

A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity

Bartłomiej Waclaw¹, Ivana Božić^{2,3}, Meredith E. Pittman⁴, Ralph H. Hruban⁴, Bert Vogelstein^{4,5} & Martin A. Nowak^{2,3,6}

Most cancers in humans are large, measuring centimetres in diameter, and composed of many billions of cells¹. An equivalent mass of normal cells would be highly heterogeneous as a result of the mutations that occur during each cell division. What is remarkable about cancers is that virtually every neoplastic cell within a large tumour often contains the same core set of genetic alterations, with heterogeneity confined to mutations that emerge late during tumour growth^{2–5}. How such alterations expand within the spatially constrained three-dimensional architecture of a tumour, and come to dominate a large, pre-existing lesion, has been unclear. Here we describe a model for tumour evolution that shows how short-range dispersal and cell turnover can account for rapid cell mixing inside the tumour. We show that even a small selective advantage of a single cell within a large tumour allows the descendants of that cell to replace the precursor mass in a clinically relevant time frame. We also demonstrate that the same mechanisms can be responsible for the rapid onset of resistance to chemotherapy. Our model not only provides insights into spatial and temporal aspects of tumour growth, but also suggests that targeting short-range cellular migratory activity could have marked effects on tumour growth rates.

Tumour growth is initiated when a single cell acquires genetic or epigenetic alterations that change the net growth rate of the cell (birth minus death), and enable its progeny to outgrow surrounding cells. As these small lesions grow, the cells acquire additional alterations that cause them to multiply even faster and to change their metabolism to survive better the harsh conditions and nutrient deprivation. This progression eventually leads to a malignant tumour that can invade surrounding tissues and spread to other organs. Typical solid tumours contain about 30–70 clonal amino-acid-changing mutations that have accumulated during this multi-stage progression¹. Most of these mutations are believed to be passengers that do not affect growth, and only ~5–10% are drivers that provide cells with a small selective growth advantage. Nevertheless, a major fraction of the mutations, particularly the drivers, are present in 30–100% of neoplastic cells in the primary tumour, as well as in metastatic lesions derived from it^{2,5}.

Most attempts at explaining the genetic make-up of tumours assume well-mixed populations of cells and do not incorporate spatial constraints^{6–10}. Several models of the genetic evolution of expanding tumours have been developed in the past^{11–14}, but they assume either very few mutations^{11,12} or one- or two-dimensional growth^{13,14}. Conversely, models that incorporate spatial limitations have been developed to help to understand processes such as tumour metabolism¹⁵, angiogenesis^{16,17} and cell migration¹², but these models ignore genetics. Here, we formulate a model that combines spatial growth and genetic evolution, and use the model to describe the growth of primary tumours and metastases, as well as the development of resistance to therapeutic agents.

We first model the expansion of a metastatic lesion derived from a cancer cell that has escaped its primary site (for example, breast or colorectal epithelium) and travelled through the circulation until it lodged at a distant site (for example, lung or liver). The cell initiating the metastatic lesion is assumed to have all the driver gene mutations needed to expand. Motivated by histopathological images (Fig. 1a), we model the lesion as a conglomerate of ‘balls’ of cells (see Methods and Extended Data Fig. 1). Cells occupy sites in a regular three-dimensional lattice (Extended Data Fig. 2a, b). Cells replicate stochastically with rates proportional to the number of surrounding empty sites

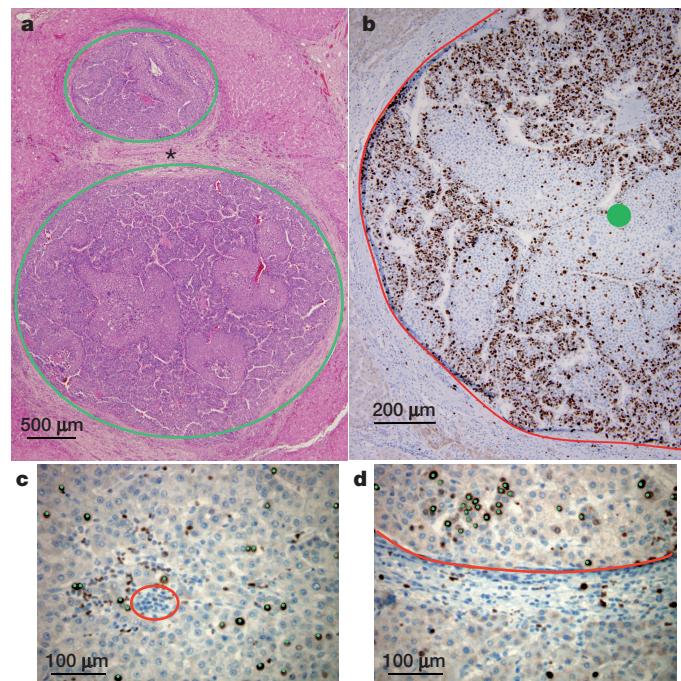
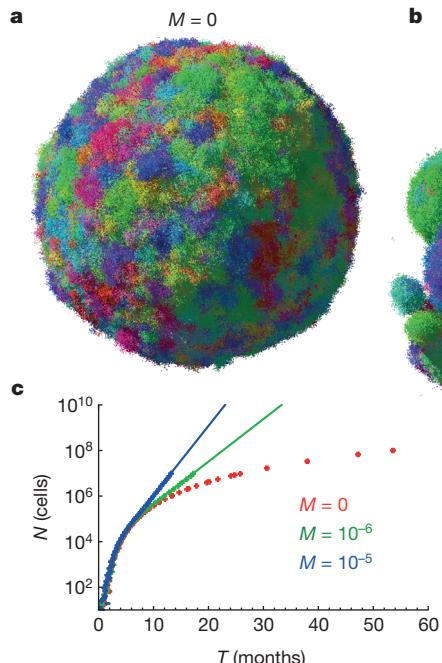


Figure 1 | Structure of solid neoplasms. **a**, Hepatocellular carcinoma composed of balls of cells (circled in green) separated by non-neoplastic tissue (asterisk). **b**, Adjacent section of the bottom tumour in **a** immunolabelled with the proliferation marker Ki67. The edge of the tumour is delineated in red; the centre is marked with a green circle. Proliferation is decreased in the centre when compared to the edge of the neoplasm. **c**, **d**, Higher magnification of the centre (**c**) and the edge (**d**) with each proliferating neoplastic cell marked by a green dot. The blue nuclei without green dots are non-proliferating. The red circle in **c** demonstrates an example of cells (inflammatory cells) that were not included in the count of neoplastic cells. The neoplastic tissue in **d** is above the red line; non-neoplastic (normal liver) is below the red line. Comparison of **c** with **d** shows that proliferation of neoplastic cells is decreased in the centre as compared to the edge of the lesion (quantified in Extended Data Table 1).

¹School of Physics and Astronomy, University of Edinburgh, JCMB, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK. ²Program for Evolutionary Dynamics, Harvard University, One Brattle Square, Cambridge, Massachusetts 02138, USA. ³Department of Mathematics, Harvard University, One Oxford Street, Cambridge, Massachusetts 02138, USA. ⁴The Sol Goldman Pancreatic Cancer Research Center, Department of Pathology, Johns Hopkins University School of Medicine, 401 North Broadway, Weinberg 2242, Baltimore, Maryland 21231, USA. ⁵Ludwig Center and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, 1650 Orleans Street, Baltimore, Maryland 21287, USA. ⁶Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.

(non-neoplastic cells or extracellular matrix), hence replication is faster at the edge of the tumour. This is supported by experimental data (Fig. 1b-d and Extended Data Table 1). A cell with no cancer cell neighbours replicates at the maximal rate of $b = \ln(2) = 0.69$ days $^{-1}$, in which b denotes the initial birth rate, equivalent to 24 h cell-doubling time, and a cell that is completely surrounded by other cancer cells does not replicate. Cells can also mutate, but we assume all mutations are passengers (they do not confer fitness advantages). After replication, a cell moves with a small probability (M) to a nearby place close to the surface of the lesion and creates a new lesion. This 'sprouting' of initial lesions could be due to short-range migration after an epithelial-to-mesenchymal transition¹⁸ and consecutive reversion to a non-motile phenotype. Alternatively, it could be the result of another process such as angiogenesis (Methods), through which the tumour gains better access to nutrients. The same model governs the evolution of larger metastatic lesions that have already developed extensive vasculature. Cells die with a death rate (d) independent of the number of neighbours, and are replaced by empty sites (non-neoplastic cells within the local tumour environment).

If there is little dispersal ($M \approx 0$), the shape of the tumour becomes roughly spherical as it grows to a large size (Fig. 2a and Supplementary Video 2). However, even a very small amount of dispersal markedly affects the predicted shape. For $M > 0$, the tumour forms a conglomerate of 'balls' (Fig. 2b, Extended Data Fig. 2c and Supplementary Video 3), much like those observed in actual metastatic lesions, with the balls separated by islands of non-neoplastic stromal cells mixed with extracellular matrix. In addition to this remarkable change in topology, dispersal strongly affects the growth rate and doubling time of the tumour. Although the size (N) of the tumour increases with time (T) from initiation as $\sim T^3$ without dispersal (Extended Data Fig. 3a, b), it grows much faster ($\sim \exp(\text{const} \times T)$ for large T) when $M > 0$ (Fig. 2c). This also remains true for long-range dispersal in which M affects the probability of escape from the primary tumour into the circulation to create new lesions in distant organs (metastasis). Using plausible estimates for the rates of cell birth, death and dispersal probability, we calculate that it takes 8 years for a lesion to grow from one cell to one billion cells in the absence of dispersal ($M = 0$), but less than 2 years with dispersal (Fig. 2c). The latter estimate is consistent with experimentally determined rates of metastasis growth as well as clinical experience, while the conventional model (without dispersal) is not.



Non-spatial models point to the size of a tumour as a crucial determinant of chemotherapeutic drug resistance^{19–21}. To determine whether a spatial model would similarly predict this dependency in a clinically relevant time frame, we calculated tumour regrowth probabilities after targeted therapies. We assume that the cell that initiates the lesion is susceptible to treatment, otherwise the treatment would have no effect on the mass, and that the probability of a resistant mutation is 10^{-7} (Methods); only one such mutation is needed for a regrowth.

Figure 3a shows snapshots from a simulation (Supplementary Video 1) performed before and after the administration of a typical targeted therapy at time $T = 0$. At first, the size of the lesion (~ 3 mm at $T = 0$) rapidly decreases, but 1 month later resistant clones begin to proliferate and form tumours of microscopic size. Such resistant subclones are predicted to be nearly always present in lesions of sizes that can be visualized by clinical imaging techniques^{21,22}. By 6 months after treatment, the lesions have regrown to their original size. The evolution of resistance is a stochastic process—some lesions shrink to zero and some regrow (Extended Data Fig. 4a). Figure 3b, c shows the probability of regrowth versus the time from the initiation of the lesion to the onset of treatment upon varying net growth rates $b-d$ and dispersal probabilities. Regardless of growth rate, the capacity to migrate makes it more likely that regrowth will occur sooner, particularly for more aggressive cancers, that is, those which have higher net growth rates (Fig. 3b). This conclusion is in line with recent theoretical work on evolving populations of migrating cells²³. If resistant mutations additionally increase the dispersal probability before or during treatment, regrowth is faster (Extended Data Fig. 4b, c).

