

Spatial patterns of tumour growth impact clonal diversification: computational modelling and evidence in the TRACERx Renal study

Xiao Fu

The Francis Crick Institute <https://orcid.org/0000-0002-8611-891X>

Yue Zhao

The Francis Crick Institute

Jose Lopez

Cruces University Hospital

Andrew Rowan

The Francis Crick Institute

Lewis Au

The Francis Crick Institute

Annika Fendler

The Francis Crick Institute

Steve Hazell

The Royal Marsden NHS Foundation Trust

Hang Xu

Stanford Cancer Institute <https://orcid.org/0000-0002-5847-655X>

Stuart Horswell

The Francis Crick Institute

Scott Shepherd

The Francis Crick Institute <https://orcid.org/0000-0001-6208-173X>

Lavinia Spain

The Royal Marsden Hospital NHS Foundation Trust

Fiona Byrne

The Francis Crick Institute <https://orcid.org/0000-0003-3800-0500>

Gordon Stamp

The Francis Crick Institute

Tim O'Brien

Guy's & St Thomas' NHS Foundation Trust

David Nicol

The Royal Marsden NHS Foundation Trust

Marcellus Augustine

The Francis Crick Institute

Ashish Chandra

Guy's & St Thomas' NHS Foundation Trust

Sarah Rudman

Guy's & St. Thomas' NHS Foundation Trust

Antonia Toncheva

Guy's and St. Thomas' NHS Foundation Trust

Andrew Furness

The Francis Crick Institute

Lisa Pickering

The Royal Marsden Hospital NHS Foundation Trust <https://orcid.org/0000-0002-7579-340X>

Matthew Orton

Institute of Cancer Research

Simon Doran

Institute of Cancer Research <https://orcid.org/0000-0001-8569-9188>

Dow-Mu Koh

Institute of Cancer Research

Christina Messiou

Institute of Cancer Research

Derfel ap Dafydd

Institute of Cancer Research

Santosh Kumar

Institute of Cancer Research

James Larkin

The Royal Marsden Hospital NHS Foundation Trust

Charles Swanton

The Francis Crick Institute <https://orcid.org/0000-0002-4299-3018>

Erik Sahai

The Francis Crick Institute

Kevin Litchfield

The Francis Crick Institute

Samra Turajlic

Francis Crick Institute <https://orcid.org/0000-0001-8846-136X>

Paul Bates (✉ paul.bates@crick.ac.uk)

The Francis Crick Institute <https://orcid.org/0000-0003-0621-0925>

Article

Keywords: Computational modelling, Cellular Automaton, Clear cell renal cell carcinoma, Spatial growth 3 patterns, Microdiversity, Parallel evolution, Time-course analysis

Posted Date: February 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-244285/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Ecology & Evolution on December 23rd, 2021. See the published version at <https://doi.org/10.1038/s41559-021-01586-x>.

1 Title:

2 Spatial patterns of tumour growth impact clonal diversification: computational modelling and
3 evidence in the TRACERx Renal study

4 Authors:

5 Xiao Fu^{1*}, Yue Zhao^{2,3,4,5*}, Jose I. Lopez⁶, Andrew Rowan², Lewis Au^{7,16}, Annika Fendler⁷,
6 Steve Hazell⁸, Hang Xu², Stuart Horswell⁹, Scott Shepherd^{7,16}, Lavinia Spain⁷, Fiona Byrne⁷,
7 Gordon Stamp¹⁰, Tim O'Brien¹¹, David Nicol¹², Marcellus Augustine², Ashish Chandra¹³,
8 Sarah Rudman¹⁴, Antonia Toncheva¹⁵, Andrew Furness^{7,16}, Lisa Pickering¹⁶, Matthew Orton¹⁷,
9 Simon Doran¹⁷, Dow-Mu Koh¹⁷, Christina Messiou¹⁷, Derfel ap Dafydd¹⁷, Santosh Kumar¹⁷,
10 James Larkin¹⁶, Charles Swanton^{2,3,18}, Erik Sahai^{19#}, Kevin Litchfield^{2,3#}, Samra Turajlic^{7,16#}
11 on behalf of the TRACERx Renal Consortium, Paul A Bates^{1#}

12
13 1. Biomolecular Modelling Laboratory, The Francis Crick Institute, 1 Midland Rd, London
14 NW1 1AT, UK

15 2. Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, 1 Midland
16 Rd, London NW1 1AT, UK

17 3. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer
18 Institute, Paul O’Gorman Building, 72 Huntley Street, London, WC1E 6BT, UK

19 4. Department of Thoracic Surgery, Fudan University Shanghai Cancer Center, Shanghai,
20 China. 200032.

21 5. Department of Oncology, Shanghai Medical College, Fudan University, Shanghai,
22 China. 200032.

23 6. Department of Pathology, Cruces University Hospital, Biocruces-Bizkaia Institute, 48903
24 Barakaldo, Bizkaia, Spain

25 7. Cancer Dynamics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1
26 1AT, UK

27 8. Department of Pathology, the Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK

28 9. Department of Bioinformatics and Biostatistics, The Francis Crick Institute, 1 Midland Rd,
29 London NW1 1AT, UK

30 10. Experimental Histopathology Laboratory, The Francis Crick Institute, London NW1 1AT,
31 UK

32 11. Urology Centre, Guy’s and St. Thomas’ NHS Foundation Trust, London SE1 9RT, UK

33 12. Department of Urology, the Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK

34 13. Department of Pathology, Guy’s and St. Thomas NHS Foundation Trust, London SE1 9RT,
35 UK

36 14. Department of Medical Oncology, Guy’s and St. Thomas’ NHS Foundation Trust, London
37 SE1 9RT, UK

38 15. Biobank, Guy’s and St. Thomas’ NHS Foundation Trust, London SE1 7EH, UK

39 16. Renal and Skin Units, The Royal Marsden Hospital, London, SW3 6JJ, UK

40 17. Division of Radiotherapy and Imaging, Institute of Cancer Research, 15 Cotswold
41 Road, Sutton, Surrey, SM2 5NG

42 18. Department of Medical Oncology, University College London Hospitals, 235 Euston Rd,
43 Fitzrovia, London, NW1 2BU, UK

44 19. Tumour Cell Biology Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1
45 1AT, UK

1 Key words:

2 Computational modelling; Cellular Automaton; Clear cell renal cell carcinoma; Spatial growth
3 patterns; Microdiversity; Parallel evolution; Time-course analysis

4

5

1 Abstract:

2

3 Intra-tumour genetic heterogeneity (ITH) fuels cancer evolution. The role of clonal diversity
4 and genetic complexity in the progression of clear-cell renal cell carcinomas (ccRCCs) has
5 been characterised, but the ability to predict clinically relevant evolutionary trajectories
6 remains limited. Here, towards enhancing this ability, we investigated spatial features of clonal
7 diversification through a combined computational modelling and experimental analysis in the
8 TRACERx Renal study. We observe through modelling that spatial patterns of tumour growth
9 impact the extent and trajectory of subclonal diversification. Moreover, subpopulations with
10 high clonal diversity, and parallel evolution events, are frequently observed near the tumour
11 margin. In-silico time-course studies further showed that budding structures on the tumour
12 surface could indicate future steps of subclonal evolution. Such structures were evident
13 radiologically in 15 early-stage ccRCCs, raising the possibility that spatially resolved sampling
14 of these regions, when combined with sequencing, may enable identification of evolutionary
15 potential in early-stage tumours.

16

17

1 Introduction:

2 The development of cancer can be viewed as an evolutionary process (Merlo et al. 2006, Zahir
3 et al. 2020). Acquisition of genomic alterations including mutations and somatic copy-number
4 alterations (SCNAs) drives the emergence of genetically heterogeneous subpopulations of
5 cancer cells or subclones (McGranahan and Swanton 2017), resulting in intra-tumour
6 heterogeneity (ITH). A small subset of genomic alterations (drivers) endows subclones with
7 increased fitness that manifests as growth and survival. Subclones compete for resources,
8 including physical space, and undergo expansion or extinction according to their fitness under
9 given selective pressures imposed by the tumour microenvironment (TME) or therapeutic
10 intervention. With recent advances in next-generation sequencing, clonal architecture and
11 evolutionary features have been elucidated in a variety of tumour types (Gerlinger et al 2012,
12 Yates et al 2015, Jamal-Hanjani et al 2017, Turajlic et al 2018). However, the ability to predict
13 clinically relevant evolutionary trajectories remains limited.

14
15 One potential to enhance this ability lies in the detection and characterisation of ongoing clonal
16 evolution. ITH provides a substrate for the selection of competent clones under evolving TME
17 and frequently underpins therapeutic failure (McGranahan and Swanton, 2015). Detection of
18 ITH relies on proper selection of samples and can be enhanced by sampling many small-size
19 regions (Lopez and Cortes, 2016). While ITH index and macrodiversity (i.e., the number of
20 subclones in the whole tumour) reflect the overall degree of established clonal diversity and
21 genetic complexity within a tumour, clone diversity at a narrow spatial scale, or microdiversity
22 (i.e., the number of subclones within a single tumour sample), could represent under-detected
23 ongoing clonal evolution and delineate future evolutionary trajectories (**Figure 1a**). Just like
24 macrodiversity, microdiversity appears to have clinical implications and has been shown to
25 underpin poor survival in paediatric kidney cancer (Mengelbier et al 2015). Microdiversity was
26 also revealed by the presence of multiple subclones both *in situ* and in co-migrating cell groups
27 during tumour invasion in breast tumours, using topographic single cell sequencing (Casasent
28 et al. 2018). Clonal diversification could be facilitated by parallel evolution, that is selection of
29 distinct mutations in the same gene in spatially separate subclones. Parallel evolution has been
30 observed in the natural evolution of various tumour types (Gerlinger et al 2012, Turajlic et al
31 2018, Melchor et al. 2014, Murugaesu et al. 2015, Yates et al 2015) and in the context of
32 therapeutic resistance under selective pressure of therapy (Voss et al 2014, Juric et al 2015).
33 More recently, parallel evolution of SCNAs was characterised comprehensively through
34 mirrored subclonal allelic imbalance (Watkins et al. 2020). As microdiversity and parallel
35 evolution are important indicators of ongoing and clinically relevant tumour evolutionary
36 features, their spatial resolution could offer further understanding of evolutionary potential.

37
38 To understand the emergence and spatial patterning of these features, we developed a coarse-
39 grained cellular automata model of tumour growth and clonal dynamics (**Figure 1b**). Previous
40 experimental evidence supports predominating proliferation at the surface of ccRCCs (Hoefflin
41 et al. 2016). However, both some cases in this study and our recent work suggest that this may
42 not be universally true (Zhao et al. 2020). As exemplified by several categories of mathematical
43 models formulated to characterise tumour growth dynamics (Rodriguez-Brenes et al. 2013,
44 Gerlee 2013), different growth patterns could have an impact on evolutionary features.
45 Moreover, previous computational studies found an impact of growth patterns on the
46 classification of neutral evolution and selection (Sun et al. 2017, Chkhaidze et al. 2019).
47 Therefore, we further evaluated the effects of different spatial patterns of tumour growth
48 (hereafter also referred to as tumour growth mode) on spatial features of clonal diversity in our
49 computational model. While distinct growth patterns may predominate in different parts of a

1 tumour, or at different stages of the same tumour growth, we restricted our investigation in the
2 current study to two simple growth models: one model with uniform growth throughout the
3 tumour volume as a “null model” assuming that all tumour mass can proliferate (referred to as
4 “Volume Growth Model”) and the other model with active proliferation restricted to the tumour
5 surface (referred to as “Surface Growth Model”) (see Methods, **Figure 1c**).

6
7 In the context of the TRACERx Renal study, we reported a workflow to evaluate both the
8 genomic profiles and spatial coordinates of 756 regions (patient tumour regions are referred to
9 as “PT regions” hereafter) from 66 tumours (Zhao et al. 2020). Here, we focus on spatial
10 analyses of microdiversity and parallel evolution in the same dataset. We found in the model
11 setting that the spatial patterns of tumour growth influenced the extent and trajectory of
12 subclonal diversification. Furthermore, subpopulations with high microdiversity, as well as
13 parallel mutational events with limited clonal expansion, were frequently detected near the
14 tumour edge, which was corroborated in the TRACERx Renal dataset. Finally, time-course
15 studies on simulated tumours suggested that evolutionary steps could be predicted at an early
16 stage and seeded by clones in the budding structures on the tumour surface in a subset of
17 tumours. These budding structures were evident in 15 early-stage ccRCC tumours, raising the
18 intriguing possibility that spatially resolved sampling, when combined with sequencing, may
19 enable identification of evolutionary potential in early-stage tumours and predict the disease
20 course.

21

1 Results:

2 Generation of an agent-based model recapitulating ccRCC evolution

3 We developed a coarse-grained cellular automaton model to explore the evolutionary dynamics
4 of ccRCCs, a brief outline for which is presented here (see Methods for detailed description).
5 The model includes 12 driver genes and 14 SCNAs (**Supplemental Figure 1**) that were
6 highlighted as canonical driver events in ccRCCs in the TRACERx Renal study (Turajlic et al.
7 2018). Each model unit, referred to as a “tumour voxel”, represents a tumour volume of 1 mm^3 .
8 Tumour voxels stochastically undergo growth, death, and, upon growth, processes of driver
9 acquisition (i.e., mutations in driver genes or driver SCNAs). To keep the model minimal,
10 several assumptions were made. In the model, time has arbitrary units, so the baseline growth
11 probability $p_{growth}^{(1)}$ was arbitrarily defined. Three levels of growth probability, $p_{growth}^{(1)} \leq$
12 $p_{growth}^{(2)} \leq p_{growth}^{(3)}$, were introduced for the 26 drivers, to broadly reflect their association with
13 the Ki67 score in tumour regions in the experimental data (**Supplemental Figure 1**). As one
14 specific implementation, growth probabilities were set at $p_{growth}^{(1)} = 0.25$, $p_{growth}^{(2)} = (1 +$
15 $s)p_{growth}^{(1)}$ and $p_{growth}^{(3)} = (1 + s)^2 p_{growth}^{(1)}$, where $0 \leq s \leq 1$ determines this advantage
16 relative to the baseline growth probability. For simplicity, mutations in all driver genes were
17 assigned with the baseline growth probability, while some SCNAs were assumed to confer
18 greater growth probabilities. Four SCNAs with strong association with Ki67 score, 7q gain,
19 20q gain, 4q loss, and 8p loss, are the most “advantageous” drivers with $p_{growth}^{(3)}$. A tumour
20 voxel grows with $p_{growth}^{(1)}$ conferred by the most advantageous driver harboured. Note that an
21 alternative implementation for fitness advantage was also explored (**Supplemental Figure 2**,
22 **Supplemental Note 1**, see Methods). Mutations in driver genes were assumed to be acquired
23 with a greater probability (p_{driver}) than SCNAs. Based on functional evidence linking them to
24 chromosome instability (Varela et al. 2011, Peng et al. 2015), as well as their association with
25 a high weighted genome instability index (wGII) in the experimental data from the TRACERx
26 Renal study (Turajlic et al. 2018), mutations in *PBRM1* or *BAP1*, were assumed to enhance the
27 probability of SCNA acquisition. Finally, a second mutation in the same gene was assumed to
28 never occur in the same tumour voxel.

