High-density sampling reveals volume growth in human tumours

Arman Angaji[1], Michel Owusu[2], Christoph Velling[*1], Nicola Dick[2,3], Donate Weghorn[†2,3], and Johannes Berg[†1]

[1]Institute for Biological Physics, University of Cologne, Zülpicher Str. 77, 50937 Cologne, Germany
[2]Centre for Genomic Regulation, C/ del Dr. Aiguader, 88, 08003 Barcelona, Spain
[3]Universitat Pompeu Fabra (UPF), Barcelona, Spain

In growing cell populations such as cancer, mutations can serve as markers that allow tracking the past evolution from current samples. The genomic analysis 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.

[*]current address: Kishony lab, Faculty of Biology, Technion, Haifa 3200003, Israel
[†]correspondence to dweghorn@crg.eu and bergj@uni-koeln.de

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, 1, 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, 1, 19, 20, 21, 22] and has been used to analyze multi-region tumour sequencing data [19, 21, 2, 23, 24].

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 [25, 26, 27, 28, 29, 30]. They also form the basis of almost all population-genetic approaches to analyzing tumour data [31, 32, 33].

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 [34, 24]. From these snapshots of late-stage 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 [34], and once from a three-dimensional microsampling [24].

In Ling *et al.* [34], 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 [34] determined by cell counting. 23 of these samples were subjected to whole-exome sequencing (average read depth $74X$), and loci with mutations found in these 23 samples were probed by genotyping (by Sequenom and Sanger sequencing) in all 285 samples [34]. 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 Materials and Methods.

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 of ca. 20 mm diameter and 160 samples from 6 sections of tumour T2 of ca. 15 mm diameter. Each sample consisted of ca. 3200 cells. 16 (9) samples from T1 (T2) were whole-exome sequenced and the remainder genotyped, analogously to the Ling *et al.* data described above.

Computational model of volume and surface growth. We use an off-lattice cell-based model, well-known from simulations of tissue mechanics [35, 36, 37, 38], 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 point in 3d space, as well as a genome. Each cell division introduces a new point, and each cell death removes a point. Cells are not allowed to overlap (implemented by a

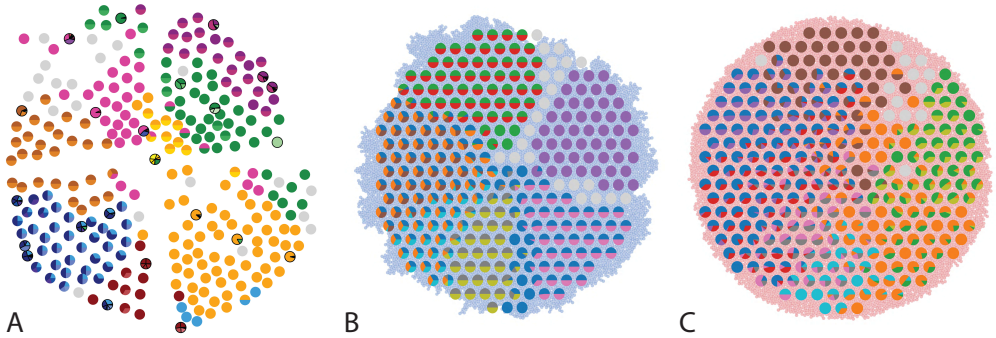

Figure 1: Multi-region sampling of a hepatocellular tumour and cell-based simulations. A shows the spatially resolved sequencing data of 285 samples of a hepatocellular carcinoma analyzed by Ling *et al.* [34]. Each sample is indicated by a small pie chart in which colors indicate specific mutations, 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. 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.

minimal distance between points) 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. 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 a lack of nutrients. We encode such a potential spatial effect 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 and its implementation, see SI S1.

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. Crucially, the distribution of direction angles is 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.* [34], and compare it to simulations of surface growth and of volume growth in Fig. 2C and D, respectively. The empirical data clearly show a uniform distribution of the direction angle, and thus no evidence for enhanced cell growth near the edge of the tumour. In a numerical simulation of surface growth, on the other hand, the histogram of direction angles shows a pronounced maximum at $\theta = 0$ produced by the radial outgrowth of clones.