Having shown that the predictions of the spatial model are consistent with metastatic lesion growth and regrowth times, we turn to primary tumours. In contrast to metastatic lesions, here the situation is considerably more complex because the tumour cells are continually acquiring new driver gene mutations that can endow them with fitness advantages over adjacent cells within the same tumour. Our model of a primary tumour assumes that it is initiated via a single driver gene mutation that provides a selective growth advantage over normal neighbouring cells. Each subsequent driver gene mutation reduces the death rate as $d = b(1-s)^k$, in which k is the number of driver mutations in the cell ($k \geq 1$), and s is the average fitness advantage per driver. Almost identical results are obtained if driver gene mutations increase cell birth rather than decrease cell death, or affect both

Figure 2 | Short-range dispersal affects size, shape and growth rate of tumours. a, b, A spherical lesion in the absence of dispersal ($M = 0$) (a) and a conglomerate of lesions (b), each initiated by a cell that has migrated from a previous lesion, for low but non-zero migration ($M = 10^{-6}$). Colours reflect the degree of genetic similarity; cells with similar colours have similar genetic alterations. The death rate is $d = 0.8b$, which corresponds to a net growth rate of $0.2b = 0.14$ days $^{-1}$, and $N = 10^7$ cells. c, Dispersal ($M > 0$) causes the tumour to grow faster in time. Each point = 100 samples, error bars (too small to be visible) are s.e.m. Continuous lines (extrapolation) are $6,000 \times 10^{0.437T}$ (green), $1,000 \times 10^{0.7T}$ (blue).

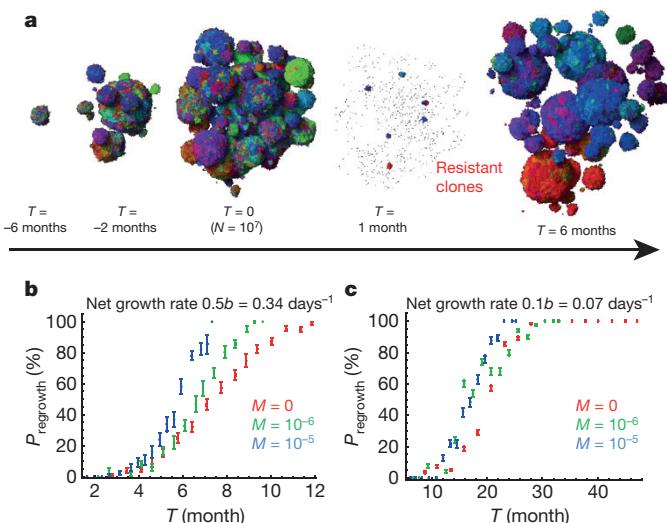


Figure 3 | Treatment success rates depend on the net growth rate of tumours. **a**, Time snapshots before and during therapy ($M = 10^{-6}$). Resistant subpopulations that cause the tumour to regrow after treatment can be seen at $T = 1$ month. **b, c**, Probability of tumour regrowth (P_{regrowth}) as a function of time after treatment initiation, for different dispersal probabilities (M) and net growth rates of the resistant cells. A higher net growth rate (**b**) leads to a high regrowth probability, so that 50% of tumours regrow 6 months after treatment is initiated when $M = 10^{-5}$. **c**, Tumours with lower net growth rates require >20 months to achieve the same probability of regrowth. Number of samples = 1 to 800 per point (282 on average). Error bars are s.e.m. See Methods for details.

cell birth and cell death (Extended Data Fig. 5b); the most important parameter is the fitness gain, s , conferred by each driver mutation.

Figure 4a shows that in the absence of any new driver mutations (as for a perfectly normal cell growing in utero), clonal subpopulations would be restricted to small, localized areas. Each of these areas has at least one new genetic alteration, but none of them confers a fitness advantage (they are ‘passengers’). In an early tumour, in which the centre cell contains the initiating driver gene mutation, the same structure would be observed—as long as no new driver gene mutations have yet appeared. The occurrence of a new driver gene mutation, however, markedly alters the spatial distribution of cells. In particular, the heterogeneity observed in normal cells (Fig. 4a) is substantially reduced (Fig. 4b and Supplementary Video 5). The degree of heterogeneity can be quantified by calculating the number of genetic alterations (passengers plus drivers) shared between two cells separated by various distances (Fig. 4d–f). The genetic diversity is markedly decreased (Fig. 4e), even with relatively small fitness advantages ($s = 1\%$). This also has implications for the number of genetic alterations that will be present in a macroscopic fraction (for example, $>50\%$) of all cells. Figure 4f shows that this number is many times larger for $s = 1\%$ than for $s = 0\%$. Furthermore, our model predicts that virtually all cells within a large tumour will have at least one new driver gene mutation after 5 years of growth (Extended Data Fig. 5a). The faster the clonal expansion occurs (the larger s is), the smaller the number of passenger mutations (Extended Data Fig. 5d, e). Our results are also robust to changes to the model (Methods and Extended Data Figs 5 and 6). We stress that an important prerequisite for limiting heterogeneity is cell turnover in the tumour, because in the spatial setting cells with driver mutations can ‘percolate’ through the tumour only if they replace other cells. In the absence of cell turnover, tumours are much more heterogeneous (Extended Data Fig. 6d).

In summary, our model accounts for many facts observed clinically and experimentally. Our results are robust and many assumptions can be relaxed without qualitatively affecting the outcome (Methods and Supplementary Information). Although tumour cell migration has

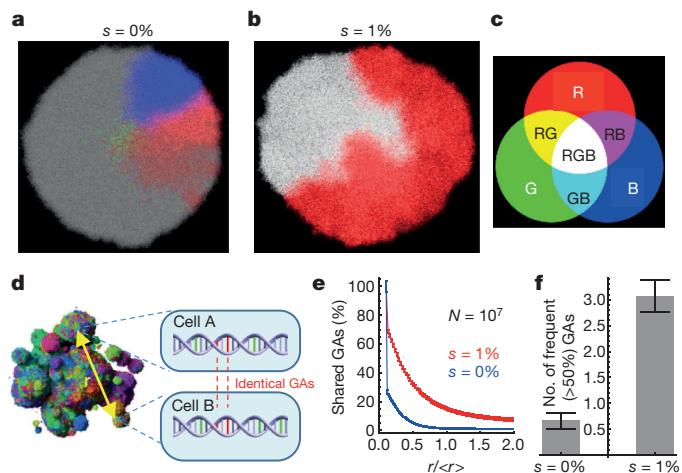


Figure 4 | Genetic diversity is strongly reduced by the emergence of driver mutations. **a–f**, For all, $M = 0$ and the initial net growth rate = 0.007 days^{-1} ($d = 0.99b$). The three most abundant genetic alterations (GAs) have been colour-coded using red (R), green (G) and blue (B) (**c**). Each section is 80 cells thick. Combinations of the three basic colours correspond to cells having two or three of these genetic alterations. **a**, No drivers—separated, conical sectors emerge in different parts of the lesion, each corresponding to a different clone. **b**, Drivers with selective advantage $s = 1\%$ lead to clonal expansions and many cells have all three genetic alterations (white area). **d**, Genetic diversity can be determined quantitatively by randomly sampling pairs of cells separated by distance r and counting the number of shared genetic alterations. **e**, The number of shared genetic alterations versus the normalized distance $r/<r>$ decreases much more slowly for the case with (red) than without (blue) driver mutations. **f**, The total number of genetic alterations present in at least 50% of all cells is much larger for $s = 1\%$ than for $s = 0\%$. Number of samples = 50 per data point. Error bars are s.e.m.

historically been viewed as a feature of cancer associated with late events in tumorigenesis, such as invasion through basement membranes or vascular walls, this classical view of migration pertains to the ability of cancer cells to migrate over large distances²⁴. Instead, our analysis reveals that even small amounts of localized cellular movement are able to markedly reshape a tumour. Moreover, we predict that the rate of tumour growth can be substantially altered by a change in dispersal rate of the cancer cells, even in the absence of any changes in doubling times or net growth rates of the cells within the tumour. Some of our predictions could be experimentally tested using new cell labelling techniques^{25,26}. Our results could also greatly inform the interpretation of mutations in genes whose main functions seem to be related to the cytoskeleton or to cell adhesion rather than to cell birth, death, or differentiation^{27,28}. For example, cells that have lost the expression of E-cadherin (a cell adhesion protein) are more migratory than normal cells with intact E-cadherin expression²⁹, and loss of E-cadherin in pancreatic cancer has been associated with poorer prognosis³⁰, in line with our predictions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 September 2014; accepted 23 July 2015.

Published online 26 August 2015.

1. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
2. Yachida, S. et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
3. Sottoriva, A. et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. USA* **110**, 4009–4014 (2013).
4. Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
5. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
6. Geteny, R. A. & Vincent, T. L. An evolutionary model of carcinogenesis. *Cancer Res.* **63**, 6212–6220 (2003).

7. Johnston, M. D., Edwards, C. M., Bodmer, W. F., Maini, P. K. & Chapman, S. J. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proc. Natl Acad. Sci. USA* **104**, 4008–4013 (2007).
8. Bozic, I. et al. Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl Acad. Sci. USA* **107**, 18545–18550 (2010).
9. Beerenwinkel, N. et al. Genetic progression and the waiting time to cancer. *PLOS Comput. Biol.* **3**, e225 (2007).
10. Durrett, R. & Moseley, S. Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Popul. Biol.* **77**, 42–48 (2010).
11. González-García, I., Sole, R. V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl Acad. Sci. USA* **99**, 13085–13089 (2002).
12. Thalhauser, C. J., Lowengrub, J. S., Stupack, D. & Komarova, N. L. Selection in spatial stochastic models of cancer: migration as a key modulator of fitness. *Biol. Direct* **5**, 21 (2010).
13. Martens, E. A., Kostadinov, R., Maley, C. C. & Hallatschek, O. Spatial structure increases the waiting time for cancer. *New J. Phys.* **13**, 115014 (2011).
14. Anderson, A. R. A., Weaver, A. M., Cummings, P. T. & Quaranta, V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell* **127**, 905–915 (2006).
15. Kim, Y., Magdalena, A. S. & Othmer, H. G. A hybrid model for tumor spheroid growth *in vitro*: theoretical development and early results. *Math. Models Methods Appl. Sci.* **17**, 1773–1798 (2007).
16. McDougall, S. R., Anderson, A. R. & Chaplain, M. A. Mathematical modeling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. *J. Theor. Biol.* **241**, 564–589 (2006).
17. Hawkins-Daarud, A., Rockne, R. C., Anderson, A. R. A. & Swanson, K. R. Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor. *Front. Oncol.* **3**, 66 (2013).
18. McDonald, O. G., Wu, H., Timp, W., Doi, A. & Feinberg, A. P. Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nature Struct. Mol. Biol.* **18**, 867–874 (2011).
19. Komarova, N. L. & Wodarz, D. Drug resistance in cancer: principles of emergence and prevention. *Proc. Natl Acad. Sci. USA* **102**, 9714–9719 (2005).
20. Bozic, I., Allen, B. & Nowak, M. A. Dynamics of targeted cancer therapy. *Trends Mol. Med.* **18**, 311–316 (2012).
21. Bozic, I. & Nowak, M. A. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc. Natl Acad. Sci. USA* **111**, 15964–15968 (2014).
22. Turke, A. B. et al. Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).
23. Komarova, N. L. Spatial interactions and cooperation can change the speed of evolution of complex phenotypes. *Proc. Natl Acad. Sci. USA* **111**, 10789–10795 (2014).
24. Talmadge, J. E. & Fidler, I. J. AACR Centennial Series: the biology of cancer metastasis: historical perspective. *Cancer Res.* **70**, 5649–5669 (2010).
25. Alcolea, M. P. et al. Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nature Cell Biol.* **16**, 615–622 (2014).
26. Weber, K. et al. RGB marking facilitates multicolor clonal cell tracking. *Nature Med.* **17**, 504–509 (2011).
27. Bordeleau, F., Alcoser, T. A. & Reinhart-King, C. A. Physical biology in cancer. 5. The rocky road of metastasis: the role of cytoskeletal mechanics in cell migratory response to 3D matrix topography. *Am. J. Physiol. Cell Physiol.* **306**, C110–C120 (2014).
28. Lawson, C. D. & Burridge, K. The on-off relationship of Rho and Rac during integrin-mediated adhesion and cell migration. *Small GTPases* **5**, e27958 (2014).
29. Gall, T. M. H. & Frampton, A. E. Gene of the month: E-cadherin (CDH1). *J. Clin. Pathol.* **66**, 928–932 (2013).
30. Winter, J. M. et al. Absence of E-cadherin expression distinguishes noncohesive from cohesive pancreatic cancer. *Clin. Cancer Res.* **14**, 412–418 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements Support from The John Templeton Foundation is gratefully acknowledged. B.W. was supported by the Leverhulme Trust Early-Career Fellowship, and the Royal Society of Edinburgh Personal Research Fellowship. I.B. was supported by Foundational Questions in Evolutionary Biology Grant RFP-12-17. M.E.P., R.H.H. and B.V. acknowledge support from The Virginia and D.K. Ludwig Fund for Cancer Research, The Lustgarten Foundation for Pancreatic Cancer Research, The Sol Goldman Center for Pancreatic Cancer Research, and NIH grants CA43460 and CA62924.