29

30 Each simulation starts from a single tumour voxel carrying *VHL* and 3p loss and is terminated
31 when the size of the simulated tumour exceeds 1 million tumour voxels. Simulated tumours
32 were analysed at three spatial scales: three-dimensional (3D) whole tumour volume, two-
33 dimensional (2D) tumour slice and tumour regions (**Figure 1d**). Exploratory simulations (50
34 per condition) were performed at 6 levels of s and 11 levels of p_{driver} . This permits the
35 understanding of the collective impact of fitness advantage and the rate of driver acquisition in
36 a spatial context.

37

38 Characterisation of clonal diversity in the whole tumour

39 Tumours under Volume Growth commonly developed a single dominant subclone that grew
40 to occupy a large proportion of tumour surface (**panel (i) in Figure 1e**). In comparison, tumours
41 under Surface Growth developed multiple advantageous subclones in different regions of the
42 tumour surface, and each of these subclones was characterised by a range of different driver
43 events (**panel (ii) in Figure 1e**). Moreover, while the tumour surface in the Volume Growth

1 model was smooth overall, the tumour under Surface Growth formed bulging structures on the
2 surface, which reflected the outgrowth of driver subclones.

3
4 Given the disparity in the quantity and spatial distribution of driver subclones in the
5 representative cases under Volume Growth and Surface Growth (**Figure 1e**), we then asked
6 how the mode of tumour growth might influence the number of subclones in the whole tumour.
7 We first counted the number of clones (including parental clone and subclones) in the whole
8 tumour (**Figure 2a-b**). In the Volume Growth model, subclones were only observed in tumours
9 with greater s and greater p_{driver} (**panel (i) in Figure 2b**). By contrast, in the Surface Growth
10 model, tumours with small to moderate p_{driver} harboured more subclones, for a wide range of
11 s (**panel (ii) in Figure 2b**). These findings suggest that the mode of tumour growth could
12 impact outgrowth of subclones and their ultimate prevalence.

13
14 Noticeably, tumours also displayed distinct proportions of subclonal populations, independent
15 of the growth mode and evolutionary conditions (i.e., identical s and p_{driver}) (**Figure 2c**). In
16 the Volume Growth model, most tumours with a small p_{driver} displayed a monoclonal
17 structure with limited evidence of clonal evolution (**panel (i) in Figure 2c**). With a large
18 p_{driver} , a greater extent of clonal evolution was evident in a higher proportion of tumours
19 (**panel (ii) in Figure 2c**). In a fraction of these cases, one dominant subclone was observed
20 with the parental clone present as a minority, suggesting early fixation of a highly fit clone and
21 a (near) clonal sweep (i.e., the entire tumour was taken up by the dominant subclone). In
22 comparison, in the Surface Growth model, extensive subclonal diversification was evident in
23 nearly all cases, even with a small p_{driver} (**panel (iii) in Figure 2d**). These features observed
24 in individual tumours were reflected by the whole-tumour cancer cell fraction (CCF) of the
25 parental clone (**Figure 2e**) and Shannon diversity index (**Figure 2f**). Across a wide range of
26 driver acquisition probabilities, the Surface Growth model, in comparison to the Volume
27 Growth model, consistently displayed a relatively lower whole-tumour CCF of the parental
28 clone and a greater Shannon diversity index, indicating a greater extent of subclonal
29 diversification. The greater extent of diversification in the Surface Growth model was also
30 noted for conditions with still smaller p_{driver} (**Supplemental Figure 3**) or smaller s
31 (**Supplemental Figure 4**), where clonal evolution was generally limited. For the interest of
32 characterising patterns of subclonal diversification and contrasting two growth modes, we
33 limited our parameter analysis, for the subsequent investigation, to $s = 1$ and a range of p_{driver}
34 from 2×10^{-4} to 1×10^{-3} .

35
36 In general, Surface Growth appears to enable more extensive subclonal diversification, leading
37 to a highly branched tumour evolution, while Volume Growth often resulted in tumours with
38 limited clonal evolution and occurrence of punctuated evolution. Interestingly, these were the
39 precise modes of evolution that underlined diverse evolutionary subtypes identified in ccRCCs,
40 ranging from “*VHL* mono-driver” tumours with limited evidence of clonal evolution, to
41 tumours with extensive subclonal diversification characterised by a range of drivers, and those
42 with early fixation of a highly fit clone resulting in rapid clonal sweep (Turajlic et al. 2018).

43
44 Having established a quantitative understanding of clonal diversity in the whole tumour and
45 the impact of different growth modes, we next strived for a spatial understanding of clonal
46 diversity.

1 Evaluation of genomic divergence between tumour regions

2 As a first step towards understanding spatial features of clonal diversity, we asked whether *in-*
3 *silico* tumours were able to recapitulate a phenomenon of spatial clonal evolution noted in
4 colorectal and liver cancers by others (Cross, et al. 2017; Zhai, et al. 2017) and, more recently,
5 in ccRCCs by our group (Zhao et al. 2020), whereby subclones are patterned within a tumour
6 such that regions farther apart exhibit more divergent genomic makeups.

7
8 To this end, we sampled uniformly spaced regions within a 2D tumour slice and measured the
9 spatial and genomic distances between each pair of regions (**Figure 3a**, see Methods).
10 Consistent with previous experimental evidence (Cross, et al. 2017; Zhai, et al. 2017; Zhao, et
11 al. 2020), between-region spatial and genomic distances were positively correlated (hereafter
12 referred to as “spatial-genomic correlation”) in representative tumours under Surface Growth
13 (**Figure 3b-c**) or under Volume Growth (**Figure 3d-e**) and also in repeat simulations across
14 multiple values of p_{driver} (**Figure 3f**). Moreover, different growth modes resulted in distinct
15 trends of variation in the spatial-genomic correlation with the change of p_{driver} . In the Volume
16 Growth model, the spatial-genomic correlation was weaker in tumours with smaller p_{driver} .
17 By contrast, in the Surface Growth model, the spatial-genomic correlation was weaker in
18 tumours with greater p_{driver} , as the extensive diversification likely caused large genomic
19 divergence even in regions separated by a short distance (**Figure 3f**). As a consequence of these
20 divergent trends, the spatial-genomic correlation was stronger in tumours under Surface
21 Growth than those under Volume Growth if p_{driver} was small, and weaker if p_{driver} was large.
22 Experimentally, the cohort of ccRCCs in the TRACERx Renal study displayed a spatial-
23 genomic correlation coefficient of 0.31 (Zhao, et al. 2020).

24
25 Overall, these data demonstrate that our model is sufficient to recapitulate the spatial-genomic
26 correlations observed across experimental data, a pattern emergent from spatial evolution of
27 subclones in tumours. We next asked how clonal diversity itself could vary spatially within a
28 tumour slice and whether the mode of tumour growth had an impact on the spatial patterns.

29 30 Spatial distribution of clonal diversity within tumour sections

31 Characterisation of clonal diversity in the whole tumour and the between-region genomic
32 divergence provided a quantitative view of the extent of subclonal diversification. We next
33 sought to examine whether clonal diversity was spatially uniform or variable within a tumour.
34 To this end, a 3mm-by-3mm sliding window was applied throughout the 2D tumour slice and
35 the number of distinct subclones (defined as “microdiversity”) was counted in that 9- mm^2 area
36 (**Figure 4a**). Subsequently, the spatial profile of microdiversity was analysed to identify
37 hotspots with microdiversity equal to or greater than 5 (defined as “microdiversity hotspots”)
38 (**Figure 4a**).

39
40 Consistent with the previous analysis on the whole-tumour CCF of subclones (**Figure 2a**),
41 tumours under Surface Growth developed multiple subclones that occupied distinct spatially
42 contiguous areas, reflective of a branched clonal structure (**panel (i) in Figure 4b**).
43 Microdiversity hotspots were frequently observed near the tumour edge within outgrowing
44 advantageous subclones but also at the boundaries of multiple adjacent subclones (**panel (ii)**
45 **in Figure 4b**). By contrast, tumours under volume growth commonly developed a single
46 dominant subclone (**panel (i) in Figure 4c**). In these tumours microdiversity hotspots were

1 observed within the area spanned by the dominant subclone with a more uniform distribution
2 along the tumour radius (**panel (ii) in Figure 4c**).

3 To corroborate these spatial features of clonal diversity with experimental data, we examined
4 66 tumours in the TRACERx Renal study. Individual PT regions that contain at least two
5 subclones were treated as a proxy for microdiversity hotspots. In total, 606 PT regions from 54
6 tumours were included in this analysis. Different spatial distributions of microdiversity
7 hotspots were observed, as highlighted in representative examples: predominantly near the
8 tumour margin (as in “G_K234”), while more uniformly distributed throughout the tumour
9 section (as in “G_K446”) (**Figure 4d**).

10
11 In order to quantify the observed spatial patterns, we measured the distance to the tumour centre
12 and the distance to the closest point at the tumour contour, for every microdiversity hotspot
13 over all repeat simulations. These two distances were then combined into a single variable
14 termed the “normalised distance to tumour centre”, denoted as d . Interestingly, the cumulative
15 probability distribution with respect to d depicted power law scaling (**Figure 4e**), suggesting
16 that the probability of observing spots with high clonal diversity along the radius of a tumour
17 could be estimated using a simple mathematical formula (i.e., $P(D \leq d) \sim d^k$, where k is the
18 power law exponent to be fitted). Furthermore, the Surface Growth model displayed a larger k
19 compared to the Volume Growth model across multiple values of p_{driver} (**Figure 4f**,
20 **Supplemental Figure 5**), indicating a greater likelihood of microdiversity hotspots being
21 located near the tumour edge under Surface Growth. The power law pattern characterising the
22 spatial distribution of microdiversity hotspots was corroborated in the experimental data
23 (**Figure 4e**); ccRCC tumours displayed a k value numerically between those in the Volume
24 Growth and Surface Growth models (**Figure 4g**). A greater error in matching data and fitted
25 power law was noted at regions approaching the margin, likely attributed to under-sampling of
26 biopsies at the tumour periphery (**Supplemental Figure 5**).

27
28 To summarise, regions with high clonal diversity within a tumour slice were increasingly
29 frequent towards the tumour margin, a spatial feature particularly enhanced in the Surface
30 Growth model and corroborated experimentally using PT regions with at least two subclones.
31 These data suggested that subpopulations with high clonal diversity were abundant near the
32 tumour margin.

34 Frequency and spatial features of parallel evolution

35 As subclonal diversification could involve acquisition of, and be facilitated by, distinct
36 mutations in the same gene at spatially separate locations, we next evaluated the frequency of
37 parallel evolution events and their spatial features.

38
39 Distinct mutational instances in the same driver gene were recorded as different driver events
40 in the model (**Figure 5a**). Guided by an exploratory whole-tumour analysis, which showed
41 parallel mutational events other than *PBRM1* and *BAP1* to be rare at relatively low resolution
42 (**Supplemental Note 2, Supplemental Figure 6a-b**), we specifically focused on these two
43 mutations. As might be expected, as a reflection of the extent of subclonal diversification in
44 the whole tumour shown above (**Figure 2b**), parallel evolution was evident in the Volume
45 Growth model only when s and p_{driver} were both large (**panel (i) in Figure 5b**), while it was
46 pervasive in the Surface Growth model under various parameter conditions (i.e., pairs of s and
47 p_{driver}) (**panel (ii) in Figure 5b**).

48 To further elucidate the spatial patterns of parallel mutations in *PBRM1* or *BAP1*, we next
49 labelled tumour regions harbouring these events (**Figure 5a**). Within a tumour section,

1 subpopulations harbouring parallel mutational events expanded to occupy a variable number
2 of regions (**Figure 5c-d**). When mutational instances in *PBRM1* or *BAP1* occurred late, they
3 led to limited clonal expansion (up to two tumour regions), despite the role in promoting
4 acquisition of SCNAs (**Supplemental Figure 6c-f**). Those instances were commonly observed
5 at a short distance to the tumour edge (**Figure 5e**). Furthermore, for a range of p_{driver} , the
6 Surface Growth model resulted in a shorter distance of the parallel evolution events to the
7 tumour edge, compared to the Volume Growth model, likely due to the birth of these subclones
8 occurring at the tumour surface.

9
10 In the TRACERx Renal study, parallel evolution was observed in 28 tumours, with parallel
11 mutations in the same gene spanning distinct sets of regions (**Supplemental Note 3**). As
12 highlighted previously (Turajlic, et al. 2018) and similar to the *in-silico* tumours, distinct
13 parallel mutational events could span a variable number of PT regions (**Figure 5f**).
14 Consistent with the *in-silico* analyses, parallel mutational events in driver genes with limited
15 clonal expansion (spanning up to two PT regions) were predominantly located near the
16 tumour edge (**Figure 5g**). This spatial distribution was also found for mutational events
17 without evidence of parallel evolution and with limited expansion (**Figure 5g**), suggesting
18 that subclones lacking significant expansion were commonly emerging near the tumour edge.
19 The lack of expansion of these subclones could be attributed to either their narrow fitness
20 advantage and/or late occurrence.

21
22 Together, these data suggest that parallel evolution in *BAP1* or *PBRM1* was more prevalent in
23 the Surface Growth model, which underlined extensive subclonal diversification. For both *in-*
24 *silico* modelling and experimental analysis of ccRCC tumours, parallel mutations with limited
25 clonal expansion were located near the tumour edge, implying that a spatially resolved
26 sampling could target the tumour margin to identify ongoing parallel evolution.

27 28 Predictive feature of future evolutionary trajectories

29 Having so far focused on quantitative and spatial characterisation of microdiversity and
30 parallel evolution, we next asked how features of clonal diversity evolved temporally and
31 whether some features could enable the prediction of future evolutionary trajectories. The
32 ability to predict likely evolutionary trajectories could aid in the clinical management of
33 ccRCC tumours. However, the challenges of obtaining serial biopsies of the same tumour
34 limit this opportunity in clinical practice. Therefore, we examined if our model could be used
35 to indicate the temporal features of evolution of clonal diversity.

36
37 To this end, the number of subclones within a tumour slice was tracked over time. Tumours
38 under Volume Growth developed a small number of subclones that were sustained over the
39 course of tumour evolution (**Figure 6a, panel (i) in Figure 6b, Supplemental Figure 7a**). By
40 comparison, the number of subclones increased over time in tumours under Surface Growth
41 (**Figure 6a, panel (ii) in Figure 6b, Supplemental Figure 7b**). Nevertheless, tumours under
42 distinct growth modes appeared indistinguishable at early stage according to this measurement.
43 We then asked whether some spatial features could indicate early the divergent patterns of
44 subsequent subclonal diversification in the two models.