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. S6 in SI S3).

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) and in the number of mutations as a function of distance from the tumour centre (SI S3).

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 [34], 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 [34] (see SI S4.2 for filters used to select variants). We find an average value of

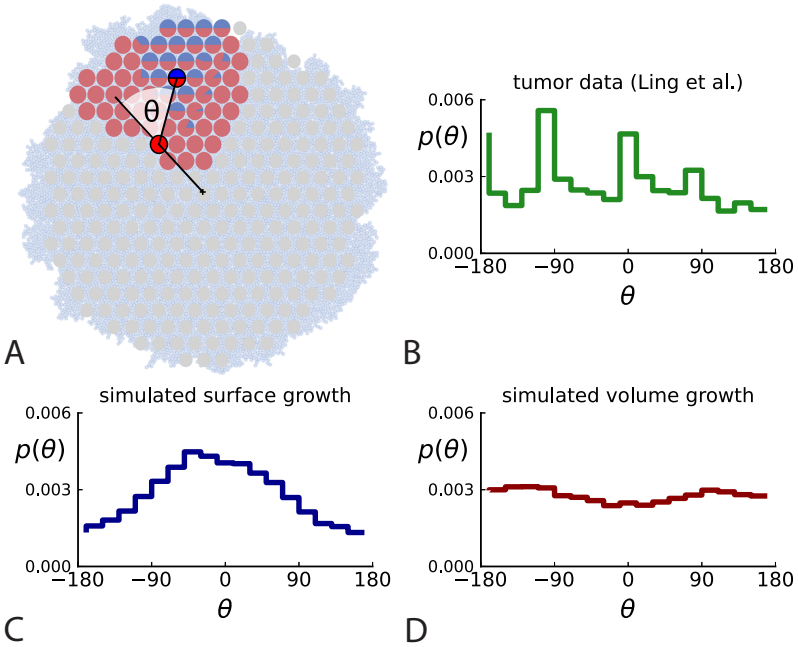

Figure 2: Relative position of mutants under different growth modes. A The direction angle θ quantifies the direction of a new mutant clone relative to its parent clone. In this illustration, a new mutant clone indicated in blue appears and grows radially outward on a red parental background, resulting 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.* [34]. 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 [34]. (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.)

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 Figure S15). 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 S4.3). 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 [34] 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 process is well known in tissue mechanics [39, 37]. 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, 40, 21], which leads to mutations that occur at high frequencies in some samples being present at low (but 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 [34]. 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.