Author Contributions B.W., M.A.N., I.B. and B.V. designed the study. B.W. wrote the computer programs and made simulations. B.W., I.B. and M.A.N. made analytic calculations. M.E.P. and R.H.H. carried out experimental work. All authors discussed the results. The manuscript was written primarily by B.W., M.A.N., I.B. and B.V., with contributions from M.E.P. and R.H.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.N. (martin_nowak@harvard.edu).

METHODS

No statistical methods were used to predetermine sample size. Experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Spatial model for tumour evolution. Tumour modelling has a long tradition³¹. Many models of spatially expanding tumours were proposed in the past^{12–14,16,32–42}, but they either assume very few^{32,34–37,39,41,43–47} or no new mutations at all^{12,15,38,48,49}, or one- or two-dimensional growth^{12,13,32,33,50–52}. On the other hand, well-mixed models with several mutations^{6,8,53,54} do not often include space, and computational models aimed at being more biologically realistic^{15,55–61} require too much computing resources (time and memory) to simulate realistically large tumours ($N \approx 10^9$ cells). Our model builds on the Eden lattice model⁶² and combines spatial growth and accumulation of multiple mutations. Since we focus on the interplay of genetics, spatial expansion and short-range dispersal of cells, for simplicity we do not explicitly model metabolism¹⁴, tissue mechanics, spatial heterogeneity of tissues, different types of cells present or angiogenesis¹⁶.

A tumour is made of non-overlapping balls (microlesions) of cells. Tumour cells occupy sites of a regular 3D square lattice (Moore neighbourhood, 26 neighbours). Empty lattice sites are assumed to be either normal cells or filled with extracellular matrix and are not modelled explicitly. Each cell in the model is described by its position and a list of genetic alterations that have occurred since the initial neoplastic cell, and the information about whether a given mutation is a passenger, driver, or resistance-carrying mutation. A passenger mutation does not affect the net growth rate whereas a driver mutation increases it by disrupting tight regulation of cellular divisions and shifts the balance towards increased proliferation or decreased apoptosis. The changes can also be epigenetic and we do not distinguish between different types of alterations. We assume that each genetic alteration occurs only once ('infinite allele model'⁶³). The average numbers of all genetic alterations, driver and resistant genetic alterations produced in a single replication event are denoted by γ , γ_d and γ_r , respectively. When a cell replicates, each of the daughter cells receives n new genetic alterations of each type (n being generally different in both cells) drawn at random from the Poisson probability distribution:

$$P(n) = \frac{e^{-\gamma_x/2} (\gamma_x/2)^n}{n!} \quad (1)$$

in which x denotes the type of genetic alteration.

In model A shown in Figs 2–4, replication occurs stochastically, with rate proportional to the number of empty sites surrounding the replicating cell, and death occurs with constant rate depending only on the number of drivers. We also simulated other scenarios (models B, C and D, see below). Driver mutations increase the net growth rate (the difference between proliferation and death) either by increasing the birth rate or decreasing the death rate by a constant factor $1 + s$, in which $s > 0$.

Dispersal is modelled by moving an offspring cell to a nearby position where it starts a new microlesion (Extended Data Fig. 1a). Microlesions repel each other; a 'shoving' algorithm^{64,65} (Extended Data Fig. 1b) ensures they do not merge.

Code availability. The computer code (available at <http://www2.ph.ed.ac.uk/~bwaclaw/cancer-code>) can handle up to 1×10^9 cells, which corresponds to tumours that are clinically meaningful and can be observed by conventional medical imaging (diameter > 1 cm). The algorithm is discussed in details in the Supplementary Information. It is not an exact kinetic Monte Carlo algorithm because such an algorithm would be too slow to simulate large tumours. A comparison with kinetic Monte Carlo for smaller tumours (Supplementary Information) shows that both algorithms produce consistent results.

Model parameters. The initial birth rate $b = \ln(2) \approx 0.69$ days⁻¹, which corresponds to a 24 h minimum doubling time. The initial death rate $d = 0 \dots 0.995b$ depends on the aggressiveness of the tumour (larger values = less aggressive lesion). In simulations of targeted therapy, we assume that, before treatment, $b = 0.69$ days⁻¹ and $d = 0.5b = 0.35$ days⁻¹, whereas during treatment $b = 0.35$ days⁻¹ and $d = 0.69$ days⁻¹, that is, birth and death rates swap places. This rather arbitrary choice leads to the regrowth time of about 6 months, which agrees well with clinical evidence. Mutation probabilities are $\gamma = 0.02$, $\gamma_d = 4 \times 10^{-5}$, $\gamma_r = 1 \times 10^{-7}$, in line with experimental evidence and theoretical work^{8,66–68}. Since there are no reliable data on the dispersal probability M , we have explored a range of values between $M = 1 \times 10^{-7}$ and 1×10^{-2} . All parameters are summarized in Extended Data Fig. 1c, see also further discussion in Supplementary Information.

Validity of the assumptions of the model. Our model is deliberately oversimplified. However, many of the assumptions we make can be experimentally justified or shown not to qualitatively affect the model.

Three-dimensional regular lattice of cells. The 3D Moore neighbourhood was chosen because it is computationally fast and introduces relatively fewer artefacts related to lattice symmetries. Real tissues are much less regular and the number of nearest neighbours is different⁶⁹. However, recent simulations of similar models of

bacterial colonies^{70,71} show that the structure of the lattice (or the lack thereof in off-lattice models) has a marginal effect on genetic heterogeneity.

Asynchronous cell division. Division times of related cells remain correlated for a few generations. However, stochastic cell division implemented in our model is a good approximation for a large mass of cells and is much less computationally expensive than modelling a full cell cycle.

Replication faster at the boundary than in the interior. Several studies have described a higher proliferation rate at the leading edge of tumours, and this has been associated with a more aggressive clinical course⁷². To estimate the range of values of death rate d for our model, we used the proliferation marker Ki67. Representative formalin-fixed, paraffin-embedded tissue blocks were selected from four small chromophobe renal cell carcinomas and six small hepatocellular carcinomas by the pathologist (M.E.P.). A section of each block was immunolabelled for Ki67 using the Ventana Benchmark XT system. Around 8–12 images, depending on the size of the lesion, were acquired from each tumour. Fields were chosen at random from the leading edge and the middle of the tumour and were not necessarily 'hot spots' of proliferative activity. Using an ImageJ macro, each Ki67-positive tumour nucleus was labelled green by the pathologist, and each Ki67-negative tumour nucleus was labelled red. Other cell types (endothelium, fibroblasts and inflammatory cells) were not labelled. The proliferation rate was then calculated using previously described methods⁷³. Statistical significance of the results was determined using a Kolmogorov-Smirnov two-sample test (significance level 0.05). The study was approved by the Institutional Review Board of the Johns Hopkins University School of Medicine. In all ten tumours, the proliferation rate at the leading edge of the tumour was greater than that at the centre by a factor of 1.25 to 6 (Extended Data Table 1). Comparing the density of proliferating cells to our model gives $d \approx 0.5b$ (range: $d = 0.17b \dots 0.8b$), which is what we assume in the simulations of aggressive lesions.

Equal fitness of all cells in metastatic lesions. We assume that cells in a metastatic lesion are already very fit since they contain multiple drivers. Indeed, studies of primary tumours and their matched metastases usually fail to find driver mutations present in the metastases that were not present in the primary lesions^{2,74}, although there are notable exceptions, see, for example, refs 75 and 76. Experimental evidence in microbes⁷⁷ and (to a lesser extent) in eukaryotes⁷⁸ suggests that fitness gains due to individual mutations are largest at the beginning of an evolutionary process and that the effects of later mutations are much smaller. It remains to be seen how well these results apply to late genetic alterations in cancer⁷⁹ but if true, new drivers occurring in the lesion are unlikely to spread through the population before the lesion reaches a clinically relevant size.

Dispersal. In our model, cells detach from the lesion and attach again at a different location in the tissue. This can be viewed either as cells migrating from one place to another one, or as a more generic mechanism that allows tumour cells to get better access to nutrients by dispersing within the tissue, hence providing a growth advantage over cells that did not disperse. Some mechanisms that do not involve active motion (that is, cells becoming motile) are discussed below.

Migration. Cancer cells are known to undergo epithelial-to-mesenchymal transition, the origin of which is thought to be epigenetic¹⁸. This involves a cell becoming motile and moving some distance. If the cell finds the right environment, it can switch back to the non-motile phenotype and start a new lesion. Motility can be enhanced by tissue fluidization due to replication and death⁸⁰. Instead of modelling the entire cycle (epithelial–mesenchymal–epithelial), we only model the final outcome (a cell has moved some distance).

Tumour buds. Many tumours exhibit focally invasive cell clusters, also known as tumour buds. Their proliferation rate is less than that of cells in the main tumour⁸¹. We propose that tumour buds contain cells that have not yet completed epithelial-to-mesenchymal transition and therefore they proliferate slower.