45
46 To shed light on this question, we evaluated the Surface Growth model for the presence of
47 early indicators of subclonal diversification. In fact, one characteristic spatial feature noticed
48 in tumours under Surface Growth was a budding structure on the tumour surface, which
49 indicated the beginning of outgrowth of an advantageous subclone (**Figure 1e, Figure 6c**).

1 Quantitatively, this budding structure reflected the rapid increase in the CCF of an
2 advantageous subclone, as it spatially outcompeted surrounding subpopulations (**Figure 6d**).
3 Depending on the driver acquisition probabilities employed in the model, the median tumour
4 size for the detection of such budding structures was around 7 cm or smaller. Exploratory
5 simulations attempting at “replaying” evolution (i.e., re-simulating clonal evolution from a
6 historical tumour state with established clonal structure as a starting point) starting from
7 different tumour sizes suggested that evolution was more repeatable if starting from a historical
8 tumour state with budding structures emerging (**Supplemental Figure 8, Supplemental Note**
9 **4**).

10
11 With respect to the above findings, we then turned to analysing 46 tumours with a size of up
12 to 7cm in the TRACERx Renal study. By qualitative examination of radiological images of
13 these tumours, 15 tumours displayed apparent budding structures at their surface. In one
14 representative case (“G_K523”), budding structures were evident both radiologically (**Figure**
15 **6e**) and in the tumour contour image with clonal diversity mapped (**Figure 6f**). Interestingly,
16 trailing the budding structures were regions with high clonal diversity, consistent with our *in-*
17 *silico* tumour simulations under Surface Growth. These findings imply ongoing subclonal
18 diversification in these regions that may delineate future evolutionary steps.

19
20 Lastly, in the TRACERx Renal study, the number of subclones depicted a non-linear
21 relationship with tumour size (**Figure 6g**). This broadly reflected different modes of evolution,
22 ranging from limited evolution (bottom-left part in **Figure 6g**), to punctuated evolution
23 (bottom-right part in **Figure 6g**) and branched evolution (top part in **Figure 6g**). Furthermore,
24 a subset of tumours (enclosed only by blue contour in **Figure 6g**) were better recapitulated by
25 the Volume Growth model, while another subset by Surface Growth model (enclosed only by
26 red contour in **Figure 6g**). A more intriguing question still is whether the future steps of clonal
27 evolution in the early-stage tumours could be predicted. When 15 early-stage tumours with
28 apparent budding structures were highlighted, a subset of these already displayed a greater
29 extent of subclonal diversification, raising the possibility that spatially resolved sampling
30 combined with sequencing in these regions may enable identification of evolutionary potential
31 in early-stage tumours.

1 Discussion:

2 Intra-tumour genetic heterogeneity arises when different parts of the tumour acquire distinct
3 genomic alterations, endowing subclones with a variety of fitness advantages. Despite our
4 understanding of how dominant subclones sculpt evolutionary trajectories and histories,
5 gleaned primarily from multi-region sampling and deep sequencing, under-represented
6 subclones that reflect ongoing and clinically relevant evolution could remain undetected.
7 Therefore, our focus on microdiversity and parallel evolution is central to the elucidation of
8 the evolutionary potential of under-represented subclones and the emergent spatial patterns
9 they form.

10

11 An important finding, via computational modelling, is that different spatial patterns of tumour
12 growth impact the extent of subclonal diversification and shape divergent modes of evolution.
13 While Surface Growth (i.e., proliferation limited to the superficial layer of a tumour) enables
14 more extensive subclonal diversification and gives rise to branched evolution, Volume Growth
15 (i.e., proliferation active throughout the tumour volume) commonly results in single dominant
16 subclone coexistent with founding clone reflective of punctuated evolution. Linear evolution
17 was not observed in our model. Experimentally, the inferences of linear evolution are generally
18 biased by the lack of sufficient sampling to detect intermediate populations. Overall, this
19 observation resonates with other computational studies to highlight the importance of
20 understanding spatial patterns of tumour growth when deconstructing tumour evolution
21 (Anderson et al 2006, Waclaw et al 2015, Sun et al 2017, Chkhaidze et al 2019, Noble et al
22 2019).

23

24 Another important finding, in both the *in-silico* tumours, in particular those under Surface
25 Growth, and tumour in the TRACERx Renal study, are frequent microdiversity hotspots and
26 parallel mutational events near the tumour edge, suggesting abundant ongoing evolution at the
27 tumour margin. Temporally tracking the parallel mutational events in *PBRM1* or *BAP1* in the
28 *in-silico* setting illuminated the rapid increase of their prevalence in small tumours leading to
29 ultimately established subclones. In those under Surface Growth, this rapid transition was
30 marked by a budding structure on the tumour surface. Interestingly, such budding structures
31 were evident in 15 early-stage ccRCC tumours. While a prospective study in a large cohort is
32 needed for further examination, this observation opens up the possibility that designing
33 spatially resolved sampling strategies in early-stage tumours with assessment of clonal
34 diversity at the tumour margin, in particular within the budding structures, could be beneficial
35 in informing upon a tumour's evolutionary potential.

36

37 Our computational modelling also uncovered power law scaling features that uniquely
38 characterise clonal diversity in a spatial context. These features add to previously reported
39 scaling phenomena describing clonal evolution (Driessens et al 2012, Williams et al 2016). In
40 addition to the scaling feature underlying the spatial distribution of microdiversity hotspots,
41 clonal diversity and sampling area also form a scaling relationship in both the model and
42 experimental data (**Supplemental Figure 9, Supplemental Note 5**). This scaling relationship
43 in ccRCC tumours is reminiscent of the taxa-area relationship in the macroevolution of species
44 (Crawley and Harral 2001, Horner-Devine et al 2004, Zhou et al 2008), drawing a parallel
45 between cancer evolution and evolution in the ecosystem.

46

47 Overall, our computational model was able to correlate with several clinical observations,
48 despite the simplification of its primary components (probabilistic birth, death, and driver
49 acquisition). Nevertheless, we are aware that our model has limitations. Clearly, many other

1 factors, such as vascularisation, necrosis, and immune predation, are all likely to (re)shape the
2 *in-silico* patterns of clonal diversity we observed. The simple implementations of fitness
3 advantage endowed by ccRCC drivers could be improved in the future with wet-lab
4 experiments aimed at measuring and comparing growth kinetics of tumours with different
5 genetic backgrounds. Future time-lapse experiments could also provide evidence for growth
6 modes of ccRCCs and elucidate whether different modes may predominate in different regions
7 or at different stages of the same tumour. Moreover, while we made an assumption that only
8 mutations in *BAP1* or *PBRM1* promoted acquisition of SCNAs for simplicity, we acknowledge
9 that other ccRCC driver mutations (e.g., *SETD2*) also link to chromosome instability. Lastly,
10 alternative mechanisms could explain the formation of budding structures and other protrusive
11 morphologies, such as mechanical properties (Fiore et al. 2020) and cell migration (Anderson
12 et al. 2006). In addition to these limitations from a modelling perspective, differences in size,
13 stage, and the number of samples among tumours in the TRACERx Renal study could
14 confound the patterns to which the computational outputs are compared. The current
15 sequencing data is panel-based with a small number of clonal markers and may have low
16 sensitivity to detect microdiversity. Future examination of whole-exome or whole-genome data
17 could lead to a more comprehensive characterisation of ongoing clonal evolution.

18
19 In conclusion, our study supports the importance of understanding spatial patterns of tumour
20 growth in deconstructing tumour evolution. It provides evidence, and has implications for,
21 focused spatial sampling of the tumour margin in order to garner richer information on ongoing
22 and clinically relevant evolution.

23

1 Methods:

2 Computational model

3 Tumour growth and clonal evolution in a spatio-temporal context have increasingly been
4 studied with the aid of computational models that incorporate rule-based spatial growth and
5 acquisition of genomic alterations (Anderson et al 2006, Waclaw et al 2015, Sun et al 2017,
6 Chkhaidze et al 2019). Spatial patterns of tumour growth (Sun et al 2017, Chkhaidze et al
7 2019) have been shown to impact the ability to classify neutral evolution in contrast to
8 selection, suggesting that spatial growth of a structured population interplays with
9 evolutionary forces (driver acquisition, selection, and genetic drift) to shape the spatial
10 patterning of subclones.

11
12 In the present study, a coarse-grained cellular automaton model has been constructed to
13 simulate tumour growth and the evolution of ccRCC drivers. A basic model unit reflects a
14 tumour volume of $1 \times 1 \times 1 \text{ mm}^3$, referred to as a “tumour voxel”. The choice of 1 mm for
15 the size of a tumour voxel reflects the observed thickness of actively proliferative surface layer
16 of some ccRCCs (Hoefflin et al. 2016). The full simulation lattice comprises $200 \times 200 \times 200$
17 lattice sites in which a tumour grows. The following sub-sections detail the model components
18 and assumptions.

19

20 Growth and death

21 Tumour voxels stochastically undergo growth and death, with baseline probabilities per
22 simulation step of $p_{growth} = 0.25$ and $p_{death} = 0.05$, respectively. Upon death, a tumour
23 voxel is removed from the simulation lattice, rendering the site empty and available for
24 accommodating new tumour voxels. Two different modes of spatial tumour growth are
25 considered: Surface Growth and Volume Growth (**Figure 1c**). For Surface Growth,
26 proliferation is only allowed to take place when space is available, namely, when at least one
27 of the 26 neighbouring lattice sites of the tumour voxel selected to divide is empty. Upon
28 duplication of a parent tumour voxel, one child tumour voxel retains the location of the parent
29 while the other is placed at a randomly selected adjacent empty site. For Volume Growth, all
30 tumour voxels are able to proliferate; upon duplication, one child tumour voxel retains the
31 location of the parent while the other is placed at a selected adjacent site according to the rule
32 described below and pushes tumour voxels in that orientation outward. The process for
33 selecting an adjacent site includes two steps: (1) to randomly sample 10 candidate positions
34 out of the 26 neighbouring lattice sites; (2) to select the orientation (i.e., pointing from the
35 position of the parent tumour voxel to the candidate position) giving the smallest distance from
36 the tumour surface, similar to the algorithm described in Waclaw et al. 2015.

37

38 Driver events

39 A panel of 26 ccRCC drivers that were highlighted in Turajlic et al. 2018, including mutations
40 in 12 driver genes and 14 somatic copy number alterations (SCNAs), are considered in the
41 present work (**Supplemental Figure 1**). For simplicity, the fitness advantage conferred by a
42 driver is assumed to manifest as growth advantage. In implementation, three levels of growth
43 probability, $p_{growth}^{(1)} \leq p_{growth}^{(2)} \leq p_{growth}^{(3)}$, were introduced for the 26 drivers, to broadly
44 reflect their association with the Ki67 score in tumour regions in the experimental data

1 **(Supplemental Figure 1)**. In a general form, $p_{growth}^{(2)} = g(s)p_{growth}^{(1)}$ and $p_{growth}^{(3)} =$
2 $h(s)p_{growth}^{(1)}$ are functions of baseline growth probability, where $h(s) \geq g(s) \geq 1$ reflect the
3 growth advantages relative to the baseline. As one specific implementation, growth
4 probabilities were set at $p_{growth}^{(1)} = 0.25$, $p_{growth}^{(2)} = (1 + s)p_{growth}^{(1)}$ and $p_{growth}^{(3)} = (1 +$
5 $s)^2 p_{growth}^{(1)}$, where $0 \leq s \leq 1$ determines this advantage relative to the baseline growth
6 probability. The baseline growth probability $p_{growth}^{(1)}$ was arbitrarily defined, as time has
7 arbitrary units in the model. For a given tumour voxel, its growth probability is defined by the
8 most advantageous driver harboured (i.e., “saturated” model of fitness advantage). For
9 simplicity, *VHL* mutation and 3p loss are considered as truncal events in the founder tumour
10 voxel, although in a clinical setting *VHL* mutation are not present universally (Turajlic et al.
11 2018). The subpopulation of tumour voxels that only harbour these two events is referred to
12 parental clone. For an exploratory purpose, an alternative implementation of growth advantage
13 endowed by drivers was also evaluated (**Supplemental Figure 2**). In this implementation, a
14 driver adds a certain amount of growth probability to the tumour voxel that acquires the driver
15 (i.e., “additive” model of fitness advantage) (**Supplemental Figure 2a**). The growth
16 probability of a tumour voxel depends on all the drivers it harbours, namely, $p_{growth} =$
17 $p_{growth}^{(1)} + \sum_k p_{growth_k}$, where p_{growth_k} reflects the amount of growth probability added by
18 driver k . p_{growth} is set to one if the calculated probability exceeds one. The amount
19 p_{growth_k} varies between drivers, reflecting different strengths of their association with Ki67
20 score (**Supplemental Figure 1a**). Three different scenarios were explored to reflect different
21 amounts of growth probability endowed by drivers on average, as determined by s_k of the
22 weakest driver, namely, $\min(s_k)$, and the difference in s_k between consecutive two drivers in
23 their advantages, namely, Δs_k . (**Supplemental Figure 2b**).

24
25 Upon proliferation of a parent tumour voxel, child tumour voxels inherit existing driver events
26 harboured by the parent tumour voxel and stochastically acquire new drivers. Mutations in
27 driver genes are acquired with a probability p_{driver} and are assumed to be acquired in a tumour
28 voxel more frequently than SCNAs, which have probability of $0.001p_{driver}$. Given the
29 functional evidence that *BAP1* (Peng et al. 2015) and *PBRM1* genes (Varela et al. 2011) guard
30 against chromosome instability and the experimental data in the TRACERx Renal study
31 demonstrating the association of mutations in these genes with high wGII (Turajlic et al. 2018),
32 we assume that the acquisition of SCNAs in a tumour voxel becomes equally likely as
33 mutations in driver genes, if the tumour voxel harbours mutations in *BAP1* or *PBRM1*. A range
34 of driver acquisition probabilities have been studied to explore its impact on patterns we
35 investigate. Lastly, only one mutation in the same driver gene is permitted in the same tumour
36 voxel, but multiple independent, distinct mutations may be acquired in parallel within a
37 simulated tumour in different tumour voxels.

38 39 Simulation

40 Each simulation starts from a single tumour voxel placed at the centre of the lattice, (x_0, y_0, z_0) ,
41 and runs until the tumour grows to at least 1 million tumour voxels after the last simulation
42 step. In each simulation step, growth and death are evaluated for every tumour voxel, in random
43 order. For those tumour voxels selected to grow, driver acquisition is evaluated for each of the
44 26 drivers. The computer code is written in CUDA C++.

45

1 Model analyses

2 Levels of analysis

3 Analyses have been conducted at three different levels (**Figure 1D**): (1) whole tumour level,
4 which takes into account all tumour voxels in the 3D volume; (2) tumour slice level, which
5 takes into account all tumour voxels within a 2D plane ($z = z_0$); (3) regional biopsy level,
6 which takes into account tumour voxels within regional biopsies. A regional biopsy is defined
7 as all tumour voxels within a region in the 2D slice. Spatially uniform sampling is performed
8 in this study. This process is carried out by locating the centres of candidate regional biopsies
9 in the $200\text{mm} \times 200\text{mm}$ 2D lattice with a spacing of 20mm and collecting all voxels within
10 a distance of 5mm from each biopsy centre.

