations present in a certain subset of samples (relative to the mutations appearing in all the samples taken together) and ask how this fraction is affected by migration. **Fig. 4A** shows the fraction of mutations detectable in a given number of samples randomly picked from the 23 whole-exome samples of [36]. One finds that typically about 75% of all the mutations can be detected with only 5 samples. We repeat this analysis but do not count mutations in a sample that occur with a higher frequency in some other sample. The rationale is that these mutations have arisen elsewhere (where they are present at higher frequency) and have migrated into the sample, where they are now present at a lower frequency. With such mutations removed, 5 samples typically only contain about one fifth of the mutations. This is in line with 5 samples representing only about one fifth of the 23 whole-exome samples. Migration of cells thus makes a single sample capture a far larger share of the genetic variability than expected from its share of the sampled tumour volume and provides an a-posteriori rationale for assessing the genetic variation found in a tumour on the basis of only a few samples.

Cell turnover

At low rates of cell death, tumours grow nearly at the rate at which tumour cells divide. Higher rates of cell death slow down the net growth of the population and lead to a constant turnover of cells. The relative rates of cell birth and death have been studied in cell cultures since the classic work of Steel [44] using thymidine labelling. Also in cell cultures, live cell imaging has been used to track individual cell divisions [45]. In the following, we use the genetic record in spatially resolved genomic data to probe the role of cell death *in vivo* during early tumour evolution.

Cell death removes cells, some of which have had offspring prior to death. This can lead to clones without extant parents (**Fig. 4B**), and the rate at which such ‘orphan clones’ arise can be used to infer the rate of cell death. We consider two metrics which quantify how frequently parental cells are removed from the population by cell death. The clone turnover gives the fraction of genotypes whose parental genotype is no longer extant in the tumour. The clade turnover gives the fraction of clades that coincide with their ancestral clade. (A clade is defined by a given mutation. Without backmutations, it comprises all individuals carrying this mutation.) These two metrics depend on the rate of cell death (relative to the birth rate) and on the mutation rate per cell division. Increasing the rate of cell death increases both clade and clone turnover, whereas increasing the mutation rate leads to a larger clone turnover only. These relationships have been determined analytically for a simple model of a growing population [46]. Inverting these relationships, the relative rate of cell death and the mutation rate can be determined from these metrics, see SI S5 and [46] for details and extensive tests.

Fig. 4C and **D** show the violin plots for the relative death rate and the mutation rate per cell division inferred in this way from the genotyped data of Ling *et al.* [36]. We find a rate of cell death nearly as high as the rate of birth (relative death rate 0.975 with 90% confidence interval [0.83, 0.995]). Hence during its early stage, the tumour was balanced nearly perfectly between growth and extinction. The inferred exome-wide mutation rate μ per cell division is 0.015 with 90% confidence interval [0.002, 0.1]. This result is at variance with a recent analysis based on the distribution of mutational distances between samples. Werner *et al.* [47] find low rates of cell death across many cancer data sets. In SI S6 we show that this discrepancy arises because different rates of cell deaths are compatible with a given distribution of pairwise distances between samples.

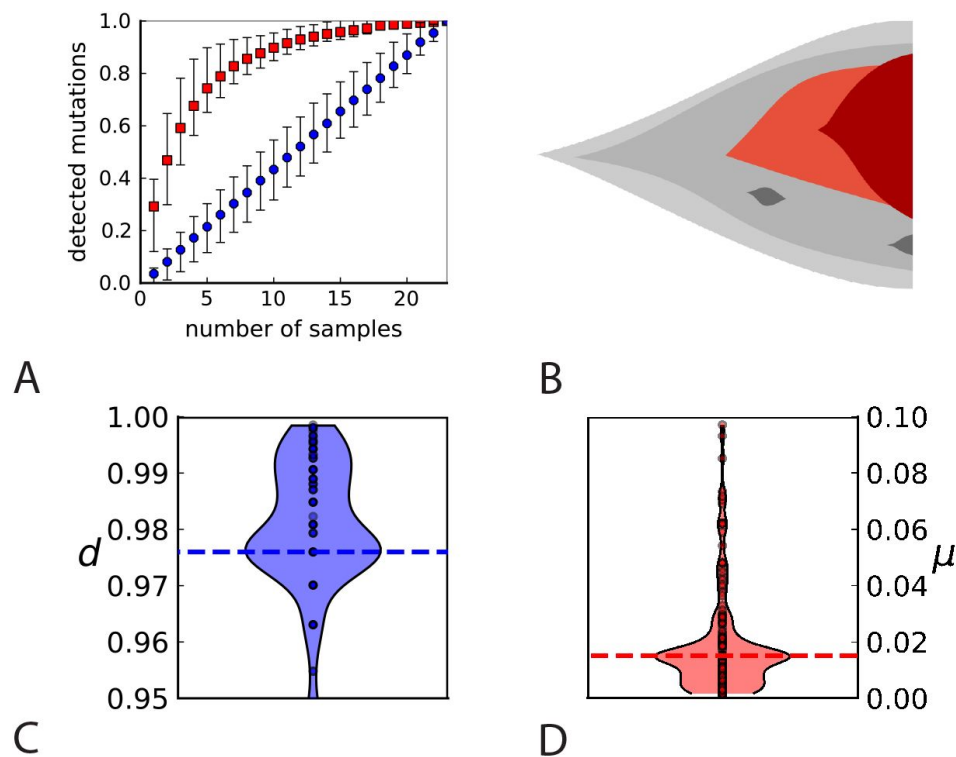


Figure 4

Rate of cell death and the mutation rate.