Single versus cluster migration. Ref. 82 found that circulating cancer cells can travel in clusters of 2–50 cells, and that such clusters can initiate metastatic foci. They report that approximately one-half of the metastatic foci they examined were initiated by single circulating cancer cells, and that circulating cancer cell clusters initiated the other half. The authors also note that the cells forming a cluster are probably neighbouring cancer cells from the primary tumour. This means that the genetic make-up of cells within a newly established lesion will be very similar, regardless of its origin (single cell versus a small cluster of cells). Therefore, the ability to travel in clusters should not affect the genetic heterogeneity or regrowth probability as compared to single-cell dispersal from our model.

Angiogenesis. We do not explicitly model angiogenesis for two reasons. First, most genetic alterations that can either change the growth rate or be detected experimentally must occur at early stages of tumour growth as explained before. Hence, the genetic make-up of the tumour is determined primarily by what happens before angiogenesis. Second, local dispersal from the model mimics tumour cells interspersing with the vascularized tissue and getting better access to nutrients, which is one of the outcomes of angiogenesis.

Biomechanics of tumours. Growth is affected by the mechanical properties of cells and the extracellular matrix. We do not explicitly include biomechanics (see, however, below), in contrast to more realistic models^{83,84}, as this would not allow us to simulate lesions larger than about 1×10^6 cells. Instead, we take experimentally determined values for birth and death rates, values that are affected by biomechanics, as the parameters of our model.

Isolated balls of cells. In our simulations, balls of cells are thought to be separated by normal, vascularized tissue which delivers nutrients to the tumour. The environment of each ball is the same, and there are no interactions between the balls other than mechanical repulsion. This represents a convenient mathematical contrivance and qualitatively recapitulates what is observed in stained sections of actual tumours (Fig. 1a). We investigated under which circumstances the balls of cancer cells would mechanically repel each other; see Extended Data Fig. 7 for a graphical summary of the results. We simulated a biomechanical, off-lattice model of normal tissue composed of ‘ducts’ lined with epithelial cells and separated by stroma (Supplementary Information, section 8). Mechanical interactions between cells were modelled using an approach similar to that described previously^{59,60,85}, with model parameters taken from refs 59, 60, 85–88. We assumed cancer cells to be of epithelial origin, as are most cancers⁸⁹. Cancer cells that invaded different areas of epithelium grew into balls that remained separated by thin slices of stroma (Supplementary Videos 8–11). This ‘encapsulation’ of tumour microlesions was possible owing to the supportive nature of stroma that is able to mechanically resist expansion of balls of cancer cells. Encapsulation is essential if the balls are to repel each other. If the tissue is ‘fluidized’ by random replication and death, the balls quickly merge (Supplementary Video 12). Another important factor are differences in mechanical properties of tumour and normal cells⁹⁰; it is known that differences in cellular adhesion and stiffness promote segregation of different types of cells^{91,92}.

In reality, microlesions within the primary tumour are less symmetric and some of them are better described as ‘protrusions’ bulging out from the main tumour tissue, owing to biomechanical instabilities; see, for example, refs 93, 94. However, stroma may still provide enough spatial separation, and the capillary network of blood vessels—either due to tumour angiogenesis or preexisting in the invaded tissue—may provide enough nutrients to the lesions so that our assumption of independently growing balls of cells remains valid. Therefore, we believe that modelling the tumour as a collection of non- or weakly-interacting microlesions is essentially correct. We also note that the existence of isolated balls is not necessary to explain our qualitative results: reduced heterogeneity and increased growth in the presence of migration. Supplementary Video 13 shows that even if the tissue is homogeneous and highly dynamical and there are no isolated balls of cells, migration leads to a considerable speedup of growth as compared to the case with no migration (Supplementary Video 14).

Tumour geometry and heterogeneity in the absence of driver mutations. Supplementary Videos 2 and 3 illustrate the process of growth of a tumour with maximally $N = 10^7$ cells, for $M = 0$ and $M = 10^{-6}$, respectively, and for $d = 0.5$. Extended Data Fig. 2 shows snapshots from a single simulation for $M = 0$, $N \sim 10^3$, and $d = 0$ (no death, Extended Data Fig. 2a) and $d = 0.9$ (Extended Data Fig. 2b). In the latter case, cells are separated by empty sites (normal cells/extracellular matrix). Extended Data Fig. 2c shows that the tumour is almost spherically symmetric for $M = 0$. The symmetry is lost for small but non-zero M , and restored for larger M when the balls become smaller and their number increases. Extended Data Fig. 2c also shows that metastatic tumours contain many clonal sectors with passenger mutations. Extended Data Fig. 8a shows that the fraction $G(r)$ of genetic alterations that are the same in two randomly sampled cells (Fig. 4) separated by distance r quickly decreases with r , indicating increased genetic heterogeneity owing to passenger mutations.

Targeted therapy of metastatic lesions. Models of cancer treatment^{21,95–100} often assume either no spatial structure or do not model the emergence of resistance. We assume that the cell that initiated the lesion was sensitive to treatment but its progeny may become resistant. Before the therapy commences, all cells have the same birth and death rates, but after the treatment resistant cells continue to proliferate with the same rate, whereas susceptible cells are assigned different rates as described above. Resistant cells can emerge before and during the therapy. The death rate of sensitive cells during treatment is greater than the birth rate, or the tumours would not be sensitive to the drug. For example, in Fig. 3 treatment increases the death rate and decreases the growth rate of susceptible cells, the growth rate of resistant cells after therapy is identical to that of the sensitive cells before treatment, $d = 0.5b$ in the absence of treatment, $M = 10^{-6}$, and treatment begins when the tumour has $N = 10^7$ cells.

Note that our model assumes the drug is uniformly distributed in the tumour⁹⁶; it is known that drug gradients can speed up the onset of resistance¹⁰¹.

Supplementary Video 1 and Extended Data Fig. 4a show that, since the process of resistance acquisition is stochastic, some tumours regrow after an initial

regression, and some do not. If only resistant cells can migrate, regrowth is faster (Extended Data Fig. 4b, c). Extended Data Fig. 4d–g shows regrowth probabilities P_{regrowth} for different treatment scenarios not mentioned in the main text, depending on whether the drug is cytostatic ($b_{\text{treatment}} = 0$) or cytotoxic ($d_{\text{treatment}} = b$), and whether $d = 0$ or $d > 0$ before treatment. In Extended Data Fig. 4d, cells replicate and die only on the surface, and the core is ‘quiescent’—cells are still alive there but cannot replicate unless outer layers are removed by treatment (Supplementary Videos 6 and 7). P_{regrowth} does not depend on the dispersal probability M at all, and is close to 100% for $N > 10^8$ cells, a size that is larger than for $d > 0$ (Extended Data Fig. 4f). It can be shown that $P_{\text{regrowth}} = 1 - \exp(-\gamma N)$. Extended Data Fig. 4e is for the cytostatic drug ($b_{\text{treatment}} = d_{\text{treatment}} = 0$); this is also equivalent to the cytotoxic drug if the tumour has a necrotic core (cells are dead but still occupy physical volume). In this case, P_{regrowth} increases with M because more resistant cells are on the surface for larger M (cells can replicate only on the surface in this scenario). Extended Data Fig. 4f, g shows models with cell death present even in the absence of treatment ($d = 0.9b$) but occurring only at the surface, unlike in Fig. 3 where cells also die inside the tumour. Death increases P_{regrowth} owing to a larger number of cellular division necessary to obtain the same size, and hence more opportunities to mutate.

Relaxing the assumptions of the model. Figure 4 shows that even a small fitness advantage substantially reduces genetic diversity through the process of clonal expansion, see also Supplementary Videos 4 and 5. We now demonstrate that this also applies to modified versions of the model, proving its robustness.

Exact values of M and s has no qualitative effect. Extended Data Fig. 8b, e shows that the average number of shared genetic alterations is larger in the presence of drivers also in the case of non-zero dispersal ($M > 0$), and its numerical value is almost the same as for $M = 0$ (Fig. 4). Extended Data Fig. 8c, f shows that as long as $s > 0$ and regardless of its exact value, driver mutations reduce genetic diversity in the tumour compared to the case $s = 0$. Extended Data Fig. 5a–c shows how many driver mutations are expected to be present in a randomly chosen cell from a tumour that is T years old. Neither dispersal nor the way drivers affect growth (via birth or death rate) has a significant effect on the number of drivers per cell (Extended Data Fig. 5b, c). A small discrepancy visible in Extended Data Fig. 5b is caused by a slightly asymmetric way death and birth is treated in our model, see the Supplementary Information.

Model B. Cells replicate with constant rate if there is at least one empty neighbour. In the absence of drivers, genetic alterations are distributed evenly throughout the lesion (Extended Data Fig. 6b) but they often occur independently and the number of frequent genetic alterations is low (Extended Data Fig. 6e). Drivers cause clonal expansion as in model A.

Model C. Cells replicate regardless of whether there are empty sites surrounding them or not. When a cell replicates, it pushes away other cells towards the surface (Supplementary Information). Extended Data Fig. 6c, e shows that this again leads to clonal expansion which decreases diversity.

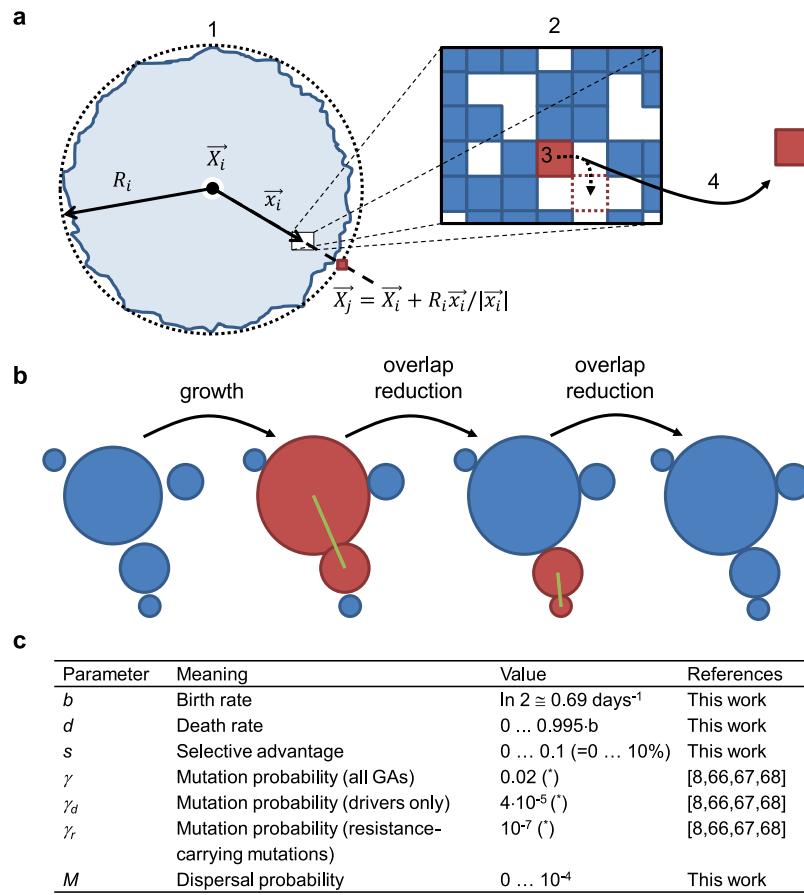
Model D. Replication/death occurs only on the surface and the core of the tumour is static. Extended Data Fig. 6d shows that driver mutations cannot spread to the other side of the lesion and conical clonal sectors can be seen even for $s > 0$. The number of frequent genetic alterations is the same for $s = 0$ and $s = 1\%$, indicating that genetic heterogeneity is not lowered by clonal expansion. This demonstrates that cell turnover inside the tumour is very important for reducing heterogeneity. To obtain the same (low) heterogeneity as for models a–c, the probability of driver mutations must be much larger in model D (Extended Data Fig. 6f).