11

12 Spatial maps of driver events

13 In representative cases, tumour voxels that harbour selected driver events are mapped, within
14 a 3D tumour surface and within a 2D tumour slice. In spatial maps of driver events, tumour
15 voxels that don't harbour drivers of interest are in grey; tumour voxels that harbour mutations
16 in driver genes are in blue; tumour voxels that harbour gain of chromosome arm events are in
17 green; tumour voxels that harbour loss of chromosome arm events are in red.

18

19 Spatial maps of subclones

20 A subclone is defined as a group of tumour voxels that harbour the same set of driver events.
21 In spatial maps of subclones, tumour voxels that belong to the parental clone are in grey, while
22 other subclones are visualised in randomly generated (R, G, B) colours.

23

24 Cancer cell fraction (CCF) of driver events

25 CCF of a driver event is calculated as the number of tumour voxels that contain the driver event
26 divided by the total number of tumour voxels in the domain of interest, depending on the level
27 of analyses. A driver event is considered detectable if the CCF is greater than 0.01.

28

29 Cancer cell fraction of subclones

30 CCF of a subclone is calculated as the number of tumour voxels that belong to a subclone
31 divided by the total number tumour voxels in the domain of interest, depending on the level of
32 analyses. A subclone is identified by a set of driver events, shared by a subpopulation of tumour
33 voxels, which are accumulated within the subclone-initiating tumour voxel. A subclone-
34 initiating tumour voxel is defined as a tumour voxel that acquires a new driver event upon birth.
35 A subclone is considered detectable if the CCF is greater than 0.01.

36

37 Shannon diversity index

38 As a measure of clone diversity, Shannon diversity index is defined as $S = \sum_i -f_i \ln f_i$, where
39 f_i is the CCF of the subclone i . All subclones are taken into account in this calculation.

40

1 Spatial and genomic distances

2 Spatial distance between two regional biopsies is calculated as the Euclidean distance between
3 the positions of the centres of two biopsies. Genomic distance between two regional biopsies
4 is calculated as $d_{gen} = \sqrt{\sum_k^{26} (G_i^k - G_j^k)^2}$, where G_i^k is the state of k th driver in the i th region
5 and has a value of 1 if the k th driver is present, otherwise a value of 0.
6

7 Microdiversity

8 Microdiversity is defined as the number of subclones contained in a 3-mm-by-3-mm region
9 within the tumour slice. In representative cases, microdiversity is spatially mapped within a
10 tumour slice, by sliding a 3-mm-by-3-mm spatial window throughout the tumour slice.
11 Microdiversity hotspots are defined as a subset of these small regions with 5 or more subclones.
12 The distance from a microdiversity hotspot to the centre of a tumour slice is referred to as the
13 distance to tumour centre (d_1). The distance from a microdiversity hotspot to the nearest point
14 along the tumour contour is referred to as the distance to tumour margin (d_2). The normalised
15 distance to tumour centre is defined as $d = d_1 / (d_1 + d_2)$. Cumulative probability distribution
16 of d is generated by combining microdiversity hotspots from repeat simulations. Power law
17 exponent is obtained by bootstrapping 100 samples of 400 hotspots per sample and fitting a
18 power law function to cumulative probability distribution of r in each sample.
19

20 Parallel evolution

21 Parallel evolution refers to the independent, distinct instances of mutations in the same driver
22 gene that are acquired at different tumour locations. At the whole tumour level, the extent of
23 parallel evolution is assessed by counting the instances of independent, distinct instances of
24 mutations in each driver gene at a given resolution of detection. The resolution of detection is
25 represented using a minimum CCF that has to be reached by a parallel mutational event to be
26 detectable. At the regional biopsy level, the extent of parallel evolution is assessed by counting
27 the number of regions that contain distinct instances of mutations for each driver gene. The
28 distance from any region containing a parallel instance to the tumour margin is measured, to
29 study the spatial distribution of parallel mutational instances with limited clonal expansion.
30

31 Temporal analysis

32 For the time-course study, 250 simulations with 5 different levels of p_{driver} (50 simulations
33 per condition) are performed for each of Surface Growth and Volume Growth models. 2D
34 tumour slices are collected every 10 steps in Surface Growth model and 5 steps in Volume
35 Growth model. The number of subclones is counted within each historical tumour slice.
36 Kernel density estimation with a Gaussian kernel is performed with respect to the number of
37 subclones and the diameter of the tumour slice, based on all simulations, to produce a
38 continuous density estimate, using `seaborn.jointplot(kind = 'kde')` in Python.
39 In a representative case, tumour voxels that harbour parallel mutational events in *BAP1* are
40 mapped within historical tumour slices. In all simulations, CCFs of subclones that harbouring
41 mutations in *BAP1* are measured within historical tumour slices over time.
42

1 Evolutionary replay

2 To perform evolutionary replay, historical tumour state (i.e., locations and clonal identities of
3 tumour voxels) at a certain time point is saved and employed as a common starting state for re-
4 growth 50 new simulated tumours. At the end of these simulations, Shannon diversity index is
5 calculated to indicate the divergence in their evolutionary outcomes.

7 Experimental analyses

9 TRACERx Renal cohort

10 79 tumour sections of 66 unique primary tumours are included in this study; see the exclusion
11 criterion in our previous publication (Zhao, et al. 2020).

13 CT images

14 Contrast-enhanced CT images were obtained using standard-of-care imaging sequences in 91
15 patients and curated using a local research PACS (based on the XNAT platform, (Marcus, et
16 al., 2007)). Outlines were drawn giving volumetric tumour coverage by an oncologist (S.S.)
17 and checked by a radiologist (D.A.), from which image strips were prepared for rapid
18 visualisation of all tumour slices for all patients using an in-house script written in python.

20 Microdiversity

21 Spatial maps of regional clone diversity are created for two representative tumour sections. In
22 these maps, regions are colour-coded based on the number of subclones. Regions that harbour
23 at least one subclone are treated as a proxy for microdiversity hotspots defined in the model
24 analysis. In total, there are 606 regions from 54 tumours that satisfy this criterion. For these
25 regions, the normalised distance to tumour centre is measured as described above in “Model
26 analysis - Microdiversity”.

28 Parallel evolution

29 Spatial maps of parallel mutational events in *PBRM1* are created for one representative tumour
30 section. In these maps, regions are coloured differently according to different parallel
31 mutational events. Regions that harbour more than one event are indicated with multiple
32 colours. To study the spatial distribution of mutational events with limited clonal expansion,
33 the maximum distance from an event spanning up to two regions to the tumour margin is
34 measured.

36 Statistical analysis

37 Two-sided Wilcoxon’s rank test is performed to compare several measurements between different
38 model conditions. Statistical significance is annotated within box plots using
39 `stat_compare_means(method = "wilcox.test", label = "p.signif")` in
40 R.

1 Pearson correlation coefficients are calculated to assess the correlation between spatial and
2 genomic distances between tumour regions, using `cor.test(variable1,`
3 `variable2, method = "pearson")` in R.
4 Two-sample Kolmogorov-Smirnov test is performed to compare two cumulative probability
5 distributions, using `scipy.stats.ks_2samp(sample1, sample2)` in Python.
6 Bootstrapping is performed to generate 100 random samples of 400 microdiversity hotspots
7 per sample with replacement, using `random.choice()` in Python. The power law
8 exponent is then determined by fitting a power law function to the cumulative probability
9 distribution from each sample, using `scipy.optimize.curve_fit()` in Python.
10 Quantile-Quantile (Q-Q) plot is employed to compare actual distribution of microdiversity
11 hotspots with a power law distribution with exponent being the median of fitted values in
12 bootstrapping, using `statsmodels.graphics.gofplots.qqplot()` in Python.
13 Kernel density estimation is performed for simulations with respect to the size of tumour
14 slice and the number of subclones, using `seaborn.jointplot(kind="kde")` in
15 Python.
16 R version 3.6.2 and Python version 3.7.7 are used for these analyses.

17

18 [Code availability](#)

19 CUDA C++ code of a representative model is available on GitHub: <https://github.com/xxxxx>

20

21

1 Acknowledgements

2 This research was funded in whole, or in part, by the Wellcome Trust (FC001003). For the
3 purpose of Open Access, the author has applied a CC BY public copyright licence to any
4 Author Accepted Manuscript version arising from this submission.

5
6 This work was supported by the Francis Crick Institute, which receives its core funding from
7 Cancer Research UK (FC001003, FC001144, FC001169, FC001988), the UK Medical
8 Research Council (FC001003, FC001144, FC001169, FC001988), and the Wellcome Trust
9 (FC001003, FC001144, FC001169, FC001988).

10
11 We thank colleagues from Biomolecular Modelling Lab, Tumour Cell Biology Lab, Cancer
12 Dynamics Lab, and Cancer Evolution and Genome Instability Lab for insightful discussion and
13 comments on this manuscript.

14
15 ST is funded by Cancer Research UK (grant reference number C50947/A18176); the Francis
16 Crick Institute, which receives its core funding from Cancer Research UK (FC10988), the UK
17 Medical Research Council (FC10988), and the Wellcome Trust (FC10988); the National
18 Institute for Health Research (NIHR) Biomedical Research Centre at the Royal Marsden
19 Hospital and Institute of Cancer Research (grant reference number A109); the Royal Marsden
20 Cancer Charity; The Rosetrees Trust (grant reference number A2204), Ventana Medical
21 Systems Inc (grant reference numbers 10467 and 10530); the National Institute of Health
22 (Bethesda) and Melanoma Research Alliance. ST has received speaking fees from Roche, Astra
23 Zeneca, Novartis and Ipsen. ST has the following patents filed: Indel mutations as a
24 therapeutic target and predictive biomarker PCTGB2018/051892 and PCTGB2018/051893
25 and Clear Cell Renal Cell Carcinoma Biomarkers P113326GB. A.F. receives funding from the
26 European Union's Horizon 2020 research and innovation programme under the Marie
27 Skłodowska-Curie grant agreement No. 892360.

28
29 We thank the TRACERx Renal trial team and the Skin and Renal Unit Research Team at The
30 Royal Marsden NHS Foundation Trust, including Eleanor Carlyle, Lyra Del Rosario, Kim
31 Edmonds, Karla Lingard, Mary Mangwende, Sarah Sarker, Charlotte Lewis, Fiona Williams,
32 Hamid Ahmod, Tara Foley, Dilruba Kabir, Justine Korteweg, Aida Murra, Nahid Shaikh,
33 Kema Peat, Sarah Vaughan and Lucy Holt. TRACERx Renal is funded by NIHR BRC at the
34 Royal Marsden Hospital and Institute of Cancer Research (A109). The Francis Crick Institute,
35 which receives its core funding from CRUK (FC010110), the UK Medical Research Council
36 (FC010110), the Wellcome Trust (FC010110).

37 38 **The TRACERx Renal Consortium**

39 The members of TRACERx Renal Consortium are: Lewis Au, Ben Challacombe, Ashish
40 Chandra, Simon Chowdhury, William Drake, Archana Fernando, Nicos Fotiadis, Andrew
41 Furness, Emine Hatipoglu, Karen Harrison-Phipps, Steve Hazell, Peter Hill, Catherine
42 Horsfield, James Larkin, Jose I. Lopez, Teresa Marafioti, David Nicol, Tim O'Brien, Jonathon
43 Olsburgh, Lisa Pickering, Alexander Polson, Sergio Quezada, Sarah Rudman, Scott Shepherd,
44 Charles Swanton, Samra Turajlic, Mary Varia, Hema Verma.

45

1 Author contributions

2 X.F., Y.Z., E.S., K.L., S.T., and P.A.B. conceived and designed the study.
3 X.F. and P.A.B. constructed the computational model and performed simulations.
4 Y.Z., C.S., K.L., and S.T. provided the tumour samples and performed imaging analysis.
5 S.D., M.O, and S.K curated and processed the radiological data in this study. S.S. performed
6 the initial outlining of the tumours and D.A. checked and refined them.
7 X.F. performed data analysis on simulations.
8 X.F., Y.Z. and J.I.L performed data analysis on tumour samples.
9 X.F., S.T., and P.A.B. wrote the manuscript with input from all authors.

10 Competing interests

11 K.L., S.T., and C.S. have a patent on indel burden and checkpoint inhibitor response pending,
12 and a patent on targeting of frameshift neoantigens for personalised immunotherapy pending.
13 K. L. reports speaker fees from Roche Tissue Diagnostics. ST reports grants from Ventana,
14 outside the submitted work. J.L. reports institutional research support from: BMS, MSD,
15 Novartis, Pfizer, Achilles Therapeutics, Roche, Nektar Therapeutics, Covance, Immunocore,
16 Pharmacyclics, Aveo, and consultancy support from: Achilles, AZ, Boston Biomedical, BMS,
17 Eisai, EUSA Pharma, GSK, Ipsen, Imugene, Incyte, iOnctura, Kymab, Merck Serono, MSD,
18 Nektar, Novartis, Pierre Fabre, Pfizer, Roche / Genentech, Secarna, Vitaccess. C.S. receives
19 grant support from Pfizer, AstraZeneca, BMS, Roche-Ventana, Boehringer-Ingelheim, and
20 Ono. C.S. has consulted for Pfizer, Novartis, GlaxoSmithKline, MSD, BMS, Celgene,
21 AstraZeneca, Illumina, Genentech, Roche-Ventana, GRAIL, Medicxi, and the Sarah Cannon
22 Research Institute. C.S. is a shareholder of Apogen Biotechnologies, Epic Bioscience, GRAIL,
23 and has stock options in and is co-founder of Achilles Therapeutics. E.S. receives research
24 support from AstraZeneca, Glaxo SmithKline, Merck Sharp & Dhome and is on the advisory
25 board of Phenomic AI.
26

27 Additional information

28 Supplemental notes and figures are included in a separate document.
29

1 Figure Legends:

2 **Figure 1.** Construction of *in silico* tumours.

3 (a) Schematic figure illustrating future evolutionary trajectories delineated by present under-
4 detected subclones.

5 (b) Schematic figure of probabilistic growth, death, and driver acquisition in a coarse-grained
6 cellular automaton model.

7 (c) Schematic figure of two growth modes: “Surface Growth” with proliferation predominating
8 at the tumour surface and “Volume Growth” with proliferation throughout the tumour volume.

9 (d) Schematic figure of three levels of measurements: from three-dimensional (3D) tumour to
10 two-dimensional (2D) tumour slice and 2D tumour regions within the slice.

11 (e) Representative *in-silico* tumours under Volume Growth (i) and Surface Growth (ii),
12 respectively, from a 3D view. Tumour voxels harbouring select drivers, as indicated in the
13 figure, are colour coded. Tumour voxels harbouring gain of chromosome arm events are in
14 green; tumour voxels harbouring loss of chromosome arm events are in red. Different shades
15 of greens or reds are employed to reflect different driver events.

16

17 **Figure 2.** Clonal diversity in the whole tumour.