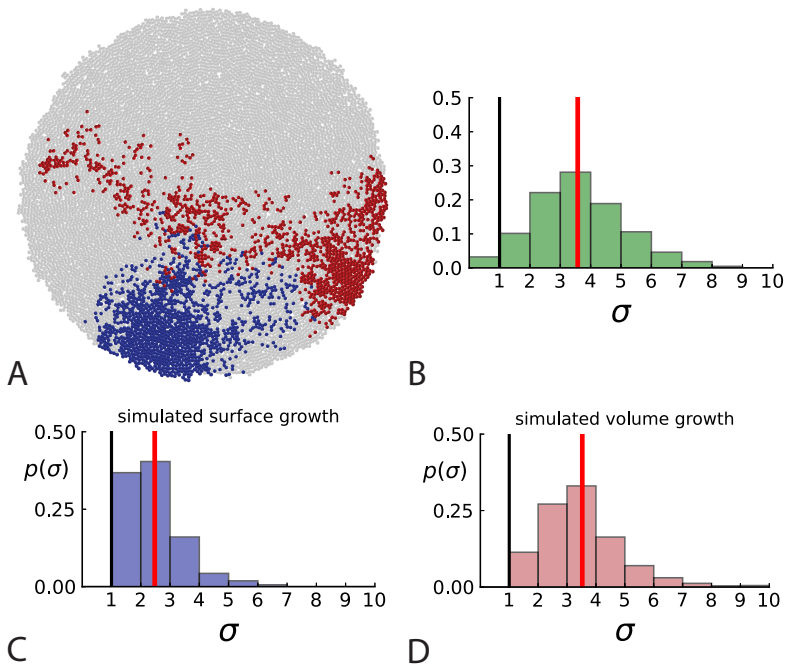

Figure 3: Dispersion of mutations within the tumour. (A) Cells with a particular mutation can form tight spatial clusters within a tumour (simulated example here: mutant shown in blue), or they can be more widely dispersed (mutant shown in red). We quantify the dispersion of a mutation using the dispersion 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 [34]. (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.

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 [41] using thymidine labelling. Also in cell cultures, live cell imaging has been used to track individual cell divisions [42]. 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.

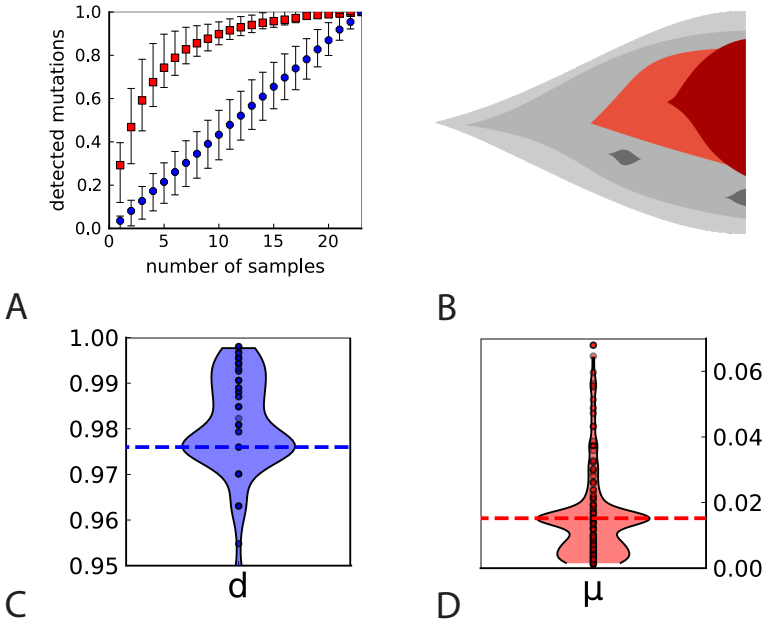

Figure 4: Rate of cell death and the mutation rate. (A) We ask how a limited number of samples identify mutations. We pick a subset of the whole-exome sequenced samples of Ling *et al.* [34] and plot the fraction of mutations present at least in one of these samples against the number of samples in the 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. In the example shown here, the clone 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.)

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. Conversely, the relative rate of cell death and the mutation rate can be determined from these metrics, see SI S5 and [43] 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 from the genotyped data of Ling *et al.* [34]. 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 $[.002, 0.1]$.

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 ([34], 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 per nucleotide is 3×10^{-8} with 90% confidence interval $[4e-9, 2e-7]$. For details, see SI S5 and S6, where we also compare our approach to the analysis of mutational distances between samples recently performed by Werner *et al.* [44].

Mutational signatures and temporal heterogeneity Somatic mutations in cancer tumours are caused by different mutational processes, which can be identified by their mutational signatures [45]. 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 [46], a herbal component used in traditional Chinese medicine. 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 $f = 1/N$ [27]. 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