Drivers affecting M . We investigated three scenarios in which drivers affect (1) only the dispersal probability $M \rightarrow (1 + q)M$, in which $q > 0$ is the ‘migration fitness advantage’ (no change in b , d), (2) both M and d , that is, $(d, M) \rightarrow (d(1 - s), (1 + q)M)$ with s , $q > 0$, (3) either M or d , with probability 1/2. Extended Data Fig. 3c shows that growth is unaffected in cases (1, 3) compared to the neutral case. For (2) the tumour growth rate increases significantly when the tumour is larger than $N = 1 \times 10^6$ cells. This shows that migration increases the overall fitness advantage, in line with ref. 102, which shows that fixation probability is determined by the product of the exponential growth rate and diffusion constant (motility) of organisms.

Six-site (von Neumann) neighbourhood. We simulated a model in which each cell has only six neighbours (von Neumann neighbourhood) instead of 26 (Moore neighbourhood). Extended Data Fig. 9 compares models A and C for the two neighbourhoods and show that there is only a small quantitative difference in the growth curves for model A (model C is unaffected), but that the shape of the ball of cells deviates more from the spherical one for the six-site neighbourhood, see also section 7 in the Supplementary Information.

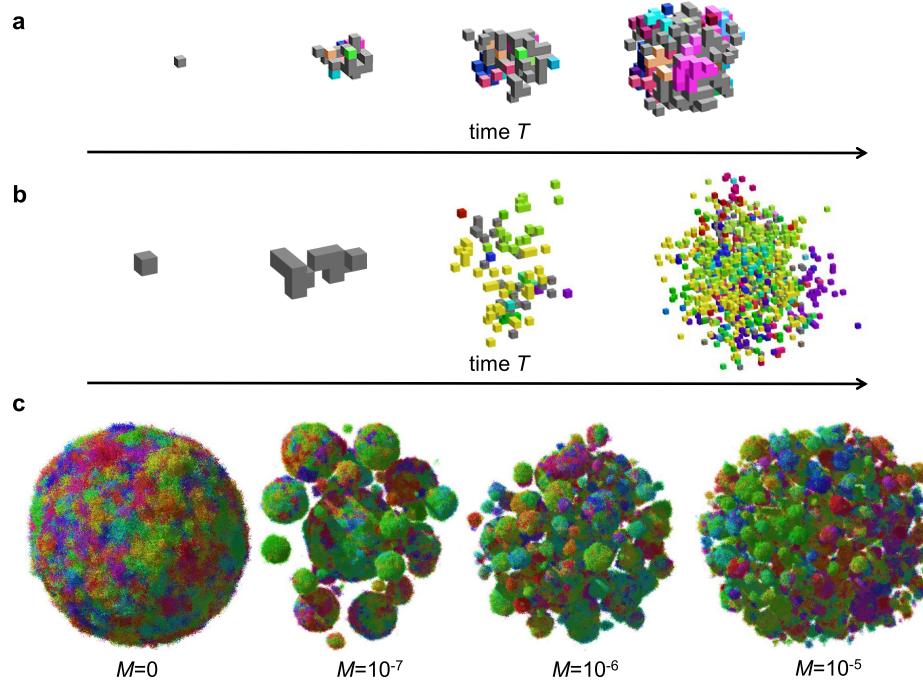
31. Rodriguez-Brenes, I. A., Komarova, N. L. & Wodarz, D. Tumor growth dynamics: insights into somatic evolutionary processes. *Trends Ecol. Evol.* **28**, 597–604 (2013).

32. Anderson, A. R. A. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Math. Med. Biol.* **22**, 163–186 (2005).
33. Komarova, N. L. Spatial stochastic models for cancer initiation and progression. *Bull. Math. Biol.* **68**, 1573–1599 (2006).
34. Lavrentovich, M. O. & Nelson, D. R. Survival probabilities at spherical frontiers. *Theor. Popul. Biol.* **102**, 26–39 (2015).
35. Komarova, N. L. Spatial stochastic models of cancer: fitness, migration, invasion. *Math. Biosci. Eng.* **10**, 761–775 (2013).
36. Manev, V. S. K., Kohandel, M., Komarova, N. L. & Sivaloganathan, S. Spatial invasion dynamics on random and unstructured meshes: implications for heterogeneous tumor populations. *J. Theor. Biol.* **349**, 66–73 (2014).
37. Durrett, R., Foo, J. & Leder, K. Spatial Moran models, II: cancer initiation in spatially structured tissue. *J. Math. Biol.* <http://dx.doi.org/10.1007/s00285-015-0912-1> (2015).
38. Gerlee, P. & Nelander, S. The impact of phenotypic switching on glioblastoma growth and invasion. *PLOS Comput. Biol.* **8**, e1002556 (2012).
39. González-García, I., Solé, R. V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl Acad. Sci. USA* **99**, 13085–13089 (2002).
40. Sehyo, C. C. et al. Model for *in vivo* progression of tumors based on co-evolving cell population and vasculature. *Sci. Rep.* **1**, 31 (2011).
41. Torquato, S. Toward an Ising model of cancer and beyond. *Phys. Biol.* **8**, 015017 (2011).
42. Reiter, J. G., Bozic, I., Allen, B., Chatterjee, K. & Nowak, M. A. The effect of one additional driver mutation on tumor progression. *Evol. Appl.* **6**, 34–45 (2013).
43. Kansal, A. R., Torquato, S., Harsh, G. R., Chiocca, E. A. & Deisboeck, T. S. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J. Theor. Biol.* **203**, 367–382 (2000).
44. Kansal, A. R., Torquato, S., Chiocca, E. A. & Deisboeck, T. S. Emergence of a subpopulation in a computational model of tumor growth. *J. Theor. Biol.* **207**, 431–441 (2000).
45. Antal, T., Krapivsky, P. L. & Nowak, M. A. Spatial evolution of tumors with successive driver mutations. *Phys. Rev. E* **92**, 022705 (2015).
46. Enderling, H., Hlatky, L. & Hahnfeldt, P. Migration rules: tumours are conglomerates of self-metastases. *Br. J. Cancer* **100**, 1917–1925 (2009).
47. Sottoriva, A. et al. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer Res.* **70**, 46–56 (2010).
48. Schaller, G. & Meyer-Hermann, M. Multicellular tumor spheroid in an off-lattice voronoi delaunay cell model. *Phys. Rev. E* **71**, 051910 (2005).
49. Radzsuweit, M., Block, M., Hengstler, J. G., Schöll, E. & Drasdo, D. Comparing the growth kinetics of cell populations in two and three dimensions. *Phys. Rev. E* **79**, 051907 (2009).
50. Moglia, B., Guisoni, N. & Albano, E. V. Interfacial properties in a discrete model for tumor growth. *Phys. Rev. E* **87**, 032713 (2013).
51. Foo, J., Leder, K. & Ryser, M. Multifocality and recurrence risk: a quantitative model of field cancerization. *J. Theor. Biol.* **355**, 170–184 (2014).
52. Poleszczuk, J., Hahnfeldt, P. & Enderling, H. Evolution and phenotypic selection of cancer stem cells. *PLOS Comput. Biol.* **11**, e1004025 (2015).
53. Durrett, R., Schmidt, D. & Schweinsberg, J. A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Probab.* **19**, 676–718 (2009).
54. Spencer, S. L., Berryman, M. J., García, J. A. & Abbott, D. An ordinary differential equation model for the multistep transformation to cancer. *J. Theor. Biol.* **231**, 515–524 (2004).
55. Kim, Y. & Othmer, H. G. A hybrid model of tumor–stromal interactions in breast cancer. *Bull. Math. Biol.* **75**, 1304–1350 (2013).
56. Ramis-Conde, I., Chaplain, M. A. J., Anderson, A. R. A. & Drasdo, D. Multi-scale modelling of cancer cell intravasation: the role of cadherins in metastasis. *Phys. Biol.* **6**, 016008 (2009).
57. Swanson, K. R. et al. Quantifying the role of angiogenesis in malignant progression of gliomas: *in silico* modeling integrates imaging and histology. *Cancer Res.* **71**, 7366–7375 (2011).
58. Taloni, A. et al. Mechanical properties of growing melanocytic nevi and the progression to melanoma. *PLoS ONE* **9**, e94229 (2014).
59. Drasdo, D. & Höhme, S. A single-cell-based model of tumor growth *in vitro*: monolayers and spheroids. *Phys. Biol.* **2**, 133 (2005).
60. Drasdo, D., Hoehme, S. & Block, M. On the role of physics in the growth and pattern formation of multi-cellular systems: what can we learn from individual-cell based models? *J. Stat. Phys.* **128**, 287–345 (2007).
61. Jiang, Y., Pjesivac-Grbic, J., Cantrell, C. & Freyer, J. P. A multiscale model for avascular tumor growth. *Biophys. J.* **89**, 3884–3894 (2005).
62. Eden, M. in *A Two-Dimensional Growth Process* (eds Family, F. & Vicsek, T.) 265–283 (World Scientific, 1961).
63. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates, 1997).
64. Kretzschmar, M. & de Bruin, H. A. *Principles of Population Genetics* (Sinauer Associates, 1997).
65. Kreft, J. U., Booth, G. & Wimpenny, J. W. T. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* **144**, 3275–3287 (1998).
66. Lardon, L. A. et al. iDyMoMiCS: Next-generation individual-based modelling of biofilms. *Environ. Microbiol.* **13**, 2416–2434 (2011).
67. Jones, S. et al. Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105**, 4283–4288 (2008).
68. Diaz, L. A. Jr et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
69. Honda, H., Morita, T. & Tanabe, A. Establishment of epidermal cell columns in mammalian skin: computer simulation. *J. Theor. Biol.* **81**, 745–759 (1979).
70. Ali, A., Somfai, E. & Grosskinsky, S. Reproduction-time statistics and segregation patterns in growing populations. *Phys. Rev. E* **85**, 021923 (2012).
71. Korolev, K. S., Xavier, J. B., Nelson, D. R. & Foster, K. R. Data from: a quantitative test of population genetics using spatio-genetic patterns in bacterial colonies. Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.3147q> (2011).
72. Gong, P., Wang, Y., Liu, G., Zhang, J. & Wang, Z. New insight into Ki67 expression at the invasive front in breast cancer. *PLoS ONE* **8**, e54912 (2013).
73. Ellison, T. A. et al. A single institution's 26-year experience with nonfunctional pancreatic neuroendocrine tumors: a validation of current staging systems and a new prognostic nomogram. *Ann. Surg.* **259**, 204–212 (2014).
74. Jones, S. et al. Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105**, 4283–4288 (2008).
75. Lindstrom, L. S. et al. Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J. Clin. Oncol.* **30**, 2601–2608 (2012).
76. Voss, M. H. et al. Tumor genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR inhibitor therapy. *Clin. Cancer Res.* **20**, 1955–1964 (2014).
77. Wiser, M. J., Ribeck, N. & Lenski, R. E. Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367 (2013).
78. White, T. C. Increased mRNA Levels of ERG16, CDR, and MDR1 correlate with increases in azole resistance in *Candida albicans* isolates from a patient infected with human immunodeficiency virus. *Antimicrob. Agents Chemother.* **41**, 1482–1487 (1997).
79. McGranahan, N. et al. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci. Transl. Med.* **7**, 283 (2015).
80. Ranft, J. et al. Fluidization of tissues by cell division and apoptosis. *Proc. Natl Acad. Sci. USA* **107**, 20863–20868 (2010).
81. LeBleu, V. S. et al. PGC-1 α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nature Cell Biol.* **16**, 992–1003 (2014).
82. Aceto, N. et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).
83. Sciumè, G. et al. A multiphase model for three-dimensional tumor growth. *New J. Phys.* **15**, 015005 (2013).
84. Charras, G. & Sahai, E. Physical influences of the extracellular environment on cell migration. *Nature Rev. Mol. Cell Biol.* **15**, 813–824 (2014).
85. Jiao, Y. & Torquato, S. Diversity of dynamics and morphologies of invasive solid tumors. *AIP Advances* **2**, 011003 (2012).
86. Galle, J., Loeffler, M. & Drasdo, D. Modeling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations *in vitro*. *Biophys. J.* **88**, 62–75 (2005).
87. Chen, E. J., Novakofski, J., Jenkins, W. K. & O'Brien, W. D. Jr. Young's modulus measurements of soft tissues with application to elasticity imaging. *Ultrascons, ferroelectrics, and frequency control. IEEE Transactions* **43**, 191–194 (1996).
88. Samani, A., Bishop, J., Luginbuhl, C. & Plewes, D. B. Measuring the elastic modulus of *ex vivo* small tissue samples. *Phys. Med. Biol.* **48**, 2183 (2003).
89. Weinberg, R. A. *The Biology of Cancer* (Garland Science, 2007).
90. Lekka, M. et al. Elasticity of normal and cancerous human bladder cells studied by scanning force microscopy. *Eur. Biophys. J.* **28**, 312–316 (1999).
91. Gonzalez-Rodriguez, D., Guevorkian, K., Douezan, S. & Brochard-Wyart, F. Soft matter models of developing tissues and tumors. *Science* **338**, 910–917 (2012).
92. Stirbat, T. V. et al. Fine tuning of tissues' viscosity and surface tension through contractility suggests a new role for α -catenin. *PLoS ONE* **8**, e52554 (2013).
93. Drasdo, D. Buckling instabilities of one-layered growing tissues. *Phys. Rev. Lett.* **84**, 4244–4247 (2000).
94. Basan, M., Joanny, J.-F., Prost, J. & Risler, T. Undulation instability of epithelial tissues. *Phys. Rev. Lett.* **106**, 158101 (2011).
95. Bozic, I. et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* **2**, e00747 (2013).
96. Stylianopoulos, T. & Jain, R. K. Combining two strategies to improve perfusion and drug delivery in solid tumors. *Proc. Natl Acad. Sci. USA* **110**, 18632–18637 (2013).
97. Foo, J. & Michor, F. Evolution of acquired resistance to anti-cancer therapy. *J. Theor. Biol.* **355**, 10–20 (2014).
98. Goldie, J. H. & Coldman, A. J. The genetic origin of drug resistance in neoplasms: implications for systemic therapy. *Cancer Res.* **44**, 3643–3653 (1984).
99. Coldman, A. J. & Goldie, J. H. A stochastic model for the origin and treatment of tumors containing drug-resistant cells. *Bull. Math. Biol.* **48**, 279–292 (1986).
100. Coldman, A. J. & Goldie, J. H. A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Math. Biosci.* **65**, 291–307 (1983).
101. Greulich, P., Waclaw, B. & Allen, R. J. Mutational pathway determines whether drug gradients accelerate evolution of drug-resistant cells. *Phys. Rev. Lett.* **109**, 088101 (2012).
102. Korolev, K. S. et al. Selective sweeps in growing microbial colonies. *Phys. Biol.* **9**, 026008 (2012).