(A) We ask how a limited number of samples identify mutations. We pick a subset of the whole-exome sequenced samples of Ling *et al.* [36] and plot the fraction of mutations present at least in one of these samples against the number of samples in the subset (red symbols, fractions are relative to the number of mutations present in at least one of the 23 samples. Mutations must be supported by at least 5 reads at a coverage of at least 150). The procedure is repeated (blue symbols) with those mutations removed that occur in some other sample with a higher frequency than the frequencies with which the mutation occurs in the subset of samples. Error bars indicate the range of the 95-percentile. (B) The schematic Muller plot shows how cell death leads to the loss of clones, some of which have extant offspring. Time runs on the horizontal axis, the vertical shows a number of different clones indicated by different colours. In the example shown here, the clone shown in light red becomes extinct, leaving behind its darker-shaded offspring clone with no parental clone. The rate of this loss of parental clones depends on the rate of cell death, and can be used for inference, see text. (C) shows the inferred rate of cell death and (D) the inferred rate of mutation per generation. The violin plots show how the inferred values vary when subsampling different fractions of all mutations and scaling the inferred mutation rate correspondingly, with the dashed lines indicating the mean inferred values. (The fraction of mutations sampled ranges from 0.5 to 0.9, with the results shown separately in SI 9.)

To obtain the mutation rate per nucleotide we need to divide μ by the effective genome size, that is the number of sites in the genome that could have acquired detectable mutations. For many sites only few or no reads are retrieved during sequencing, so dividing μ by the size of the whole exome would underestimate the mutation rate. Ling *et al.* apply a set of criteria to filter out sites with poor coverage ([36], SI): A site must have 1) a total coverage of more than 150 reads when summed over all samples, 2) at least one sample with a coverage of 10 reads or more, and 3) a coverage of 6 reads or more in the normal tissue sample. We find that 4.8×10^5 sites in the exome fulfill these criteria, and the resulting mutation rate estimate per nucleotide is 1.6×10^{-8} with 90% confidence interval $[2e-9, 1e-7]$. Comparing this result to a mutation rate estimate from cell cultures of healthy human fibroblasts of 2.7×10^{-9} [?], this would mean the tumour mutation rate was enhanced by a factor of 5. We note that the relatively small size of the tree likely affects these estimates, and that analyses of larger trees in the future may lead to more accurate results.

Mutational signatures and temporal heterogeneity

Somatic mutations in cancer tumours are caused by different mutational processes, which can be identified by their mutational signatures [48]. To investigate which processes were active in the tumours analysed here, we decomposed the mutational profiles of all three tumours into mutational signature components. **Figure 5** shows that the strongest signature in all tumours is the single-base-substitution signature 22 (SBS22). This signature has been associated with exposure to the exogenous mutagen aristolochic acid [49], a herbal component used in traditional Chinese medicine. (The cancer patients of the Ling *et al.* and Li *et al.* lived in China.) To ask if there is a pseudo-temporal heterogeneity in the mutational processes, we divided the mutations from each tumour into clonal (early) and subclonal (late) mutations. While the Li *et al.* tumours maintain similar mutational signatures over pseudo-time, we find that the exposure to aristolochic acid of the patient from Ling *et al.* appears to have ceased over the course of tumour evolution: while the mutational signature SBS22 dominates the clonal mutations, it is absent in the subclonal mutations, see **Fig. 5**. The probability of not observing subclonal SBS22 due to sampling noise (p -value) was estimated to be less than 10^{-4} , see SI S9.

Population size

The results on the mode of evolution, migration, and cell turnover characterize the early stages of the tumours analysed here. The reason is that in an exponentially growing population, late-stage mutations have frequencies too low to be detected: Under neutral evolution and neglecting fluctuations in clone sizes, the final frequency (in the entire population) of a mutation that arose when a tumour consisted of N cells is given by $f = 1/N$ [29]. A lower limit on the frequency f thus corresponds to an upper limit on the population size. In order to determine the population size the tumour had when the mutations we analysed here arose, we generalize this result to include the effects of spatial sampling and stochastic population dynamics, see SI S7. The higher the spatial resolution, the lower the whole-tumour frequency of a detectable mutation can be. For this reason, a high spatial resolution means that late-coming (low frequency) mutations can be detected. We find that the mutations detected in the samples with high spatial resolution (surface growth/volume growth analysis and genetic turnover) arose when the tumour consisted of about 80000 cells. The mutations in the whole-exome samples used to measure the dispersion of mutations arose when the tumour size was around 7000 cells.