18 (a) Schematic figure for the whole-tumour analysis of clonal diversity.

19 (b) Heatmap showing the average number of clones (i.e., parental clone and subclones) with
20 respect to driver acquisition probability and proliferative advantage in the Volume Growth (i)
21 and Surface Growth (ii) models. The average is calculated from 50 *in silico* tumours per
22 parameter condition. Clones with a whole-tumour cancer cell fraction (CCF) of at least 0.05
23 are counted.

24 (c) Whole-tumour CCF of parental and largest subclones in *in silico* tumours under Volume
25 Growth (i-ii) and Surface Growth (iii), respectively. Driver acquisition probabilities in these
26 sets of simulations are $p_{driver} = 2 \times 10^{-4}$ in (i), 1×10^{-3} in (ii), 2×10^{-4} in (iii),
27 respectively. “Parental (3p loss, *VHL*)” clone is shown along with up to five subclones with a
28 whole-tumour CCF of 0.01 or higher. All remaining subclones are represented in the “other”
29 group.

30 (d) Whole-tumour CCF of parental subclones in *in silico* tumours under Volume Growth and
31 Surface Growth with varying driver acquisition probabilities. $N = 100$ for each condition.

32 (e) Shannon diversity index in *in silico* tumours under Volume Growth and Surface Growth
33 with varying driver acquisition probabilities. $N = 100$ for each condition.

34 Statistical annotations in (d-e) reflect two-sided Wilcoxon tests: “****” indicates $P \leq 0.0001$.

35

36 **Figure 3.** Between-region genomic divergence.

37 (a) Schematic figure and procedure for the analysis in 2D tumour regions.

38 (b) Spatial locations of regional samples with tumour edge indicated in a representative *in silico*
39 tumour under Surface Growth.

40 (c) Between-region spatial distance against genomic distance derived from the representative
41 *in silico* tumour in panel b. Overlapping data points are counted (“n”). Pearson correlation (“r”)
42 and p value are indicated in the figure.

43 (d-e) The same analysis as described in (b-c) for a representative *in silico* tumour with Volume
44 Growth.

45 (f) Pearson correlation between spatial distance and genomic distance in *in silico* tumours
46 under Volume Growth and Surface Growth with varying driver acquisition probabilities. $N =$
47 100 for each condition.

1 Statistical annotations in (f) reflect two-sided Wilcoxon tests: “****” indicates $P \leq 0.0001$
2 and “ns” indicates no statistical significance.

3
4 **Figure 4.** Spatial features of clonal diversity.

5 (a) Schematic figure and procedure for the analysis within a 2D tumour.

6 (b) Spatial maps of subclones (i) and microdiversity (ii) in a representative *in-silico* tumour
7 under Surface Growth.

8 (c) The same analysis as described in (b) for a representative *in-silico* tumour under Volume
9 Growth.

10 (d) Maps of regional biopsies with the number of subclones within a biopsy colour coded in
11 two cases (G_K234 and G_K446) in the TRACERx Renal study. Hues from red to purple to
12 blue reflect decreasing number of subclones. “Low” reflect zero subclones, while “High”
13 reflects the maximum number of subclones found in any region (i.e., 4 subclones in G_K234,
14 3 subclones in G_K446).

15 (e) Cumulative probability distribution, $P(D \leq d)$, of the normalised distance to tumour centre
16 in *in silico* tumours under Surface Growth and Volume Growth and in ccRCC tumours. Three
17 sets of *in-silico* tumours with different driver acquisition probabilities are shown. $N = 100$ for
18 each model condition. “S” and “V” in the figure reflect Surface Growth and Volume Growth,
19 respectively. “p=2e-4” reflects a driver acquisition probability of $2e-4$. 606 patient tumour (PT)
20 regions from 54 ccRCC tumours are considered for the experimental analysis.

21 (f) Bootstrapped power law exponent k , as in $P(D \leq d) \sim d^k$, fitted to cumulative probability
22 distribution of normalised distance to tumour centre in each of bootstrap samples. For
23 microdiversity hotspots under each model condition, 100 bootstrap samples are generated by
24 randomly sampling 400 hotspots with replacement.

25 (g) Bootstrapped power law exponent k in ccRCC tumours.

26 Driver acquisition probabilities used in representative simulations are: $p_{driver} = 2 \times 10^{-4}$ in
27 both *in-silico* tumours under Surface Growth (b) and *in-silico* tumours under Volume Growth
28 (c).

29 Statistical annotations in (f) reflect two-sided Wilcoxon tests: “****” indicates $P \leq 0.0001$.

30
31 **Figure 5.** Frequency and spatial features of parallel evolution.

32 (a) Schematic figure for measuring parallel evolution in the 3D tumour (i) and for analysing
33 parallel mutational events within 2D tumour regions (ii).

34 (b) Heatmap showing the fraction of *in silico* tumours that have parallel evolution in *PBRM1*
35 with respect to driver acquisition probability and growth advantage in the Volume Growth (i)
36 and Surface Growth (ii) models. The fraction is calculated based on 50 *in silico* tumours per
37 parameter condition. *In silico* tumours are counted, if there are two or more parallel mutational
38 events in *PBRM1* with a whole-tumour CCF of at least 0.05.

39 (c-d) Spatial distribution of *PBRM1* or *BAP1* mutation (i) and its parallel events (ii) in a
40 representative *in silico* tumour under Surface Growth (c) and Volume Growth (d), respectively.
41 In panel (ii), different colours overlaid with “*PBRM1*” or “*BAP1*” reflect different parallel
42 mutation events.

43 (e) Maximum distances to the tumour edge from parallel mutations in *PBRM1* or *BAP1* that
44 span up to two regions in *in silico* tumours under Volume Growth and Surface Growth with
45 varying driver acquisition probabilities. $N = 100$ for each condition.

46 (f) Maps of regions containing parallel mutations in *PBRM1* in a representative case (G_K520)
47 in the TRACERx Renal study. Distinct parallel mutations are coloured differently. For regions
48 containing more than two parallel mutations, two colours are applied simultaneously. Double-
49 headed arrow indicates a measurement of distance to tumour edge in this example.

1 (g) Maximum distances from mutational events that span up to two regions in ccRCC tumours
2 in the TRACERx Renal study. For comparison, all mutations (“All mutations”), mutations in
3 driver genes (“Drivers”), and parallel mutations in driver genes (“Drivers (P.E.)”) are shown
4 as separate groups.

5 Driver acquisition probabilities used in representative simulations are: $p_{driver} = 2 \times 10^{-4}$ in
6 *in silico* tumours under Surface Growth and $p_{driver} = 1 \times 10^{-3}$ in *in silico* tumours under
7 Volume Growth.

8 Statistical annotations in (e) reflect two-sided Wilcoxon tests: “**” indicates $P \leq 0.01$, “*”
9 indicates $P \leq 0.05$, and “ns” indicates no statistical significance.

10
11 **Figure 6.** Early predictive indicators of evolutionary trajectories.

12 (a) The number of subclones as a function of the diameter of a 2D tumour slice in *in silico*
13 tumours under Surface Growth and under Volume Growth, respectively. $N = 50$ simulations
14 with $p_{driver} = 6 \times 10^{-4}$ are shown for each condition.

15 (b) Kernel density estimation (KDE) with respect to the number of subclones and the
16 diameter of a 2D tumour slice in *in silico* tumours under Volume Growth (i) and under
17 Surface Growth (ii). Each KDE plot is based on 250 simulations (50 per condition) under 5
18 conditions with $p_{driver} = 2 \times 10^{-4}, 4 \times 10^{-4}, 6 \times 10^{-4}, 8 \times 10^{-4}, 1 \times 10^{-3}$.

19 (c) The spatial patterns of parallel mutations in *BAP1* over time in a representative *in silico*
20 tumour under Surface Growth. The red arrow indicates a budding structure at early stage of
21 subclonal expansion.

22 (d) The CCFs of *BAP1* clones within the 2D tumour slice as a function of the diameter of a
23 2D tumour slice in each of 50 *in silico* tumours. Only *BAP1* clones with an ultimate CCF of
24 0.01 or higher are shown. For those with an ultimate CCF of 0.1 or higher, the diameters of
25 tumour slices at which these clones first became detectable are recorded, with the median
26 indicated by the dashed line. The others with an ultimate CCF of 0.1 or lower are shown in
27 grey.

28 (e) Axial image in the corticomedullary contrast phase of a representative case (G_K523)
29 showing budding structure on the tumour surface (red arrow). Outlines in red were drawn
30 giving volumetric tumour coverage by an oncologist (S.S.) and a radiologist (D.A.)

31 (f) Maps of tumour regions with the number of subclones colour coded in a representative
32 case (G_K523). Hues from red to purple to blue reflect decreasing number of subclones.
33 “Low” reflect zero subclones, while “High” reflects the maximum number of subclones
34 found in any region (i.e., 4 subclones in G_K523).

35 (g) The number of subclones as a function of ultimate tumour size in the TRACERx Renal
36 study, overlaid with kernel density estimation based on simulated data. Tumours with a size
37 smaller than 7 cm and with radiologically evident budding structures on the tumour surface
38 are highlighted (orange). Contours reflect 90% probability density based on *in silico* tumours
39 under Surface Growth (red) and under Volume Growth (blue), respective in Figure 6b.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

References

1. Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nat Rev Cancer*. 2006;6(12):924-935. doi:10.1038/nrc2013
2. Zahir N, Sun R, Gallahan D, Gatenby RA, Curtis C. Characterizing the ecological and evolutionary dynamics of cancer. *Nat Genet*. 2020;52(8):759-767. doi:10.1038/s41588-020-0668-4
3. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. 2017;168(4):613-628. doi:10.1016/j.cell.2017.01.018
4. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing [published correction appears in *N Engl J Med*. 2012 Sep 6;367(10):976]. *N Engl J Med*. 2012;366(10):883-892. doi:10.1056/NEJMoa1113205
5. Yates, L., Gerstung, M., Knappskog, S. et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 21, 751–759 (2015). <https://doi.org/10.1038/nm.3886>
6. Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med*. 2017;376(22):2109-2121. doi:10.1056/NEJMoa1616288
7. Turajlic S, Xu H, Litchfield K, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*. 2018;173(3):595-610.e11.
8. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution [published correction appears in *Cancer Cell*. 2015 Jul 13;28(1):141]. *Cancer Cell*. 2015;27(1):15-26. doi:10.1016/j.ccell.2014.12.001
9. Lopez JI, Cortes JM. A divide-and-conquer strategy in tumor sampling enhances detection of intratumor heterogeneity in routine pathology: A modeling approach in clear cell renal cell carcinoma. *F1000Res*. 2016;5:385. Published 2016 Mar 22. doi:10.12688/f1000research.8196.2
10. Mengelbier LH, Karlsson J, Lindgren D, et al. Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer. *Nat Commun*. 2015;6:6125. Published 2015 Jan 27. doi:10.1038/ncomms7125

- 1 11. Casasent AK, Schalck A, Gao R, et al. Multiclonal Invasion in Breast Tumors Identified
2 by Topographic Single Cell Sequencing. *Cell*. 2018;172(1-2):205-217.e12.
3 doi:10.1016/j.cell.2017.12.007
- 4 12. Melchor L, Brioli A, Wardell CP, et al. Single-cell genetic analysis reveals the
5 composition of initiating clones and phylogenetic patterns of branching and parallel
6 evolution in myeloma. *Leukemia*. 2014;28(8):1705-1715. doi:10.1038/leu.2014.13
- 7 13. Murugaesu N, Wilson GA, Birkbak NJ, et al. Tracking the genomic evolution of
8 esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov*.
9 2015;5(8):821-831. doi:10.1158/2159-8290.CD-15-0412
- 10 14. Voss MH, Hakimi AA, Pham CG, et al. Tumor genetic analyses of patients with
11 metastatic renal cell carcinoma and extended benefit from mTOR inhibitor
12 therapy. *Clin Cancer Res*. 2014;20(7):1955-1964. doi:10.1158/1078-0432.CCR-13-
13 2345
- 14 15. Juric D, Castel P, Griffith M, et al. Convergent loss of PTEN leads to clinical resistance
15 to a PI(3)K α inhibitor. *Nature*. 2015;518(7538):240-244. doi:10.1038/nature13948
- 16 16. Watkins TBK, Lim EL, Petkovic M, et al. Pervasive chromosomal instability and
17 karyotype order in tumour evolution. *Nature*. 2020;587(7832):126-132.
18 doi:10.1038/s41586-020-2698-6
- 19 17. Zhao Y, Fu X, Lopez J, et al. Selection of metastasis competent subclones in the tumour
20 interior: TRACERx renal. *Research Square*. 2020. doi:10.21203/rs.3.rs-61979/v1
- 21 18. Rodriguez-Brenes IA, Komarova NL, Wodarz D. Tumor growth dynamics: insights
22 into evolutionary processes. *Trends Ecol Evol*. 2013;28(10):597-604.
23 doi:10.1016/j.tree.2013.05.020
- 24 19. Gerlee P. The model muddle: in search of tumor growth laws. *Cancer Res*.
25 2013;73(8):2407-2411. doi:10.1158/0008-5472.CAN-12-4355
- 26 20. Sun R, Hu Z, Sottoriva A, et al. Between-region genetic divergence reflects the mode
27 and tempo of tumor evolution. *Nat Genet*. 2017;49(7):1015-1024. doi:10.1038/ng.3891
- 28 21. Chkhaidze K, Heide T, Werner B, et al. Spatially constrained tumour growth affects
29 the patterns of clonal selection and neutral drift in cancer genomic data. *PLoS Comput*
30 *Biol*. 2019;15(7):e1007243. Published 2019 Jul 29. doi:10.1371/journal.pcbi.1007243
- 31 22. Hoefflin R, Lahrmann B, Warsow G, et al. Spatial niche formation but not malignant
32 progression is a driving force for intratumoural heterogeneity. *Nat Commun*.
33 2016;7:ncomms11845. Published 2016 Jun 13. doi:10.1038/ncomms11845