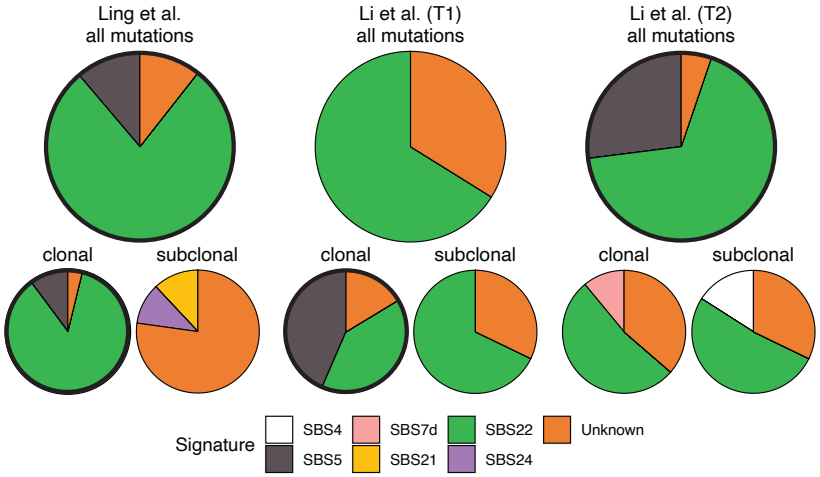

Figure 5: Mutational signature decomposition. Relative weights of single-base substitution (SBS) mutational signatures [45] in all three tumours were derived with the in-house method SigNet [47] where possible (highlighted pie charts), otherwise with non-negative least squares (Methods). (left) Ling *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. SBS5 and SBS40 are endogenous mutational processes. Signatures with non-positive weights were combined into a single category (Unknown).

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

Modes of tumour evolution We have used genetic data taken from a hepatocellular tumour at high spatial resolution [34] 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 [48, 49, 50, 51]. Analyses of multiple samples from the same tumour have established intra-tumour heterogeneity and found neutral evolution in many cases [3, 19, 34, 32, 21] (see [52] for a critique), as well as instances of selection [33, 53]. 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 an enhanced rate of growth near the surface of the tumour. 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.* [34] 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 [54]. 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 [32] or was under negative selection [55].

Our results are based on three well-encapsulated hepatocellular tumours, resected and sampled in two and three dimensions [34, 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.

Acknowledgment This work was supported by the DFG SFB 1310. 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. 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 [34], respectively, and Martin Peifer and Michael Lässig for discussions.

Materials and Methods

Alignment and variant calling

Fastq files of the sequences were downloaded from `https://bigd.big.ac.cn/search/?dbId=gsa&q=PRJCA000091&page=2` [34]. Fastq files were processed using our in-house mapping pipeline based on GATK4 (version 4.1.0.0 from DockerHUB). We used GRCh37.p13 as the reference genome and gencode.v19 as the annotation file. Singletons, and reads with less than 70 bp were discarded. Additionally, we excluded known variants in the human population (taken from https://data.broadinstitute.org and ftp://ftp.ncbi.nih.gov). We used the interval bed file S02972011 from SureSelect all human exon version 3. Variant calling and initial filtering was performed with Mutect2 (GATK), using the joint variant caller mode and applying the optional filter.

Decomposition of mutational profiles

We decomposed the mutational profiles of the tumours from the two studies using our in-house method SigNet [47]. SigNet is based on an artificial neural network and predicts the weights of the COSMIC v3 mutational signature catalog [45] in individual samples. If the sample profile does not correspond to the training data, it is decomposed with non-negative least squares. SBS1, SBS5 and SBS40 are associated with endogenous mutational processes. Since the samples that contain SBS25 were decomposed with a non-negative least squares approach (as they were classified as out-of-distribution by SigNet), it is likely that SBS25 is an erroneous assignment derived from SBS5 and SBS22 mutations [45]. The high number of signatures classified as Unknown in the separate clonal and subclonal mutations in T2 also suggest that the non-negative least squares approach used there leads to some spurious assignments. An example is SBS7d associated with UV radiation.

References

[1] Waclaw, B. *et al.* A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525**, 261–264 (2015).

[2] Chkhaidze, K. *et al.* Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. *PLoS Computational Biology* **15**, e1007243 (2019).

[3] Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl j Med* **366**, 883–892 (2012).

[4] McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer cell* **27**, 15–26 (2015).