Discussion

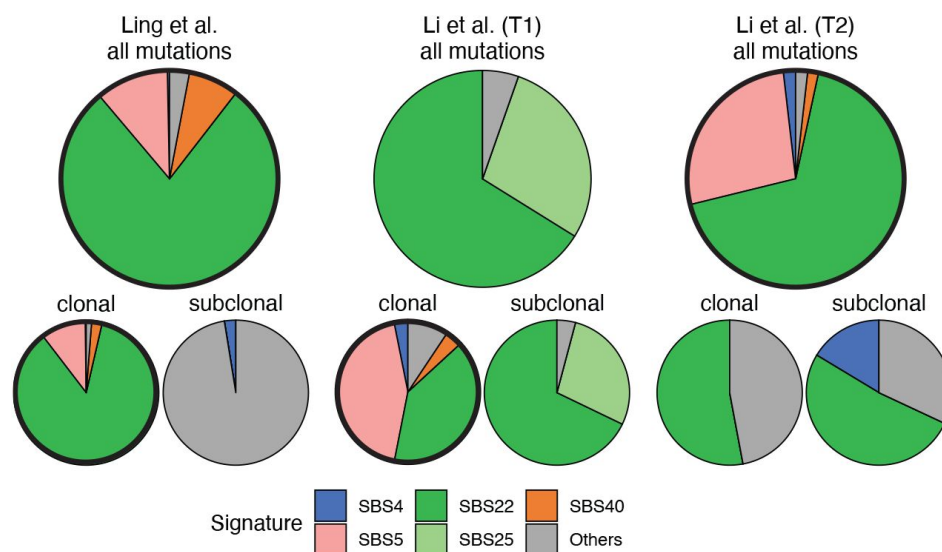


Figure 5

Mutational signature decomposition.

Relative weights of single-base substitution (SBS) mutational signatures [48] in all three tumours were derived with SigNet Refitter where possible (highlighted pie charts), otherwise with non-negative least squares [50] (Methods). (left) Ling et al., (centre) tumour T1 of Li et al., (right) tumour T2 of Li et al. Top: all mutations, bottom: mutations stratified by their clonality. Signature SBS22 is associated with exposure to aristolochic acid. In the Ling et al. data, this signature is prominent among clonal mutations, but absent in subclonal mutations. Also shown are signatures with relative weight larger than 1% and attributed to endogenous mutational processes (SBS5 and SBS40), tobacco smoking (SBS4) and a signature that resembles SBS22 (SBS25). All other signatures were combined into a single category ("Others").

Modes of tumour evolution

We have used genetic data taken from hepatocellular tumours at high spatial resolution [36, 24] to analyse the early evolution of a tumour *in vivo*. We are not the first to link spatial tumour sampling with evolution: on a larger spatial scale, phylogenetic comparison of samples from primary and metastatic tumours has elucidated the dynamics of metastasis [50, 51, 52, 53]. Analyses of multiple samples from the same tumour have established intra-tumour heterogeneity and found neutral evolution in many cases [3, 19, 36, 34, 21] (see [54] for a critique), as well as instances of selection [35, 55]. The mode of growth of a tumour affects the resulting genetic diversity of a tumour [25, 26].

Here, we have used the spatial information contained in data based on hundreds of samples from single tumours. We found that the tumour evolved under uniform volume growth; there was no evidence of spatial constraints leading to radially outward growth. We also found a substantial turnover of cells with a rate of cell death comparable to the rate of cell division.

The rate of cell death being comparable to that of cell division implies that the tumour grew much more slowly than suggested by the rate of cell division. Instead, in its early stage, the tumour was balanced precariously between continued growth to macroscopic size and shrinkage: A small increase in the rate of cell death, or a small decrease in the growth rate would have caused the small net growth rate to become negative, possibly even leading to the extinction of the tumour.

The high spatial resolution also allows to track changes in the mutational processes over evolutionary time far more easily than in bulk data. Based on the high-resolution data of Ling *et al.* [36] we found a drastic difference between the mutational signature of early and late mutations. In this data, the dominant mutational signature in early mutations is associated with an exogenous mutagen (aristolochic acid) found in herbal traditional Chinese medicine and thought to cause liver cancer [56]. This signature is absent in late mutations, compatible with a discontinued dose of the mutagen. In the data set of Li *et al.* [24], the same signature is dominant both in the early and late mutations, compatible with a continuous dose of the mutagen.

Mutation dispersion and sampling

We have found mutations in the tumour that are widely dispersed throughout the tumour; as a result they can have low frequencies in one particular sample but high frequencies in other parts of the tumour. We find the same level of dispersal of mutations in off-lattice simulations of growing tissues, where it is due to cells pushing each other out of the way as a part of the growth process. Hence, cells that were initially adjacent to each other and share a mutation can get pushed apart during the course of tumour growth and end up with a large distance between them. We found that this effect can be highly advantageous when assessing the mutations in a tumour from a limited number of samples: cells carrying a particular mutation can migrate into a spatial region that is later sampled. Mutations might thus be present at low frequency in a particular sample, because the cells carrying it migrated from afar, rather than because the mutation either arose late during the growth of the tumour [34] or was under negative selection [57].

Our results are based on three well-encapsulated hepatocellular tumours, resected and sampled in two and three dimensions [36, 24]. Potentially, different types of tumours or tumours at different stages may exhibit other growth modes. For instance, tumours that are not encapsulated might have a more heterogeneous environment, and evolve under surface growth. Also, the spatial and genomic resolution of any data set imposes limits on how late stage events can be observed. We estimate that the data analysed here capture the evolutionary dynamics up to 80000 cells. Increasing the genomic and spatial resolution will allow to identify mutations of lower frequency and thus characterize the *in-vivo* evolution beyond the early stages. Future techniques

to sequence resected tumours at higher spatial and genomic resolution than currently possible will allow to trace key later-stage evolutionary changes like angiogenesis, effects of the tumour microenvironment, or the development of genetic instability.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant SFB1310/2 - 325931972. We acknowledge support of the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI /10.13039/501100011033), and the Generalitat de Catalunya through the CERCA programme. We also acknowledge funding by the Spanish Ministry of Science and Innovation through grants PGC2018-100941-A-I00 and PID2021-128976NB-I00. This work was also partially funded by the FWF Austrian Science Fund (Erwin-Schrödinger postdoctoral fellowship, J4366). We thank Xuemei Lu and Chung-I Wu for discussions on the data sets [24] and [36], respectively, and Martin Peifer and Michael Lässig for discussions. Many thanks to Alison Feder and Nicola Müller for discussions on their SDevo algorithm.