- 1 23. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of
2 the SWI/SNF complex gene PBRM1 in renal carcinoma [published correction appears
3 in Nature. 2012 Apr 5;484(7392):130]. Nature. 2011;469(7331):539-542.
4 doi:10.1038/nature09639
- 5 24. Peng J, Ma J, Li W, et al. Stabilization of MCRS1 by BAP1 prevents chromosome
6 instability in renal cell carcinoma. Cancer Lett. 2015;369(1):167-174.
7 doi:10.1016/j.canlet.2015.08.013
- 8 25. Cross W, Kovac M, Mustonen V, et al. The evolutionary landscape of colorectal
9 tumorigenesis. Nat Ecol Evol. 2018;2(10):1661-1672. doi:10.1038/s41559-018-0642-
10 z
- 11 26. Zhai W, Lim TK, Zhang T, et al. The spatial organization of intra-tumour heterogeneity
12 and evolutionary trajectories of metastases in hepatocellular carcinoma. Nat Commun.
13 2017;8:4565. Published 2017 Feb 27. doi:10.1038/ncomms14565
- 14 27. Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and
15 phenotypic evolution driven by selective pressure from the microenvironment. Cell.
16 2006;127(5):905-915. doi:10.1016/j.cell.2006.09.042
- 17 28. Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. A spatial
18 model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature.
19 2015;525(7568):261-264. doi:10.1038/nature14971
- 20 29. Noble R, Burri D, Kather JN, et al. Spatial structure governs the mode of tumour
21 evolution. 2019. doi: <https://doi.org/10.1101/586735>
- 22 30. Driessens G, Beck B, Ccauwe A, Simons BD, Blanpain C. Defining the mode of tumour
23 growth by clonal analysis. Nature. 2012;488(7412):527-530. doi:10.1038/nature11344
- 24 31. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral
25 tumor evolution across cancer types. Nat Genet. 2016;48(3):238-244.
26 doi:10.1038/ng.3489
- 27 32. Crawley MJ, Harral JE. Scale dependence in plant biodiversity. Science.
28 2001;291(5505):864-868. doi:10.1126/science.291.5505.864
- 29 33. Horner-Devine MC, Lage M, Hughes JB, Bohannon BJ. A taxa-area relationship for
30 bacteria. Nature. 2004;432(7018):750-753. doi:10.1038/nature03073
- 31 34. Zhou J, Kang S, Schadt CW, Garten CT Jr. Spatial scaling of functional gene diversity
32 across various microbial taxa. Proc Natl Acad Sci U S A. 2008;105(22):7768-7773.
33 doi:10.1073/pnas.0709016105

- 1 35. Fiore VF, Krajnc M, Quiroz FG, et al. Mechanics of a multilayer epithelium instruct
2 tumour architecture and function [published correction appears in Nature. 2020 Sep
3 11;:]. Nature. 2020;585(7825):433-439. doi:10.1038/s41586-020-2695-9
- 4 36. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The Extensible Neuroimaging
5 Archive Toolkit: an informatics platform for managing, exploring, and sharing
6 neuroimaging data. Neuroinformatics. 2007;5(1):11-34. doi:10.1385/ni:5:1:11