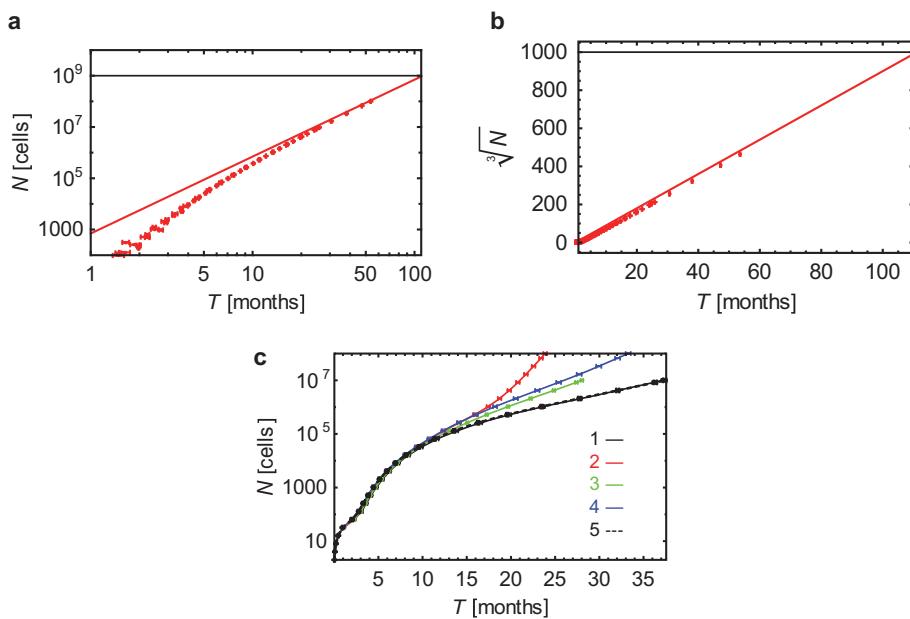
Extended Data Figure 1 | Details of the model. **a**, A sketch showing how dispersal is implemented: (1) A ball of cells of radius R_i , in which the centre is at X_i , is composed of tumour cells and normal cells (blue and empty squares in the zoomed-in rectangle (2)). A cell at position x_i with respect to the centre of the ball attempts to replicate (3). If replication is successful, the cell migrates with probability M and creates a new microlesion (4). The position X_j of this new ball of cells is determined as the endpoint of the vector that starts at X_i and has direction x_i and length R_i . **b**, Overlap reduction between the balls of cells. When a growing ball begins to overlap with another ball (red), they

are both moved apart along the line connecting their centres of mass (green line) by as much as necessary to reduce the overlap to zero. The process is repeated for all overlapping balls as many times as needed until there is no overlap. **c**, Summary of all parameters used in the model. If, for a given parameter, many different values have been used in different plots, a range of values used is shown. Birth and death rates can also depend on the number of driver mutations, see Methods. Asterisk, parameter estimated from other quantities available in the literature.



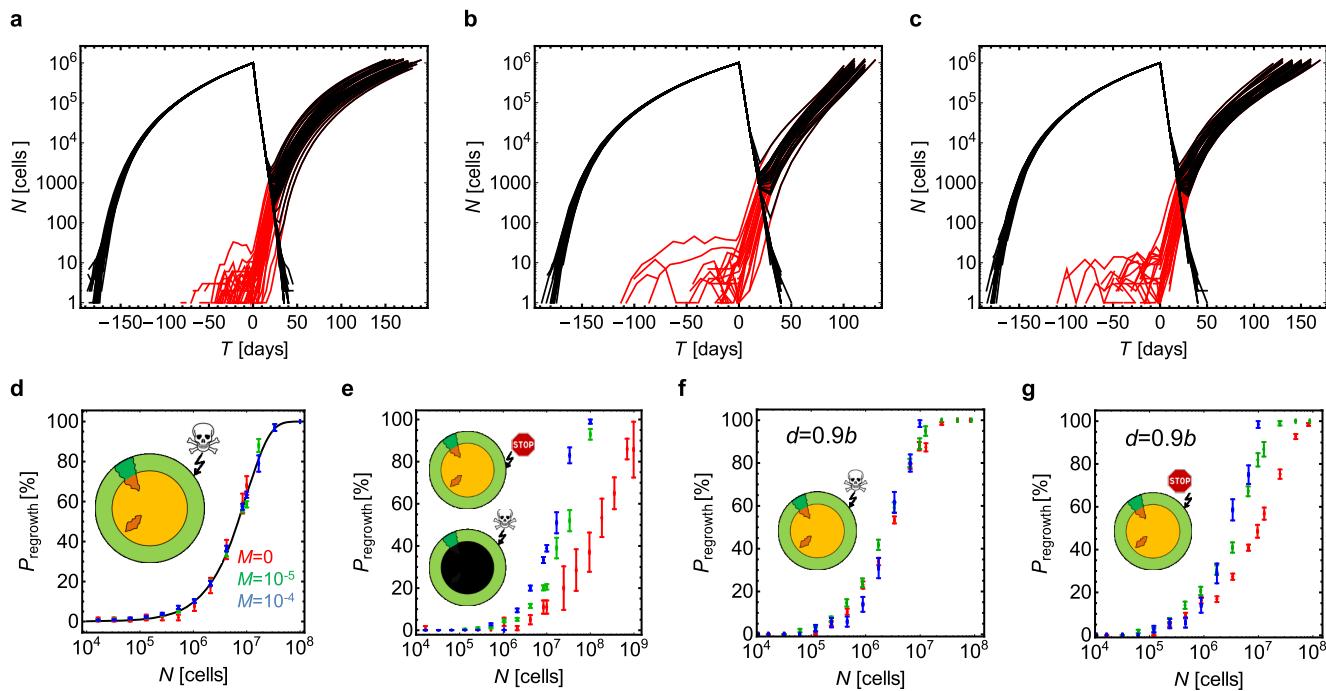
Extended Data Figure 2 | Simulation snapshots. **a, b,** A few snapshots of tumour growth for no dispersal, and $d = 0$ (a) and $d = 0.9b$ (b). To visualize clonal sectors, cells have been colour-coded by making the colour a heritable trait and changing each of its RGB components by a small random fraction whenever a cell mutates. The initial cell is grey. Empty space (white)

are non-cancer cells mixed with extracellular matrix. Note that images are not to scale. **c,** Tumour shapes for $N = 1 \times 10^7$, $d = 0.9b$, and different dispersal probability M . Images not to scale; the tumour for $M = 1 \times 10^{-5}$ is larger than the one for $M = 0$.