References

- [1] Waclaw B., et al. (2015) **A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity** *Nature* **525**:261–264
- [2] Chkhaidze K., et al. (2019) **Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data** *PLoS Computational Biology* **15**
- [3] Gerlinger M., et al. (2012) **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing** *N Engl J Med* **366**:883–892
- [4] McGranahan N., Swanton C. (2015) **Biological and therapeutic impact of intratumor heterogeneity in cancer evolution** *Cancer cell* **27**:15–26
- [5] Takei Y., et al. (2021) **Integrated spatial genomics reveals global architecture of single nuclei** *Nature* **590**:344–350
- [6] Zhao T., et al. (2022) **Spatial genomics enables multi-modal study of clonal heterogeneity in tissues** *Nature* **601**:85–91
- [7] Lomakin A., et al. (2021) **Spatial genomics maps the structure, character and evolution of cancer clones** *bioRxiv* <https://doi.org/10.1101/2021.04.16.439912>
- [8] Shraiman B. I. (2005) **Mechanical feedback as a possible regulator of tissue growth** *Proceedings of the National Academy of Sciences* **102**:3318–3323
- [9] Montel F., et al. (2011) **Stress clamp experiments on multicellular tumor spheroids** *Physical Review Letters* **107**
- [10] Hallatschek O., Hersen P., Ramanathan S., Nelson D. R. (2007) **Genetic drift at expanding frontiers promotes gene segregation** *Proceedings of the National Academy of Sciences* **104**:19926–19930
- [11] Brú A., Albertos S., Subiza J. L., García-Asenjo J. L., Brú I. (2003) **The universal dynamics of tumor growth** *Biophysical Journal* **85**:2948–2961
- [12] Hoefflin R., et al. (2016) **Spatial niche formation but not malignant progression is a driving force for intratumoural heterogeneity** *Nature Communications* **7**:1–12
- [13] Greenspan H. (1972) **Models for the growth of a solid tumor by diffusion** *Studies in Applied Mathematics* **51**:317–340
- [14] Ward J. P., King J. (1997) **Mathematical modelling of avascular-tumour growth** *Mathematical Medicine and Biology: A Journal of the IMA* **14**:39–69
- [15] Baish J. W., Jain R. K. (2000) **Fractals and cancer** *Cancer Research* **60**:3683–3688
- [16] González-García I., Solé R.V., Costa J. (2002) **Metapopulation dynamics and spatial heterogeneity in cancer** *Proceedings of the National Academy of Sciences* **99**:13085–13089

- [17] Komarova N. L. (2006) **Spatial stochastic models for cancer initiation and progression** *Bulletin of Mathematical Biology* **68**:1573–1599
- [18] Sottoriva A., et al. (2010) **Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity** *Cancer research* **70**:46–56
- [19] Sottoriva A., et al. (2015) **A Big Bang model of human colorectal tumor growth** *Nature Genetics* **47**:209–216
- [20] Iwasaki W. M., Innan H. (2017) **Simulation framework for generating intratumor heterogeneity patterns in a cancer cell population** *PLoS ONE* **12**
- [21] Sun R., et al. (2017) **Between-region genetic divergence reflects the mode and tempo of tumor evolution** *Nature genetics* **49**:1015–1024
- [22] van der Heijden M., et al. (2019) **Spatiotemporal regulation of clonogenicity in colorectal cancer xenografts** *Proceedings of the National Academy of Sciences* **116**:6140–6145
- [23] Sinha S., Li X., Thirumalai D. (2022) **Statistical mechanical theory for spatio-temporal evolution of intra-tumor heterogeneity in cancers: Analysis of multiregion sequencing data** *bioRxiv* <https://doi.org/10.1101/2022.02.19.481164>
- [24] Li G., et al. (2022) **Evolution under spatially heterogeneous selection in solid tumors** *Molecular Biology and Evolution* **39**
- [25] Noble R., et al. (2022) **Spatial structure governs the mode of tumour evolution** *Nature ecology & evolution* **6**:207–217
- [26] Fu X., et al. (2022) **Spatial patterns of tumour growth impact clonal diversification in a computational model and the tracex renal study** *Nature ecology & evolution* **6**:88–102
- [27] Luebeck E. G., Moolgavkar S. H. (2002) **Multistage carcinogenesis and the incidence of colorectal cancer** *Proceedings of the National Academy of Sciences* **99**:15095–15100
- [28] Michor F., Iwasa Y., Nowak M. A. (2004) **Dynamics of cancer progression** *Nature Reviews Cancer* **4**:197–205
- [29] Durrett R. (2013) **Population genetics of neutral mutations in exponentially growing cancer cell populations** *The Annals of Applied Probability* **23**
- [30] Bozic I., et al. (2013) **Evolutionary dynamics of cancer in response to targeted combination therapy** *eLife* **2**
- [31] Foo J., Michor F. (2014) **Evolution of acquired resistance to anti-cancer therapy** *Journal of Theoretical Biology* **355**:10–20
- [32] Heyde A., Reiter J. G., Naxerova K., Nowak M. A. (2019) **Consecutive seeding and transfer of genetic diversity in metastasis** *Proceedings of the National Academy of Sciences* **116**:14129–14137
- [33] Attolini C. S.-O., et al. (2010) **A mathematical framework to determine the temporal sequence of somatic genetic events in cancer** *Proceedings of the National Academy of Sciences* **107**:17604–17609