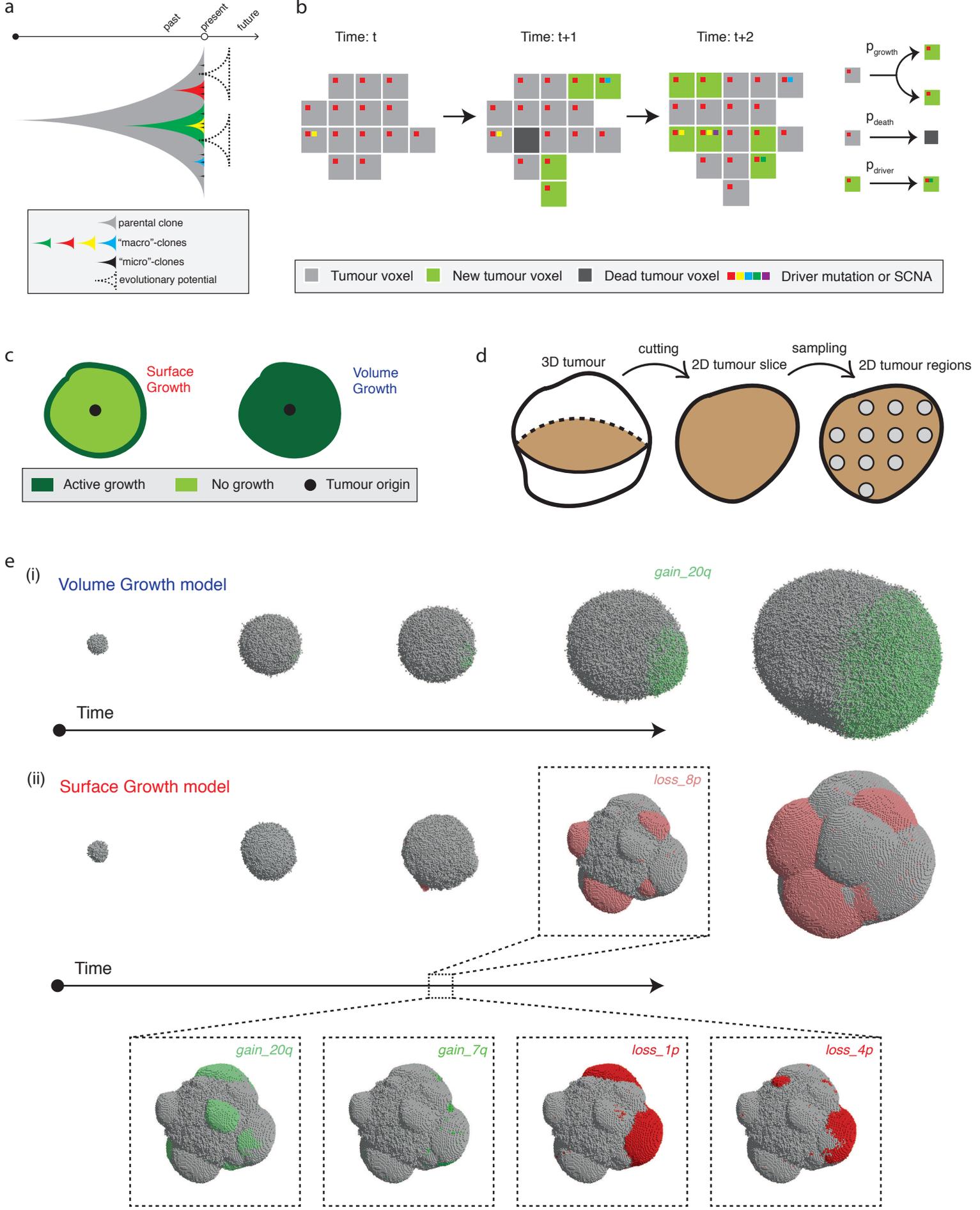


Figure 1

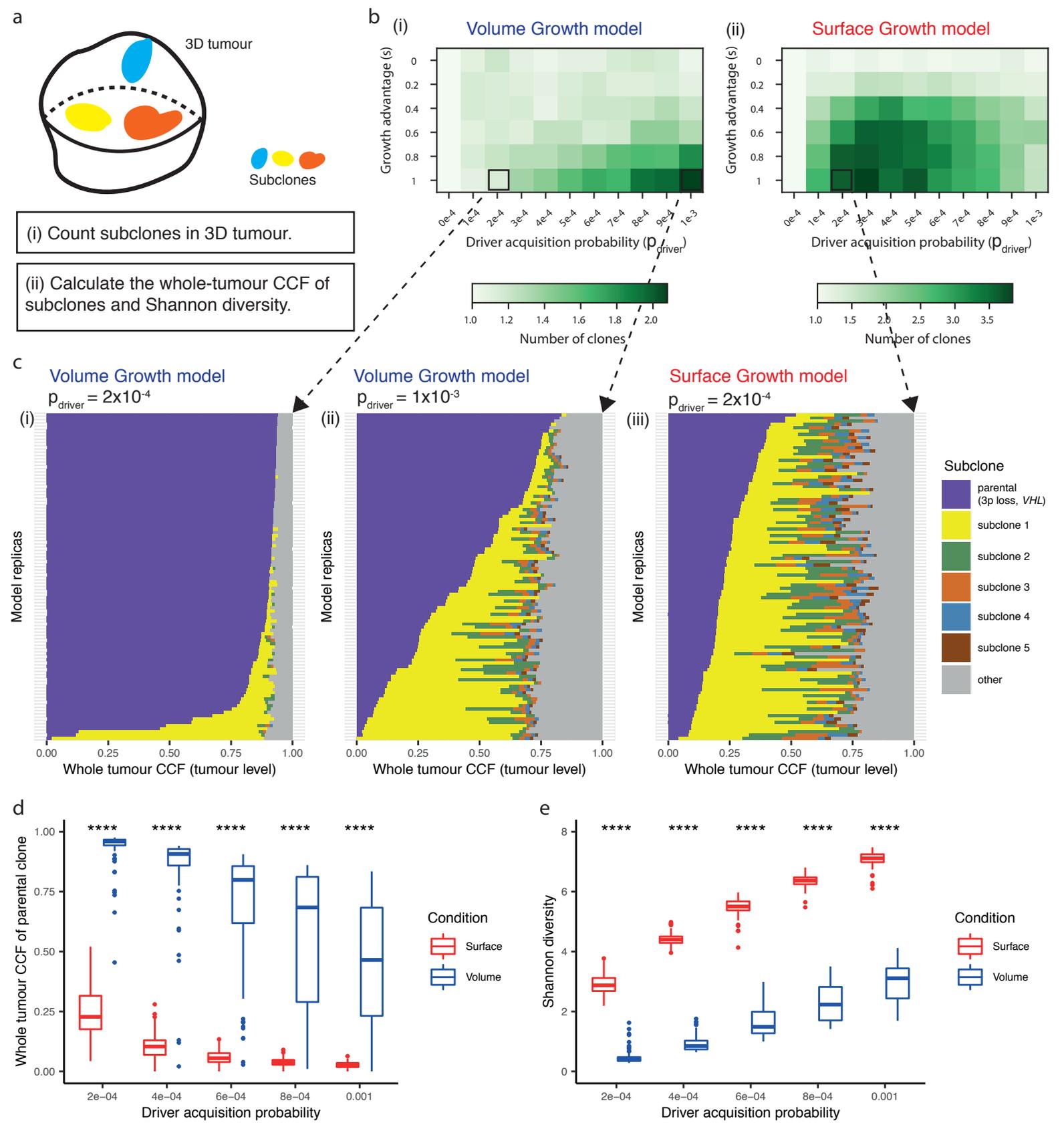


Figure 2

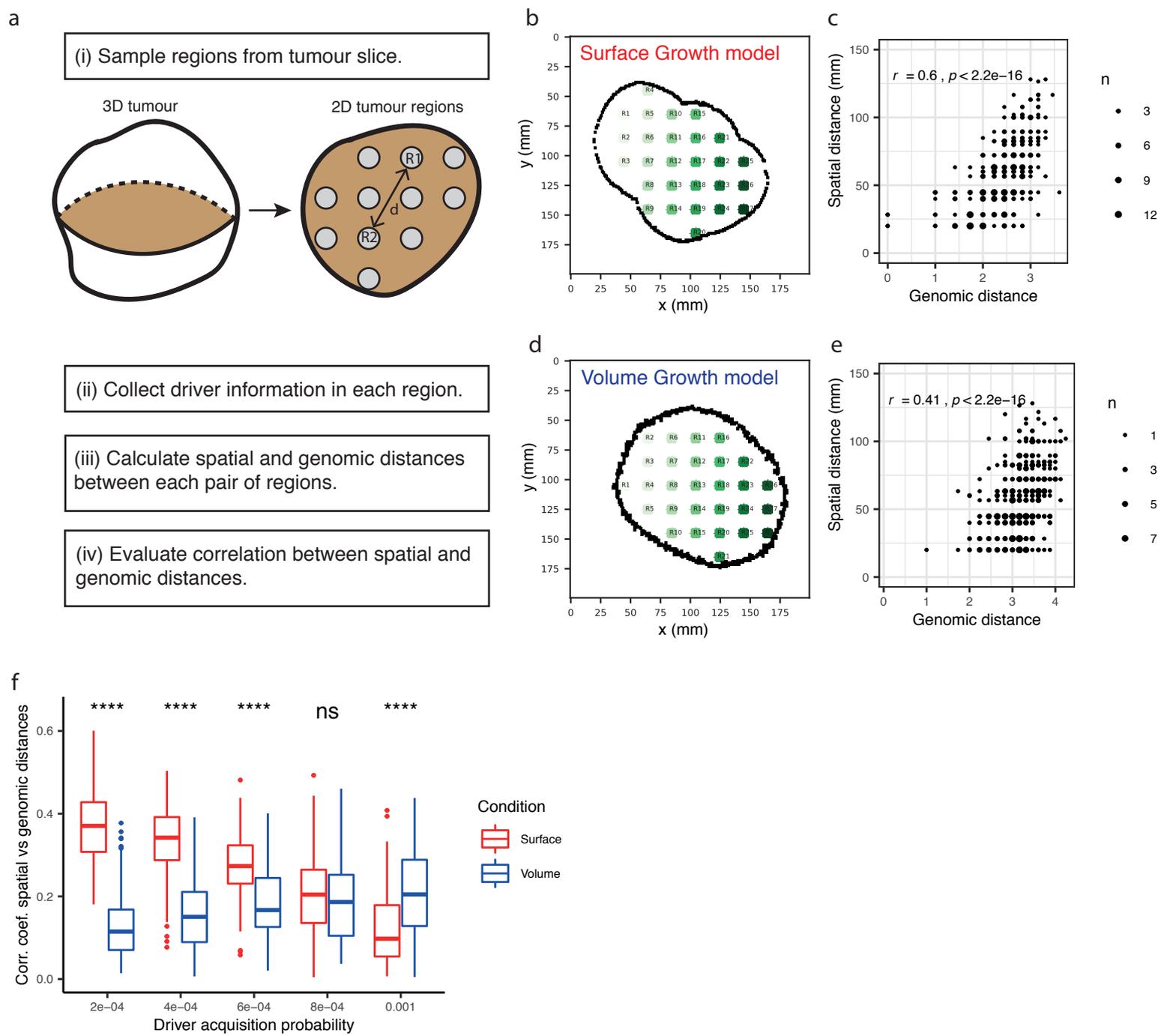


Figure 3

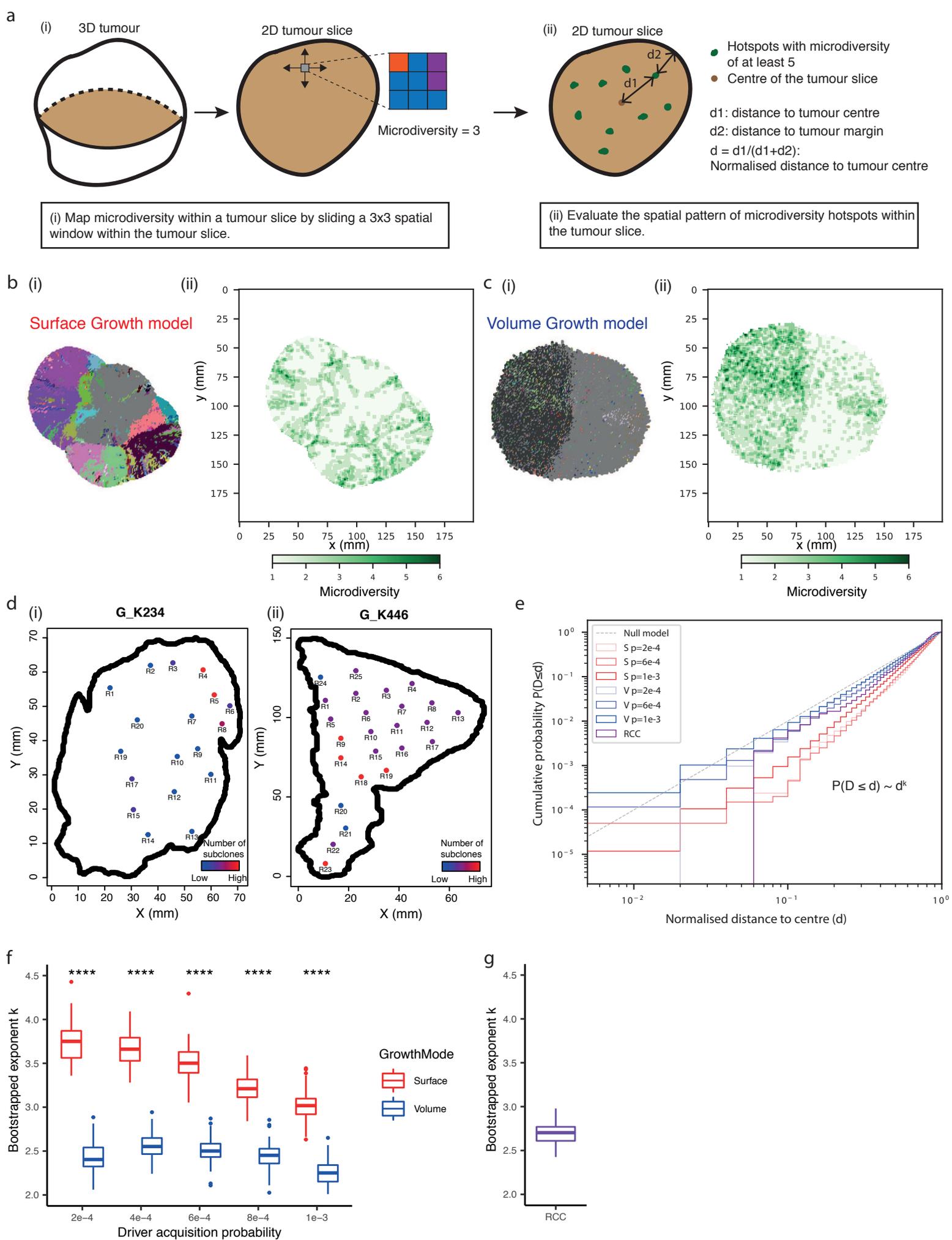


Figure 4

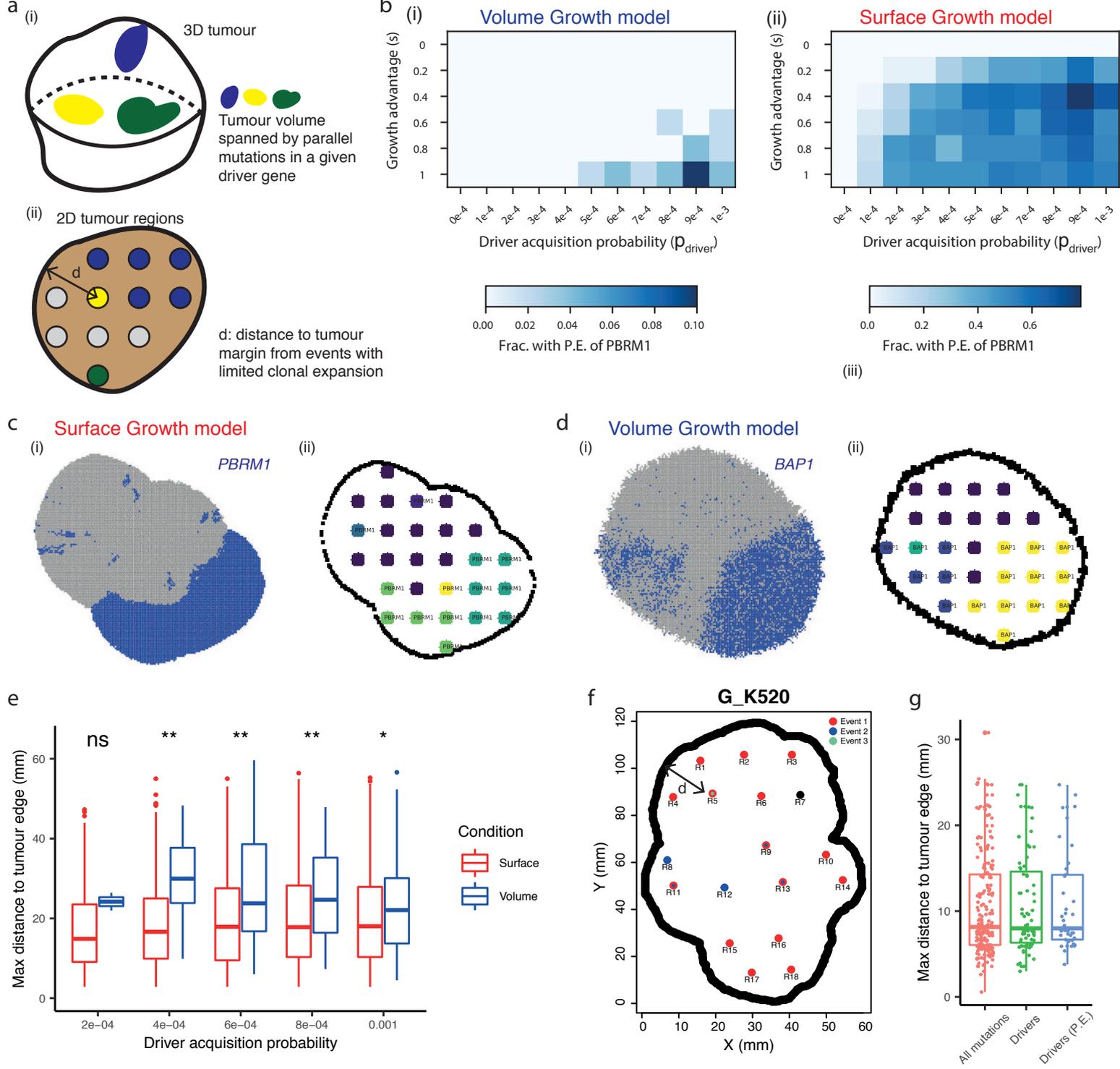


Figure 5

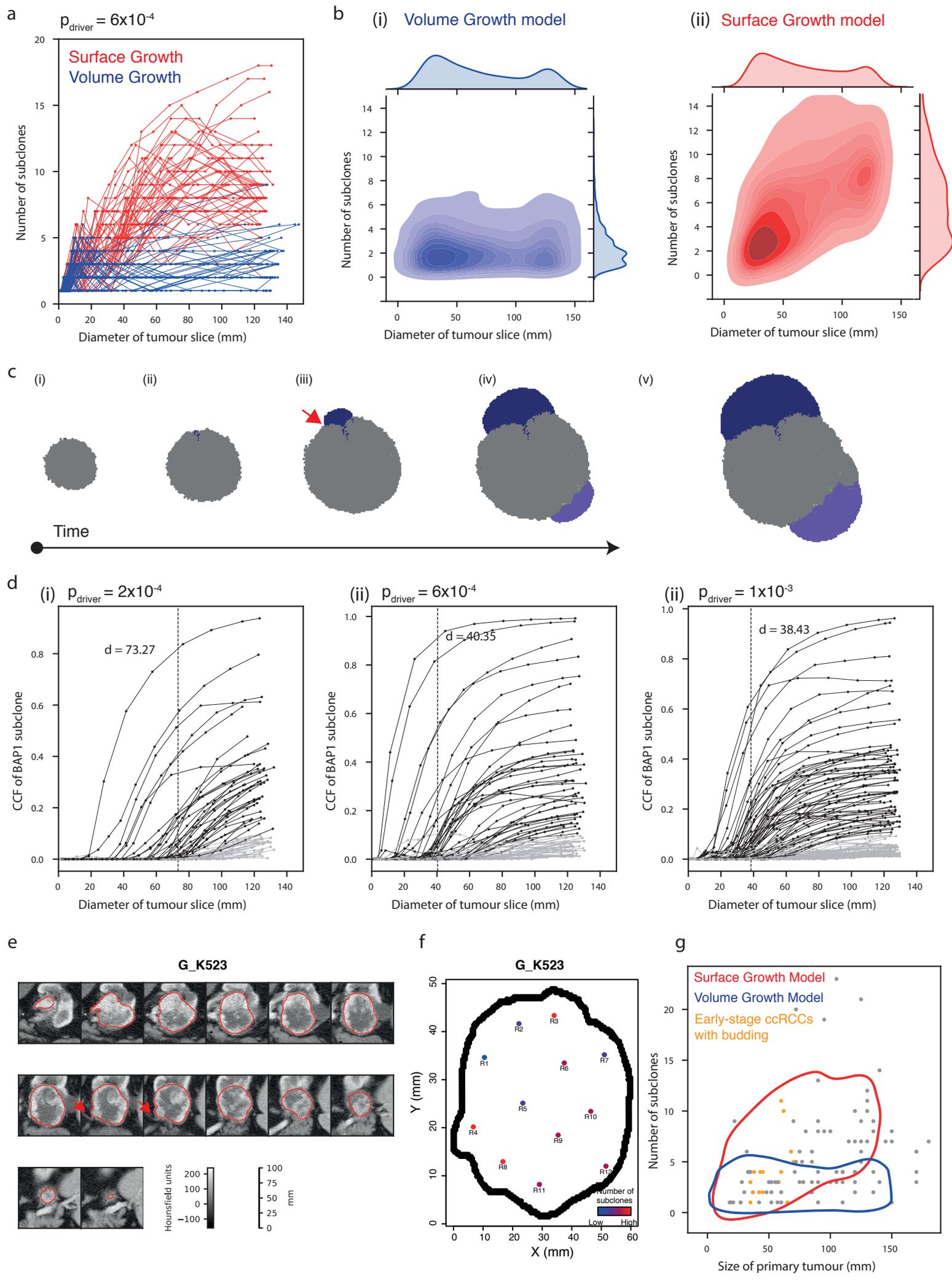


Figure 6

Figures

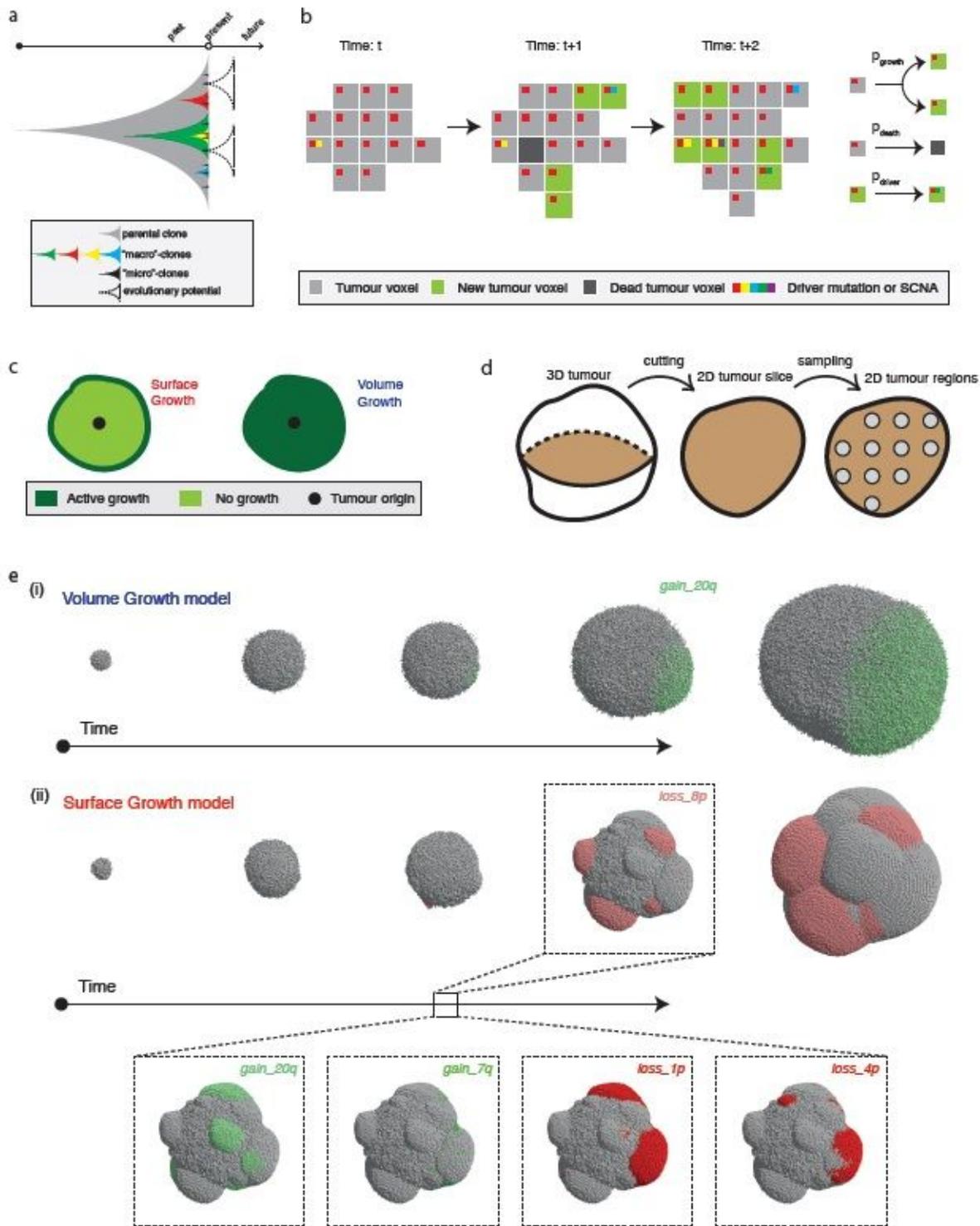


Figure 1

Construction of in silico tumours. (a) Schematic figure illustrating future evolutionary trajectories delineated by present under detected subclones. (b) Schematic figure of probabilistic growth, death, and driver acquisition in a coarse-grained cellular automaton model. (c) Schematic figure of two growth

modes: "Surface Growth" with proliferation predominating at the tumour surface and "Volume Growth" with proliferation throughout the tumour volume. (d) Schematic figure of three levels of measurements: from three-dimensional (3D) tumour to two-dimensional (2D) tumour slice and 2D tumour regions within the slice. (e) Representative in-silico tumours under Volume Growth (i) and Surface Growth (ii), respectively, from a 3D view. Tumour voxels harbouring select drivers, as indicated in the figure, are colour coded. Tumour voxels harbouring gain of chromosome arm events are in green; tumour voxels harbouring loss of chromosome arm events are in red. Different shades of greens or reds are employed to reflect different driver events.