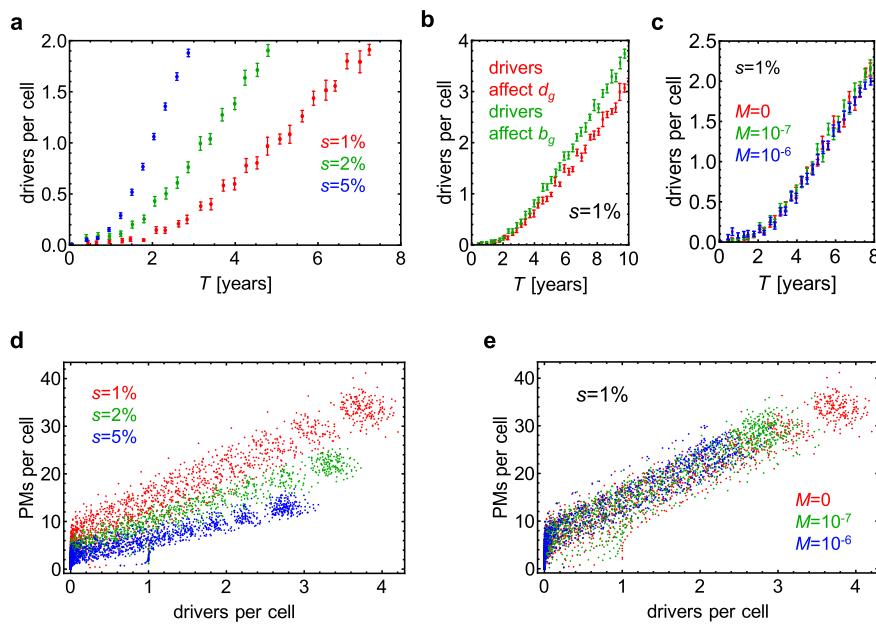
Extended Data Figure 3 | Tumour size as a function of time. **a**, Growth of a tumour without dispersal ($M = 0$), for $d = 0.8b$. For large times (T), the number of cells grows approximately as $\text{const} \times T^3$. The tumour reaches size $N = 1 \times 10^9$ cells (horizontal line) after about 100 months (8 years) of growth. **b**, The same data are plotted in the linear scale, with N replaced by ‘linear extension’ $N^{1/3}$. **c**, Tumour size versus time when drivers affect the dispersal probability. In all cases, $d = 0.9b$, and (1, black) drivers increase the dispersal rate tenfold ($q = 9$) but have no effect on the net growth rate; (2, red)

drivers increase both the net growth rate ($s = 10\%$) and M ; (3, green) drivers either (with probability 1/2) increase M tenfold ($q = 9$) or increase the net growth rate by $s = 10\%$; (4, blue) drivers increase only the net growth rate by $s = 10\%$; and (5, black dashed line) neutral case with $M = 1 \times 10^{-7}$, which is indistinguishable from (1). In all cases (1–3) the initial dispersal probability $M = 1 \times 10^{-7}$. Points represent average value over 40–100 simulations per data point, error bars are s.e.m.



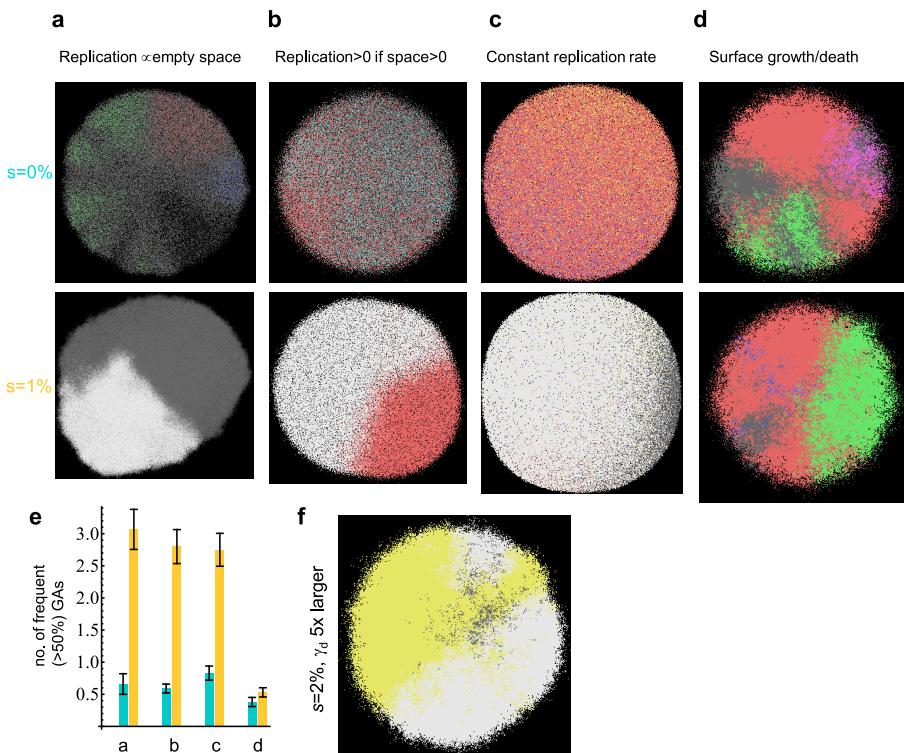
Extended Data Figure 4 | Simulation of targeted therapy. **a–c**, The total number of cells in the tumour (black) and the number of resistant cells (red) versus time, during growth ($T < 0$) and treatment ($T > 0$), for ~ 100 independent simulations, for $d = 0.5b$ for $T < 0$. Therapy begins when $N = 1 \times 10^6$ cells. After treatment, many tumours die out (N decreases to zero) but those with resistant cells will regrow sooner or later. **a**, $M = 0$ for all cells at all times. **b**, $M = 0$ for all cells for $T < 0$ and $M = 10^{-5}$ for resistant cells for $T > 0$. **c**, $M = 0$ for non-resistant and $M = 10^{-5}$ for resistant cells at all times. In all three cases, P_{regrowth} is very similar: $36 \pm 5\%$ (mean \pm s.e.m.) (**a**), $25 \pm 4\%$ (**b**), and $27 \pm 6\%$ for (**c**). **d–g**, Regrowth probability for four treatment scenarios not discussed in the main text. Data points correspond to three dispersal probabilities: $M = 0$ (red), $M = 1 \times 10^{-5}$ (green), and $M = 1 \times 10^{-4}$ (blue).

The probability is plotted as a function of tumour size N just before the therapy commences. **d**, Before treatment, cells replicate only on the surface. Cells in the core are quiescent and do not replicate. Therapy kills cells on the surface and cells in the core resume proliferation when liberated by treatment. **e**, As in **d**, but drug is cytostatic and does not kill cells but inhibits their growth. The results for P_{regrowth} are identical if the drug is cytotoxic and the tumour has a necrotic core (cells die inside the tumour and cannot replicate even if the surface is removed). **f**, Before treatment, cells replicate and die on the surface. The core is quiescent. Therapy kills cells on the surface (cytotoxic drug). **g**, As in **f**, but therapy only inhibits growth (cytostatic drug). In all cases (**d–g**) error bars represent s.e.m. from 8–1,000 simulations per point.



Extended Data Figure 5 | Accumulation of driver and passenger genetic alterations. **a–c,** The number of drivers per cell in the primary tumour plotted as a function of time (10–100 simulations per point, error bars denote s.e.m.). **a,** $M = 0$ and three different driver selective advantages. For $s = 1\%$, cells accumulate on average one driver mutation within 5 years. The time can be significantly lower for very strong drivers ($s > 1\%$). **b,** The rate at which drivers accumulate depends mainly on their selective advantage and not on whether

they affect death or birth rate. **c,** Dispersal does not affect the rate of driver accumulation. **d, e,** The number of passenger mutations (PMs) per cell versus the number of driver mutations per cell. More passenger mutations are present for smaller driver selective advantage (**d**), and this is independent of the dispersal probability M (**e**) in the regime of small M . Data points correspond to independent simulations.

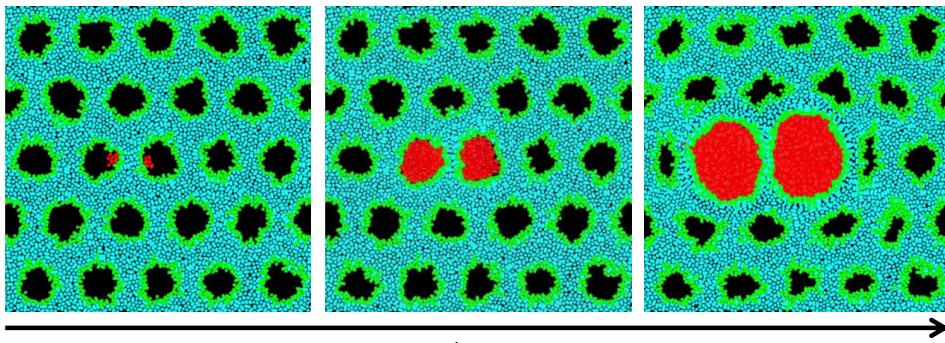
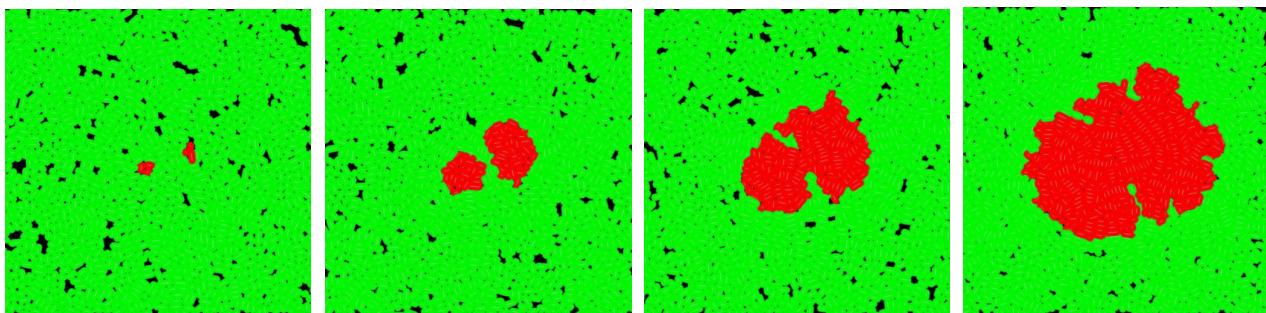
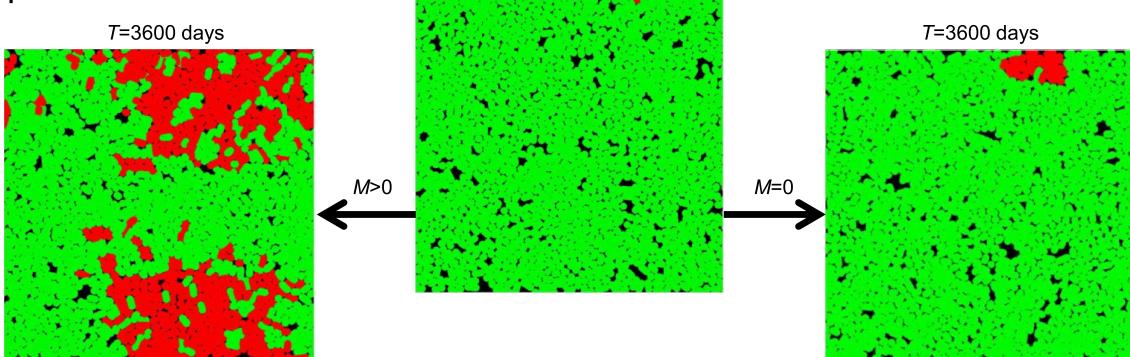