- [34] Williams M. J., Werner B., Barnes C. P., Graham T. A., Sottoriva A. (2016) **Identification of neutral tumor evolution across cancer types** *Nature Genetics* **48**:238–244
- [35] Williams M. J., et al. (2018) **Quantification of subclonal selection in cancer from bulk sequencing data** *Nature Genetics* **50**:895–903
- [36] Ling S., et al. (2015) **Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution** *Proceedings of the National Academy of Sciences* **112**:E6496–E6505
- [37] Drasdo D., Höhme S. (2005) **A single-cell-based model of tumor growth in vitro: monolayers and spheroids** *Physical Biology* **2**
- [38] Van Liedekerke P., et al. (2019) **Quantitative cell-based model predicts mechanical stress response of growing tumor spheroids over various growth conditions and cell lines** *PLoS computational biology* **15**
- [39] Malmi-Kakkada A. N., Li X., Samanta H. S., Sinha S., Thirumalai D. (2018) **Cell growth rate dictates the onset of glass to fluidlike transition and long time superdiffusion in an evolving cell colony** *Physical Review X* **8**
- [40] Metzcar J., Wang Y., Heiland R., Macklin P. (2019) **A review of cell-based computational modeling in cancer biology** *JCO Clinical Cancer Informatics* **2**:1–13
- [41] Lewinsohn M. A., Bedford T., Müller N. F., Feder A. F. (2023) **State-dependent evolutionary models reveal modes of solid tumour growth** *Nature Ecology & Evolution* **7**:581–596
- [42] Ranft J., et al. (2010) **Fluidization of tissues by cell division and apoptosis** *Proceedings of the National Academy of Sciences* **107**:20863–20868
- [43] Suzuki Y., et al. (2017) **Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer** *Molecular oncology* **11**:124–139
- [44] Steel G. (1967) **Cell loss as a factor in the growth rate of human tumours** *European Journal of Cancer (1965)* **3**:381–387
- [45] Lorenzo C., et al. (2011) **Live cell division dynamics monitoring in 3d large spheroid tumor models using light sheet microscopy** *Cell division* **6**:1–8
- [46] Angaji A., Velling C., Berg J. (2021) **Stochastic clonal dynamics and genetic turnover in exponentially growing populations** *Journal of Statistical Mechanics: Theory and Experiment* **2021**
- [47] Werner B., et al. (2020) **Measuring single cell divisions in human tissues from multi-region sequencing data** *Nature Communications* **11**:1–9
- [48] Alexandrov L. B., et al. (2020) **The repertoire of mutational signatures in human cancer** *Nature* **578**:94–101
- [49] Hoang M. L., et al. (2013) **Mutational signature of aristolochic acid exposure as revealed by wholeexome sequencing** *Science translational medicine* **5**:197–197

- [50] Jones S., et al. (2008) **Comparative lesion sequencing provides insights into tumor evolution** *Proceedings of the National Academy of Sciences* **105**:4283–4288
- [51] Zhao Z.-M., et al. (2016) **Early and multiple origins of metastatic lineages within primary tumors** *Proceedings of the National Academy of Sciences* **113**:2140–2145
- [52] Alves J. M., Prado-Lopez S., Cameselle-Teijeiro J. M., Posada D. (2019) **Rapid evolution and biogeographic spread in a colorectal cancer** *Nature communications* **10**:1–7
- [53] Hu Z., Li Z., Ma Z., Curtis C. (2020) **Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases** *Nature genetics* **52**:701–708
- [54] Tarabichi M., et al. (2018) **Neutral tumor evolution?** *Nature genetics* **50**:1630–1633
- [55] Caravagna G., et al. (2018) **Detecting repeated cancer evolution from multi-region tumor sequencing data** *Nature methods* **15**:707–714
- [56] Ng A. W., et al. (2017) **Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia** *Science translational medicine* **9**
- [57] Weghorn D., Sunyaev S. (2017) **Bayesian inference of negative and positive selection in human cancers** *Nature genetics* **49**:1785–1788

Editors

Reviewing Editor

Jennifer Flegg

The University of Melbourne, Melbourne, Australia

Senior Editor

Wafik El-Deiry

Brown University, Providence, United States of America

Reviewer #1 (Public review):

Summary:

Arman Angaji and his team delved into the intricate world of tumor growth and evolution, utilizing a blend of computer simulations and real patient data from liver cancer.

Strengths:

Their analysis of how mutations and clones are distributed within tumors revealed an interesting finding: tumors don't just spread from their edges as previously believed. Instead, they expand both from within and the edges simultaneously, suggesting a unique growth mode. This mode naturally indicates that external forces may play a role in cancer cells dispersion within the tumor. Moreover, their research hints at an intriguing phenomenon - the high death rate of progenitor cells and extremely slow pace in growth in the initial phase of tumor expansion. Understanding this dynamic could significantly impact our comprehension of cancer development.