[5] Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).

[6] Zhao, T. *et al.* Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* **601**, 85–91 (2022).

[7] Lomakin, A. *et al.* Spatial genomics maps the structure, character and evolution of cancer clones. *BioRxiv* (2021). URL `https://www.biorxiv.org/content/10.1101/2021.04.16.439912v1`.

[8] Shraiman, B. I. Mechanical feedback as a possible regulator of tissue growth. *Proceedings of the National Academy of Sciences* **102**, 3318–3323 (2005).

[9] Montel, F. *et al.* Stress clamp experiments on multicellular tumor spheroids. *Physical Review Letters* **107**, 188102 (2011).

[10] Hallatschek, O., Hersen, P., Ramanathan, S. & Nelson, D. R. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences* **104**, 19926–19930 (2007).

[11] Brú, A., Albertos, S., Subiza, J. L., García-Asenjo, J. L. & Brú, I. The universal dynamics of tumor growth. *Biophysical Journal* **85**, 2948–2961 (2003).

[12] Hoefflin, R. *et al.* Spatial niche formation but not malignant progression is a driving force for intratumoural heterogeneity. *Nature Communications* **7**, 1–12 (2016).

[13] Greenspan, H. Models for the growth of a solid tumor by diffusion. *Studies in Applied Mathematics* **51**, 317–340 (1972).

[14] Ward, J. P. & King, J. Mathematical modelling of avascular-tumour growth. *Mathematical Medicine and Biology: A Journal of the IMA* **14**, 39–69 (1997).

[15] Baish, J. W. & Jain, R. K. Fractals and cancer. *Cancer Research* **60**, 3683–3688 (2000).

[16] González-García, I., Solé, R. V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *Proceedings of the National Academy of Sciences* **99**, 13085–13089 (2002).

[17] Komarova, N. L. Spatial stochastic models for cancer initiation and progression. *Bulletin of Mathematical Biology* **68**, 1573–1599 (2006).

[18] Sottoriva, A. *et al.* Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer research* **70**, 46–56 (2010).

[19] Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nature Genetics* **47**, 209–216 (2015).

[20] Iwasaki, W. M. & Innan, H. Simulation framework for generating intratumor heterogeneity patterns in a cancer cell population. *PLoS ONE* **12**, e0184229 (2017).

[21] Sun, R. *et al.* Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature genetics* **49**, 1015–1024 (2017).

[22] van der Heijden, M. *et al.* Spatiotemporal regulation of clonogenicity in colorectal cancer xenografts. *Proceedings of the National Academy of Sciences* **116**, 6140–6145 (2019).

[23] Sinha, S., Li, X. & Thirumalai, D. Statistical mechanical theory for spatio-temporal evolution of intra-tumor heterogeneity in cancers: Analysis of multiregion sequencing data (2022). URL `https://www.biorxiv.org/content/10.1101/2022.02.19.481164v1`.

[24] Li, G. *et al.* Evolution under spatially heterogeneous selection in solid tumors. *Molecular Biology and Evolution* **39**, msab335 (2022).

[25] Luebeck, E. G. & Moolgavkar, S. H. Multistage carcinogenesis and the incidence of colorectal cancer. *Proceedings of the National Academy of Sciences* **99**, 15095–15100 (2002).

[26] Michor, F., Iwasa, Y. & Nowak, M. A. Dynamics of cancer progression. *Nature Reviews Cancer* **4**, 197–205 (2004).

[27] Durrett, R. Population genetics of neutral mutations in exponentially growing cancer cell populations. *The Annals of Applied Probability* **23**, 230 (2013).

[28] Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* **2**, e00747 (2013).

[29] Foo, J. & Michor, F. Evolution of acquired resistance to anti-cancer therapy. *Journal of Theoretical Biology* **355**, 10–20 (2014).

[30] Heyde, A., Reiter, J. G., Naxerova, K. & Nowak, M. A. Consecutive seeding and transfer of genetic diversity in metastasis. *Proceedings of the National Academy of Sciences* **116**, 14129–14137 (2019).