Extended Data Figure 6 | Genetic diversity in a single lesion for different models. **a–d**, Representative simulation snapshots, with genetic alterations colour-coded as in Fig. 4. Top: $s = 0$, bottom: $s = 1\%$. **a**, Model A from the main text in which cells replicate with rates proportional to the number of empty nearby sites. **b**, Model B, the replication rate is constant and non-zero if there is at least one empty site nearby, and zero otherwise. **c**, Model C, cells replicate at a constant rate and push away other cells to make space for their progeny. **d**, Model D, cells replicate/die only on the surface, the interior of the tumour

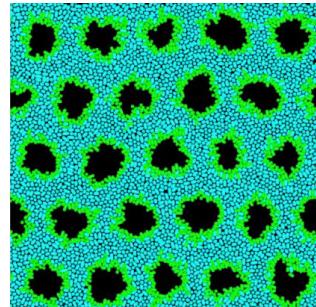
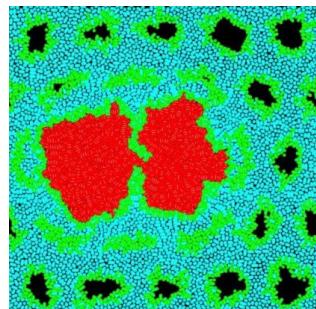
(‘necrotic core’) is static. In all cases, $N = 1 \times 10^7$, $d = 0.99$. **e**, Number of genetic alterations present in at least 50% of cells for identical parameters as in **a–d**. In all cases except surface growth (**d**), drivers increase genetic homogeneity, as measured by the number of most frequent genetic alterations. Results averaged over 50–100 simulations, error bars denote s.e.m. **f**, Model D, with $\gamma_d = 2 \times 10^{-4}$ instead of 4×10^{-5} , that is, drivers occur five times more often. In this case, driver mutations arise earlier than in **d**, and the tumour becomes more homogeneous.

a

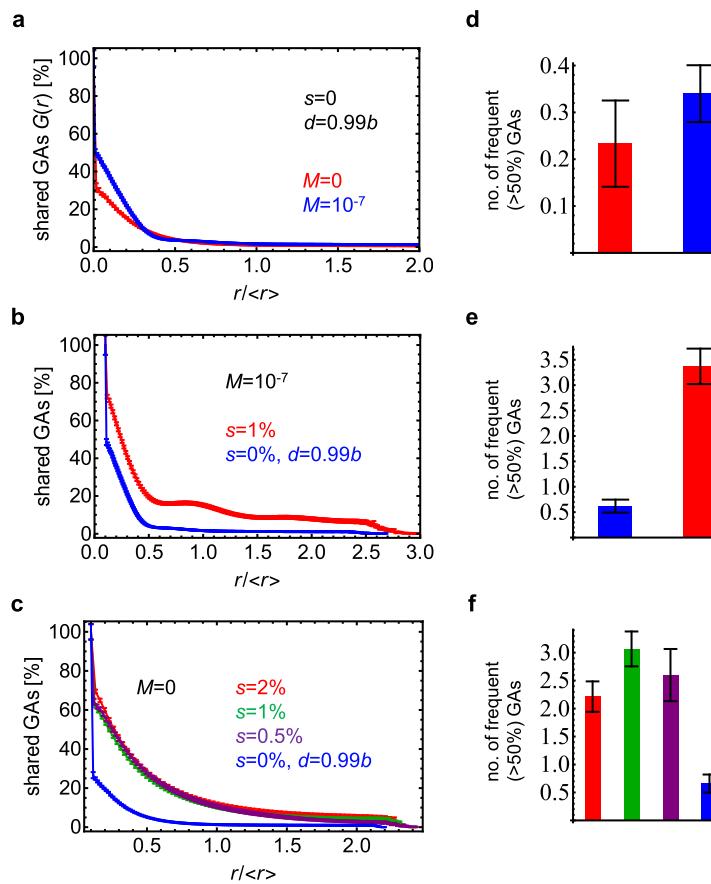
Parameter	Meaning	Value	References
L	Length of the simulation box	400–800 μm	This work
d	Cell diameter	10 μm (*)	[59,60,86]
g	Elongation speed	0.208 [$\mu\text{m}/\text{h}$] (†)	This work
μ	Dynamic viscosity of the intracellular fluid	$2 \cdot 10^5 \text{ Pa}\cdot\text{s}$	[86]
E_{ec}	Young modulus of epithelial/cancer cells	1 kPa (*)	[86]
E_s	Young modulus of stroma	1 kPa (‡)	[87, 88]
ν	Poisson's ratio	1/3 (*)	[59,60,86]
σ	Cell-cell adhesion energy	200 $\mu\text{J}/\text{m}^2$ (*)	[86]

c**e****f**

Extended Data Figure 7 | The off-lattice model. **a**, Summary of all parameters used in the model. Asterisk, typical value, varies between different types of tissues; dagger symbol, equivalent to 24 h minimal doubling time; double dagger symbol, based on the assumption that macroscopic elastic properties of tissues such as liver, pancreases or mammary glands are primarily determined by the elastic properties of stroma. **b**, Simulation snapshot of a normal tissue before the invasion of cancer cells. **c**, Two balls of cancer cells in

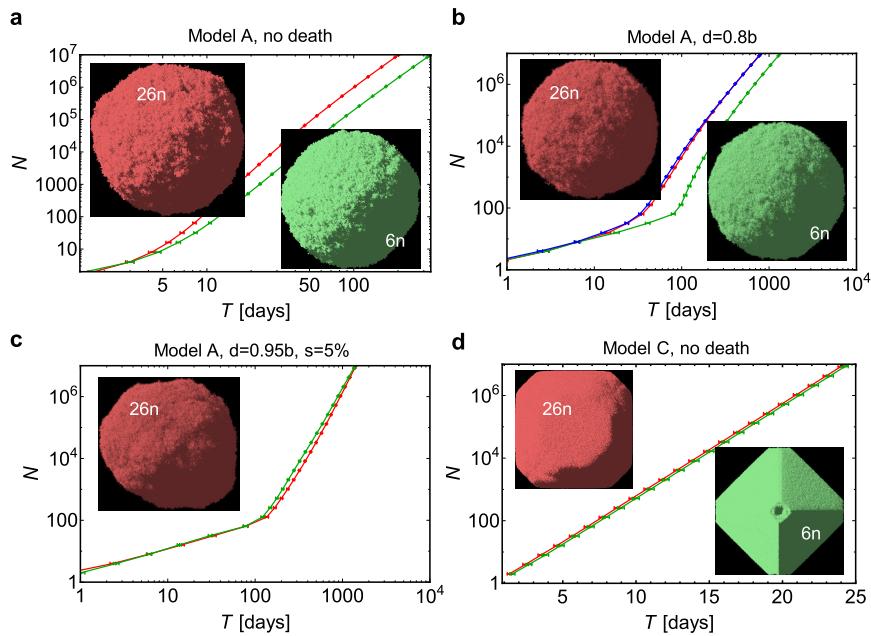
b**d**

two nearby ducts repel each other as they grow as a consequence of mechanical forces exerted on each other. **d**, The balls coalesce if growth is able to break the separating extracellular matrix. **e**, If the balls are not encapsulated, they quickly merge. **f**, Isolated balls of cells are not required to speed up growth; migration (left) can cause the tumour to expand much faster even if individual microlesions merge together, as opposed to the case with no migration (right).



Extended Data Figure 8 | Genetic diversity quantified. **a**, Tumours are much more genetically heterogeneous in the absence of driver mutations ($s = 0$) (see Fig. 4). The plot shows the fraction $G(r)$ of genetic alterations (GAs) shared between the cells as function of their separation (distance r) in the tumour. The fraction quickly decreases with increasing r . The distance in the figure is normalized by the average distance $\langle r \rangle$ between any two cells in the tumour. For a spherical tumour, $\langle r \rangle$ is approximately equal to half of the tumour diameter. **b**, Fraction of shared genetic alterations for $s = 1\%$ and $s = 0\%$,

$N = 1 \times 10^7$, and $M = 1 \times 10^{-7}$. In the presence of drivers, $G(r)$ decays slower, indicating more homogeneous tumours. **c**, The exact value of the selective advantage of driver mutations is not important (all curves $G(r)$ look the same, except for $s = 0$) as long as $s > 0$. **d–f**, Number of genetic alterations present in at least 50% of cells for identical parameters as in **a–c**, correspondingly. Drivers substantially increase the level of genetic homogeneity. In all panels the results have been averaged over 30–100 simulations, with error bars as s.e.m.



Extended Data Figure 9 | Growth curves for the 26-nearest neighbours (26n, red curves) and the 6-nearest neighbours (6n, green curves) models. **a**, Model A (as in the main text), no death. The tumour grows about twice as slow in the 6n model. Pictures show tumour snapshots for both models; there is no visible difference in the shape. **b**, Model A, death $d = 0.8b$. The additional blue curve is for the 6n model, with modified replication probability to account for missing neighbours as explained in the Supplementary Information. **c**, Model A, with death $d = 0.95b$, and drivers $s = 5\%$. There is

very little difference in the growth curves between the 6n and 26n models. A small asymmetry in the shape is caused by faster-growing cells with driver mutations. **d**, Model C (exponential growth). Growth is the same in both 6n and 26n models, but the shape is more aspheric for the 6n model. This is probably caused by shifting cells along the coordinate axes and not along the shortest path to the surface when making space for new cells. All plots show the mean (average over 50–100 simulations) and s.e.m.

Extended Data Table 1 | Experimental results for the percentage of proliferating cells in the centre versus the edge of solid tumours

Case	Tumor type	Edge			Center			Ratio center:edge	p-value
		Images	KI %	Total no. of cells	Images	KI %	Total no. of cells		
1	Chromophobe RCC	8	3.46	1013	3	1.11	2561	0.32	0.05
2	Chromophobe RCC	5	3.07	508	3	1.05	938	0.34	0.28
3	Chromophobe RCC	5	2.63	524	3	0.44	697	0.17	0.03
4	Chromophobe RCC	7	1.58	581	2	0.53	958	0.34	0.17
5	HCC	7	17.14	892	2	9.74	1637	0.57	0.05
6	HCC	7	51.71	1079	4	32.84	2562	0.64	0.03
7	HCC	6	47.37	435	3	19.97	1397	0.42	0.09
8	HCC	7	19.02	895	4	13.78	1191	0.72	0.35
9	HCC	6	15.09	1074	3	11.98	1094	0.79	0.33
10	HCC	9	29.84	1305	2	20.87	2457	0.70	0.22
Summary									
1-4	Chromophobe RCC	25	2.69	—	11	0.81	—	0.30	0.00002
5-10	HCC	42	30.0	—	18	19.1	—	0.64	0.007

A representative section of each tumour was labelled for the proliferation marker Ki67 (KI), and images of the tumour at the leading edge and the centre were acquired as described (Methods). Proliferation is markedly increased at the leading edge, and this is statistically significant ('Summary', Kolmogorov–Smirnov two-sample test, $P < 0.05$). The average ratio of the number of proliferating cells in the centre/at the edge is 0.50 (range 0.17–0.79). HCC, hepatocellular carcinoma; RCC, renal cell carcinoma