Weaknesses:

It's important to note, however, that this study relies on specific computer models, metrics derived from inferred clones, and a limited number of patient data. While the insights gained are promising, further investigation is essential to validate these findings. Nonetheless, this work opens up exciting avenues for comprehending the evolution of cancers.

Comments on revised submission:

The authors have effectively addressed my concerns. This revision is excellent.

<https://doi.org/10.7554/eLife.95338.2.sa2>

Reviewer #2 (Public review):

Summary:

The article uses a cell-based model to investigate how mutations and cells spread throughout a tumour. The paper uses published data and the proposed model to understand how growth and death mechanisms lead to the observed data. This work provides an insight into the early stages of tumour development. From the work provided here, the results are solid, showing a thorough analysis. The article is well written and presents a very suitable and rigorous analysis to describe the data. The authors did a particularly nice job of the discussion and decision of their "metrics of interest", though this is not the main aim of this work.

Strengths:

Due to the particularly nice and tractable cell-based model, the authors are able to perform a thorough analysis to compare the published data to that simulated with their model. They then used their computational model to investigate different growth mechanisms of volume growth and surface growth. With this approach, the authors are able to compare the metric of interest (here, the direction angle of a new mutant clone, the dispersion of mutants throughout the tumour) to quantify how the different growth models compare to the observed data. The authors have also used inference methods to identify model parameters based on the data observed. The authors performed a rigorous analysis and have chosen the metrics in an appropriate manner to compare the different growth mechanisms.

Context:

Improved mechanistic understanding into the early developmental stages of tumours will further assist in disease treatment and quantification. Understanding how readily and quickly a tumour is evolving is key to understanding how it will develop and progress. This work provides a solid example as to how this can be achieved with data alongside simulated models.

<https://doi.org/10.7554/eLife.95338.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public reports:

In the public reports there is only one point we would like to discuss. It concerns our use of a computational model to analyse spatial tumour growth. Citing from the eLife assessment, which reflects several comments of the referees:

The paper uses published data and a proposed cell-based model to understand how growth and death mechanisms lead to the observed data. This work provides an important insight into the early stages of tumour development. From the work provided here, the results are solid, showing a thorough analysis. However, the work has not fully specified the model, which can lead to some questions around the model's suitability.

The observables we use to determine the (i) growth mode and the (ii) dispersion of cells are model-independent. The method to determine the (iii) rate of cell death does not use a spatial model. Throughout, our computational model of spatial growth is not used to analyze data. Instead, it is used to check that the observables we use can actually discriminate between different growth modes given the limitations of the data. We have expanded the description of the computational model in the revised version, and have released our code on Github. However, the conclusions we reach do not rely on a computational model. Instead, where we estimate parameters, we use population dynamics as described in section S5. The other observables are parameter free and model-independent. We view this as a strength of our approach.

Recommendations for the authors:

Reviewer #1:

(1.1) In Figure 1, the data presented by Ling et al. demonstrate a distinctive “comb” pattern. While this pattern diverges from the conventional observations associated with simulated surface growth, it also differs from the simulated volume growth pattern. Is this discrepancy attributable to insufficient data? Alternatively, could the emergence of such a comb-like structure be feasible in scenarios featuring multiple growth centers, wherein clones congregate into spatial clusters?

We are unsure what you are referring to. One possibility is you refer to the honey-comb structure formed by the samples of the Ling et al. data shown in Fig. 1A of the main text. This is an artefact arising from the cutting of the histological cut into four quadrants, see Fig. S1 in the SI of Ling et al. The perceived horizontal and vertical “white lines” in our Fig. 1A stems from the lack of samples near the edges of these quadrants. We have added this information to the figure caption.

An alternative is you are referring to the peaks in Fig 2A of the main text. The three of these peaks indeed stem from individual clones. We have placed additional figures in the SI (S2 B and S2 C) to disentangle the contribution from different clones. The peaks have a simple explanation: each clone contributes the same weight to the histogram. If a clone only has few offspring, this statistical weight is concentrated on a few angles only, see SI Figure S2 B.

(1.2) I am not sure why there are two sections about “Methods” in the main text: Line 50 as well as Line 293. Furthermore, the methods outlined in the main paper lack the essential details necessary for readers to navigate through the critical aspects of their analysis. While these details are provided in the Supplementary Information, they are not adequately referenced within the methods section of the main text. I would recommend that the authors revise the method sections of the main text to include pertinent descriptions of key concepts or innovations, while also directing readers to the corresponding supplementary method section for further elucidation.

We have merged the Section “Materials and Methods” at the end of the main text with the SI description of the data in SI 4.2 and placed a reference to this material in the main body.

(1.3) The impact of the particular push method (proposed in the model) on the resultant spatial arrangement of clones remains unclear. For instance, it's conceivable that

employing a different pushing method (for example, with more strict constraints on direction) could yield a varied pattern of spatial diversity. Furthermore, there is ambiguity regarding the criteria for determining the sequence of the queue housing overlapping cells.

Regarding the off-lattice dynamics we use, there are indeed many variants one could use. In nonexhaustive trials, we found that the details of the off-lattice dynamics did not affect the results. The reason may be that at each computational step, each cell only moves a very small amount, and differences in the dynamics tend to average out over time.