[31] Attolini, C. S.-O. *et al.* A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. *Proceedings of the National Academy of Sciences* **107**, 17604–17609 (2010).

[32] Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nature Genetics* **48**, 238–244 (2016).

[33] Williams, M. J. *et al.* Quantification of subclonal selection in cancer from bulk sequencing data. *Nature Genetics* **50**, 895–903 (2018).

[34] Ling, S. *et al.* 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 (2015).

[35] Drasdo, D. & Höhme, S. A single-cell-based model of tumor growth in vitro: monolayers and spheroids. *Physical Biology* **2**, 133 (2005).

[36] Van Liedekerke, P. *et al.* Quantitative cell-based model predicts mechanical stress response of growing tumor spheroids over various growth conditions and cell lines. *PLoS computational biology* **15**, e1006273 (2019).

[37] Malmi-Kakkada, A. N., Li, X., Samanta, H. S., Sinha, S. & Thirumalai, D. Cell growth rate dictates the onset of glass to fluidlike transition and long time superdiffusion in an evolving cell colony. *Physical Review X* **8**, 021025 (2018).

[38] Metzcar, J., Wang, Y., Heiland, R. & Macklin, P. A review of cell-based computational modeling in cancer biology. *JCO Clinical Cancer Informatics* **2**, 1–13 (2019).

[39] Ranft, J. *et al.* Fluidization of tissues by cell division and apoptosis. *Proceedings of the National Academy of Sciences* **107**, 20863–20868 (2010).

[40] Suzuki, Y. *et al.* Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Molecular oncology* **11**, 124–139 (2017).

[41] Steel, G. Cell loss as a factor in the growth rate of human tumours. *European Journal of Cancer (1965)* **3**, 381–387 (1967).

[42] Lorenzo, C. *et al.* Live cell division dynamics monitoring in 3d large spheroid tumor models using light sheet microscopy. *Cell division* **6**, 1–8 (2011).

[43] Angaji, A., Velling, C. & Berg, J. Stochastic clonal dynamics and genetic turnover in exponentially growing populations. *Journal of Statistical Mechanics: Theory and Experiment* **2021**, 103502 (2021).

[44] Werner, B. *et al.* Measuring single cell divisions in human tissues from multi-region sequencing data. *Nature Communications* **11**, 1–9 (2020).

[45] Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

[46] Hoang, M. L. *et al.* Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Science translational medicine* **5**, 197ra102–197ra102 (2013).

[47] Serrano Colome, C., Canal Anton, O., Seplyarskiy, V. & Weghorn, D. Mutational signature decomposition with deep neural networks reveals origins of clock-like processes and hypoxia dependencies (2023). URL `https://www.biorxiv.org/content/10.1101/2023.12.06.570467v1`.

[48] Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences* **105**, 4283–4288 (2008).

[49] Zhao, Z.-M. *et al.* Early and multiple origins of metastatic lineages within primary tumors. *Proceedings of the National Academy of Sciences* **113**, 2140–2145 (2016).

[50] Alves, J. M., Prado-Lopez, S., Cameselle-Teijeiro, J. M. & Posada, D. Rapid evolution and biogeographic spread in a colorectal cancer. *Nature communications* **10**, 1–7 (2019).

[51] Hu, Z., Li, Z., Ma, Z. & Curtis, C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nature genetics* **52**, 701–708 (2020).

[52] Tarabichi, M. *et al.* Neutral tumor evolution? *Nature genetics* **50**, 1630–1633 (2018).

[53] Caravagna, G. *et al.* Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature methods* **15**, 707–714 (2018).

[54] Ng, A. W. *et al.* Aristolochic acids and their derivatives are widely implicated in liver cancers in taiwan and throughout asia. *Science translational medicine* **9**, eaan6446 (2017).

[55] Weghorn, D. & Sunyaev, S. Bayesian inference of negative and positive selection in human cancers. *Nature genetics* **49**, 1785–1788 (2017).