We deliberately do not give constraints on the direction. Such constraints emerge in lattice-based models (when preferred directions arise from the lattice symmetry), but these are artifacts of the lattice.

At cell division the offspring is placed in a random direction next to the parent regardless of whether this introduces an overlap. Cells then push each other along the axis connecting their two centers of mass – unlike in lattice based models a sequence of pushes does not propagate through the tumor straight away but sets off of a cascade of pushes. Equal pushing of two cells (i.e. two initial displacements as opposed to pushing one of the two) results in the same patterns of directed, low dispersion surface and undirected, high dispersion volume growth but is much harder computationally as it reintroduces overlaps that have been resolved in the previous step.

We have rewritten the description of the pushing queue in the SI Section 1. The choice of the pushing sequence is somewhat arbitrary but we found that it also has no noticeable effect on the growth mode. Maybe putting it in contrast to depth-first approaches helps to illustrate this: We tried two queueing schemes for iterating through overlapping cells, width-first and depth-first. In both cases, we begin by scanning a given cell's (the root's) neighborhood for overlaps and shuffle the list of overlapping neighbours. In a width-first approach we then add this list to the queue. Subsequent iterations append their lists of overlapping cells to the queue, such that we always resolve overlaps within the neighborhood of the root first. A depth-first approach follows a sequence of pushes by immediately checking a pushed cell's neighborhood for new overlaps and adding these to the front of the queue (which works more like a stack then). This can be efficiently implemented by recursion but has no noticeable performance advantage and results in the same patterns of directed, low dispersion surface and undirected, high dispersion volume growth. In our opinion the width-first approach of first resolving overlaps in the immediate neighborhood is more intuitive, which is why we adopted it for our simulation model.

(1.4) For the example presented in S5.1, how can the author identify from genomic data that mutation 3 does not replace its ancestral clade mutation 2? In other words, if mutation 2, 3 and 4 are linked meaning clone 4 survives but 2 and 3 dies, how does one know if clone 3 dies before clone 2? I understand that this is a conceptual example, but if one cannot identify this situation from the real data, how can the clade turnover be computed?

Thank you for this comment, which points to an error of ours in the turnover example of the SI: Clade 3 does in fact replace 2 and contributes to the turnover! (The algorithm correctly annotated clade 3 as orphaned and computes a turnover of 3/15 for this example). We have corrected this.

In this example, it does not matter for the clade turnover whether clone 3 dies before clone 2. As long as its ancestor (clone 2) becomes extinct it adds to the clade turnover. The term “replaces” applies to the clade of 3 which has a surviving subclone and thereby eventually

replaces clade 2. The clade turnover is solely based on the presence of the mutations (which define their clade) and not on the individual clones.

(1.5) After reviewing reference 24 (Li et al.), I noticed that the assertions made therein contradict the findings presented in S3 (Mutation Density on Rings). Specifically, Li et al. state that “peripheral regions not only accumulated more mutations, but also contained more changes in genes related to cell proliferation and cell cycle function” (Page 6) and “Phylogenetic trees show that branch lengths vary greatly with the long-branched subclones tending to occur in peripheral regions” (Page 4). However, upon re-analysis of their data, the authors demonstrated a decrease in mutation density near the surface. It is crucial to comprehend the underlying cause of such a disparity.

The reason for this disparity is the way Li et al. labelled samples as belonging to peripheral or central regions of the tumour. We have added a new figure in the SI to show this: Fig. S14 shows the number of mutations found in samples of Li et al. against their distances from the centre, along with the classification of samples as center/periphery given in Li et al. In the case of tumor T1, the classification of a sample in reference Li et al. does not agree with the distance from the center: samples classified as core are often more distant from the center than those classified as peripheral. Furthermore, Lewinsohn et al. (see below) show in their Fig. 5 that samples classified as ‘center’ by Li all fall into a single clade, and we believe this affects all results derived from this classification. For this reason, we do not consider the classification in reference 24 (Li et al.) further. We now briefly discuss this in Section S3.3.

(1.6) The authors consider coinciding mutations to occur when offspring clades align with an ancestral clade. Nevertheless, since multiple mutations can arise simultaneously in a single generation (such as kataegis), it becomes essential to discern its impact on clade turnover and, consequently, the estimation of d/b.

The mutational signatures found here show no sign of kataegis. Also, the number of polymorphic sites in the whole-exome data is small and the mutations are uniformly spread across the exome. The point is well taken, however, the method requires single mutations per generation. In practice, this can be achieved by subsampling a random part of the genome or exome (see [45]). We tested this point by processing the data from only a fraction of the exome; this did not change the results. In particular, Figure S30 shows the turnover-based inference for different subsampling rates L of the Ling et al. data. Subsampling of sites reduces the exome-wide mutation rate, the inferred rate scales linearly with L , as expected.

(1.7) I could not understand Step 2 in Section S2.1, an illustration may be helpful.

We have added figure S2 explaining the directional angle algorithm to Section S2.1 in the supplementary information.

(1.8) Figure S2, does a large ρ_{hoc} lead to volume growth rather than surface growth, not the other way around?

Thank you for catching this mix-up!

Reviewer #2

I do have a few minor comments/questions, but I am confident the authors will be able to address them appropriately.

(2.1) Line 56: I am not sure what the units of “average read depth 74X” is in terms of SI units?

This number gives the number of sequence reads covering a particular nucleotide and is dimensionless. We have added this information.

(2.2) Lines 63 - 68: I am unsure what is meant by the terms “T1 of ca.” and “T2 of ca.”. Can these also be explained/defined please?

These refer to the approximate (circa) diameters of tumor 1 and tumor 2 in the data by Li et al. We have expanded the abbreviations.

(2.3) Line 69: I would like to see a more extensive description of the cell-based model here in the main text, such as how do the cells move. Moreover, do cells have a finite reach in space, do they have a volume/area?

We have expanded the model description in the main body of the paper and placed information there that previously was only in the SI.

(2.4) Line 76: You have said cells can “push” one another in your model. Do they also “pull” one another? Cell adhesion is known to contribute to tumour integrity - so this seems important for a model of this nature.

We have not implemented adhesive forces between pairs of cells so far. This would cause a higher pressure under cell growth (which can have important physiological consequences). However, the hard potential enforcing a distance between adjacent cells would still lead to cells pushing each other apart under population growth, so we expect to see the dispersion effect we discuss even when there is adhesion.

(2.5) Line 80-81: “due to lack of nutrient”. Is nutrient included in this work? It is my understanding it is not. No problem if so, it is just that this line makes it seem like it is and important. If it is not, the authors should mention this in the same sentence.

Thank you for pointing out this source of misunderstanding, your understanding is correct and we have modified the text to remove the ambiguity.

(2.6) Line 94-95: Since you are interested in tissue growth, recent work has indicated how the cell boundary (and therefore tissue boundary) description influences growth. Please also be sure to indicate this when you describe the model.

We presume you refer to the recent paper by Lewinsohn et al. (Nature Ecology and Evolution, 2023), which reports a phylogenetic analysis based on the Li et al. data. Lewinsohn et al. find that cells near the tumour boundary grow significantly faster than those in the tumour's core. This is at variance with what we find; we were not aware of this paper at the time of submission. We now refer to this paper in the main text, and also have included a new section S3.4 in the SI accounting for this discrepancy. If you refer to a different paper, please let us know.

Briefly, we repeat the analysis of Lewinsohn et al., using their algorithm on artificial data generated by our model under volume growth. Samples were placed precisely like they were placed in the tumor analyzed by Li et al. We find that, even though the data was generated by volume growth, the algorithm of Lewinsohn et al. finds a signal of surface growth, in many cases even stronger compared to the signal which Lewinsohn et al. find in the empirical data. We have added subsection S3.4 with new figure S15 in the Supplementary Information.

(2.7) Line 107: “thus no evidence for enhanced cell growth near the edge of the tumour”. It is unclear to me how this tells us information relative to the tumour edge. It seems to

me this is an artifact that at the edge of the tumour, there are less cells to compare with? Could you please expand on this a bit?

The direction angles tell us if new mutations arise predominantly radially outwards. With this observable, surface growth would lead to a non-uniform distribution of these angles even if we restrict the analysis to samples from the interior of the tumor (which, under surface growth, was once near the surface). So the effect is not linked to fewer cells for comparison. Also, we have checked the direction angles in simulations under different growth modes with the samples placed in the same way as in the data (see Figs. S3 and S4 right panels). We have expanded the text in the main text, section Results accordingly.

(2.8) I really enjoyed the clear explanation between lines 119 and 122 regarding cell dispersion!

Thank you!

(2.9) Figure 2B: Since you are looking at a periodic feature in theta, I would have expected the distribution to be periodic too, and therefore equal at $\theta = -180 = 180$. Can you explain why it is different, please? Interestingly, you simulated data does seem to obey this!

The distribution of theta is periodic but the binning and midpoints of bins were chosen badly. We have replotted the diagram with bin boundaries that handle the edge-points $-180/180$ correctly. Thank you for pointing this out.

(2.10) Figure 3B: This plot does not have a title. Also, what do the red vertical lines in plots 3B, 3C and 3D indicate?

We have added the title. The red lines indicate the expectation values of the distributions.

(2.11) Figure 4: I am unsure how to read the plot in 4B. Also, what does the y-axis represent in 4C and 4D?

We have added explanations for 4B and have placed the labels for 4C and 4D in the correct position on the y-axes.

(2.12) Lines 194-199: you discuss your inferred parameters here, but you do not indicate how you inferred these parameters. May you please briefly mention how you inferred these, please?

These were inferred using the turnover method explained in the paragraph above, we have expanded the information. A full account is given in the SI Section S5.

(2.13) Line 258-260: "... mutagen (aristolochic acid) found in herbal traditional Chinese medicine and thought to cause liver cancer." I do not see what this sentence adds to the work. Could you please be clearer with the claim you are making here?

Mutational signatures allow to infer underlying mutational processes. The strongest signature found in the data is associated with a mutagen that has in the past been used in traditional Chinese medicines. The patients from whom the tumours were biopsied were from China, so past exposure to this potent mutagen is possible. We are not making a big claim here, the mutational signature of aristolochic acid and its cancerogenic nature has been well studied and is referenced here. The result is interesting in our context because in one of the datasets (Li et al.) the signature is present in early (clonal) mutations but absent in later

ones, allowing to make inferences from present data on the past. We have added the information that the patients were from China.

(2.14) In your Supplementary Information, S1, I believe your summation should not be over i , as you state in the following it is over cells within 7 cell radii. Please fix this by possibly defining a set which are those within 7 cell radii.

We have done this.

<https://doi.org/10.7554/eLife.95338.2.sa0>