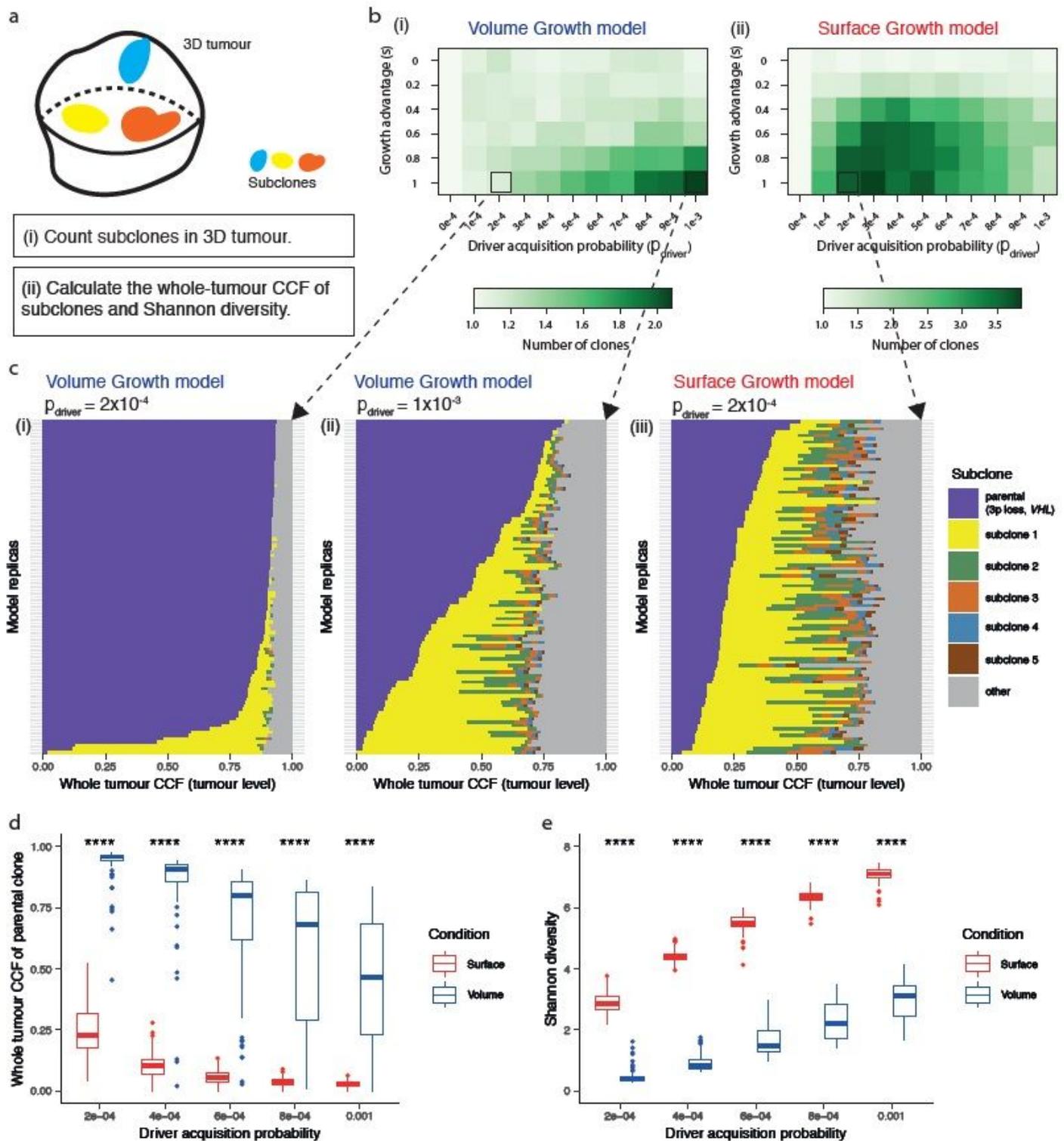


Figure 2

Clonal diversity in the whole tumour. (a) Schematic figure for the whole-tumour analysis of clonal diversity. (b) Heatmap showing the average number of clones (i.e., parental clone and subclones) with respect to driver acquisition probability and proliferative advantage in the Volume Growth (i) and Surface Growth (ii) models. The average is calculated from 50 in silico tumours per parameter condition. Clones with a whole-tumour cancer cell fraction (CCF) of at least 0.05 are counted. (c) Whole-tumour CCF

of parental and largest subclones in in silico tumours under Volume Growth (i-ii) and Surface Growth (iii), respectively. Driver acquisition probabilities in these sets of simulations are $\$,\#-./\# = 2 \times 1001$ in (i), $1 \times 100!$ in (ii), 2×1001 in (iii), respectively. “Parental (3p loss, VHL)” clone is shown along with up to five subclones with a whole-tumour CCF of 0.01 or higher. All remaining subclones are represented in the “other” group. (d) Whole-tumour CCF of parental subclones in in silico tumours under Volume Growth and Surface Growth with varying driver acquisition probabilities. $N = 100$ for each condition. (e) Shannon diversity index in in silico tumours under Volume Growth and Surface Growth with varying driver acquisition probabilities. $N = 100$ for each condition. Statistical annotations in (d-e) reflect two-sided Wilcoxon tests: “****” indicates $1 \leq 0.0001$.

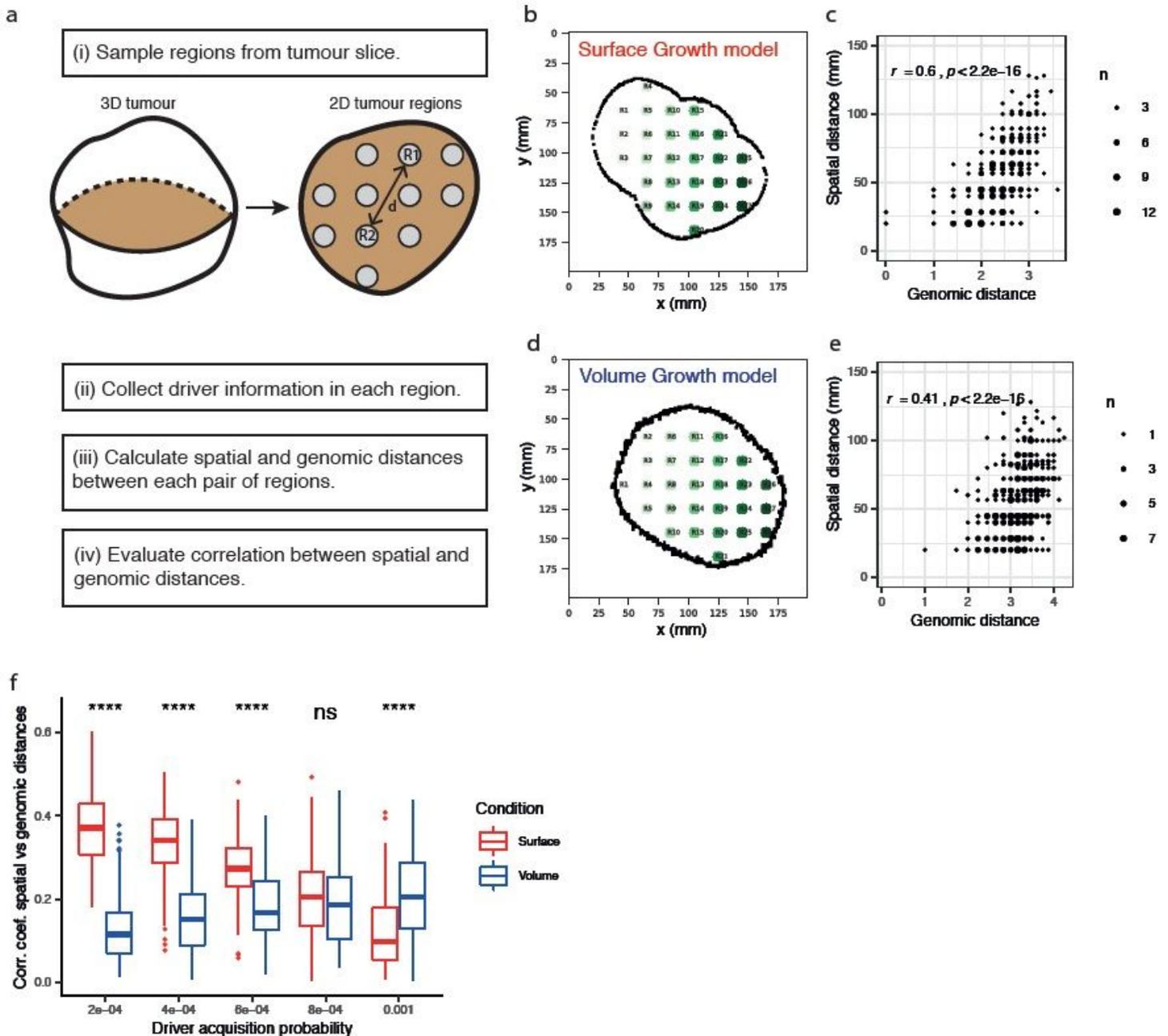


Figure 3

Between-region genomic divergence. (a) Schematic figure and procedure for the analysis in 2D tumour regions. (b) Spatial locations of regional samples with tumour edge indicated in a representative in silico tumour under Surface Growth. (c) Between-region spatial distance against genomic distance derived from the representative in silico tumour in panel b. Overlapping data points are counted (“n”). Pearson correlation (“r”) and p value are indicated in the figure. (d-e) The same analysis as described in (b-c) for a representative in silico tumour with Volume Growth. (f) Pearson correlation between spatial distance and genomic distance in in silico tumours under Volume Growth and Surface Growth with varying driver acquisition probabilities. N = 100 for each condition. Statistical annotations in (f) reflect two-sided Wilcoxon tests: “****” indicates $1 \leq 0.0001$ and “ns” indicates no statistical significance.

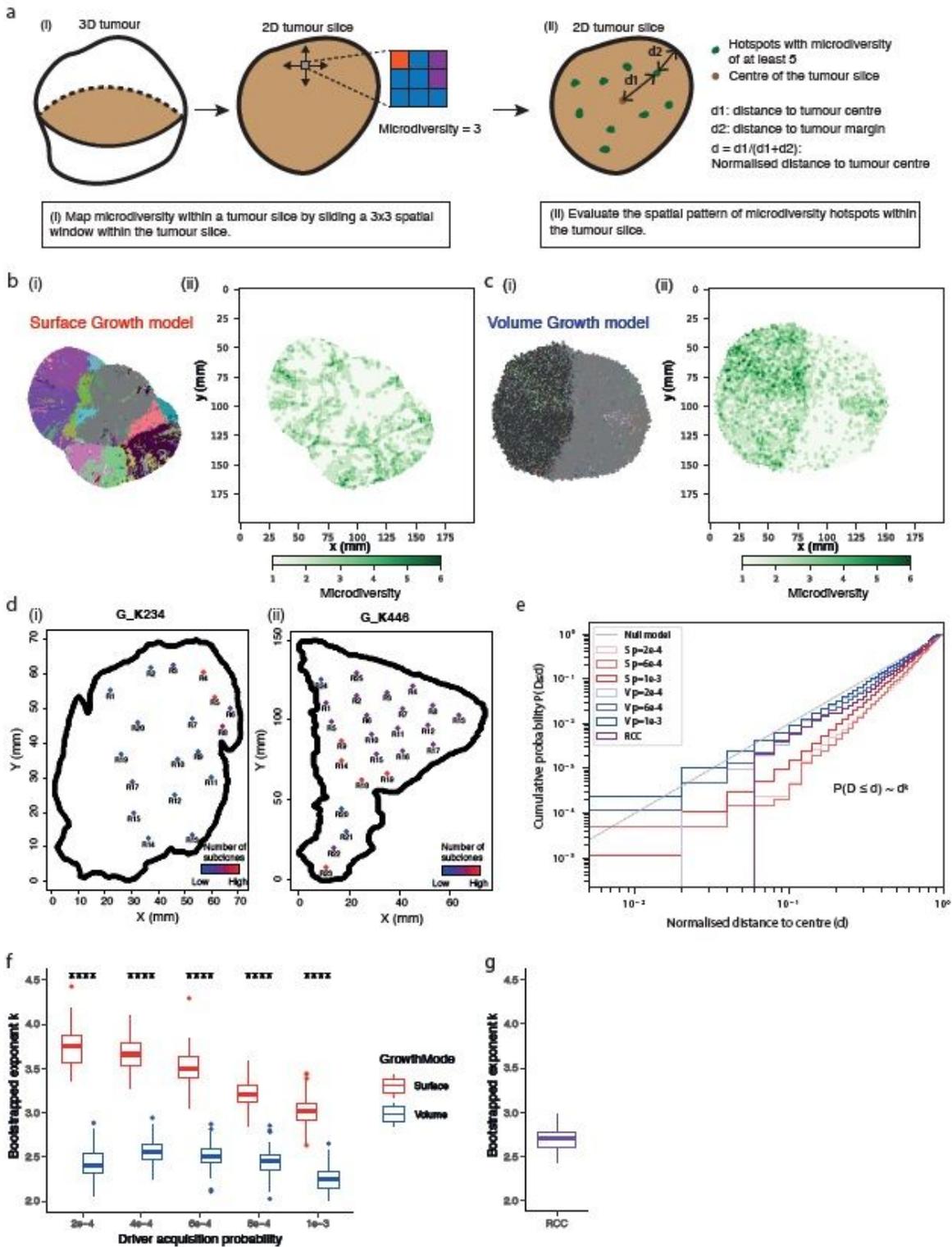


Figure 4

Spatial features of clonal diversity. (a) Schematic figure and procedure for the analysis within a 2D tumour. (b) Spatial maps of subclones (i) and microdiversity (ii) in a representative in-silico tumour under Surface Growth. (c) The same analysis as described in (b) for a representative in-silico tumour under Volume Growth. (d) Maps of regional biopsies with the number of subclones within a biopsy colour coded in two cases (G_K234 and G_K446) in the TRACERx Renal study. Hues from red to purple to blue

reflect decreasing number of subclones. “Low” reflect zero subclones, while “High” reflects the maximum number of subclones found in any region (i.e., 4 subclones in G_K234, 3 subclones in G_K446). (e) Cumulative probability distribution, $1(2 \leq 0)$, of the normalised distance to tumour centre in in silico tumours under Surface Growth and Volume Growth and in ccRCC tumours. Three sets of in-silico tumours with different driver acquisition probabilities are shown. $N = 100$ for each model condition. “S” and “V” in the figure reflect Surface Growth and Volume Growth, respectively. “ $p=2e-4$ ” reflects a driver acquisition probability of $2e-4$. 606 patient tumour (PT) regions from 54 ccRCC tumours are considered for the experimental analysis. (f) Bootstrapped power law exponent 4, as in $1(2 \leq 0) \sim 02$, fitted to cumulative probability distribution of normalised distance to tumour centre in each of bootstrap samples. For microdiversity hotspots under each model condition, 100 bootstrap samples are generated by randomly sampling 400 hotspots with replacement. (g) Bootstrapped power law exponent 4 in ccRCC tumours. Driver acquisition probabilities used in representative simulations are: $\$ \# \cdot / \# = 2 \times 1001$ in both in-silico tumours under Surface Growth (b) and in-silico tumours under Volume Growth (c). Statistical annotations in (f) reflect two-sided Wilcoxon tests: “****” indicates $1 \leq 0.0001$.

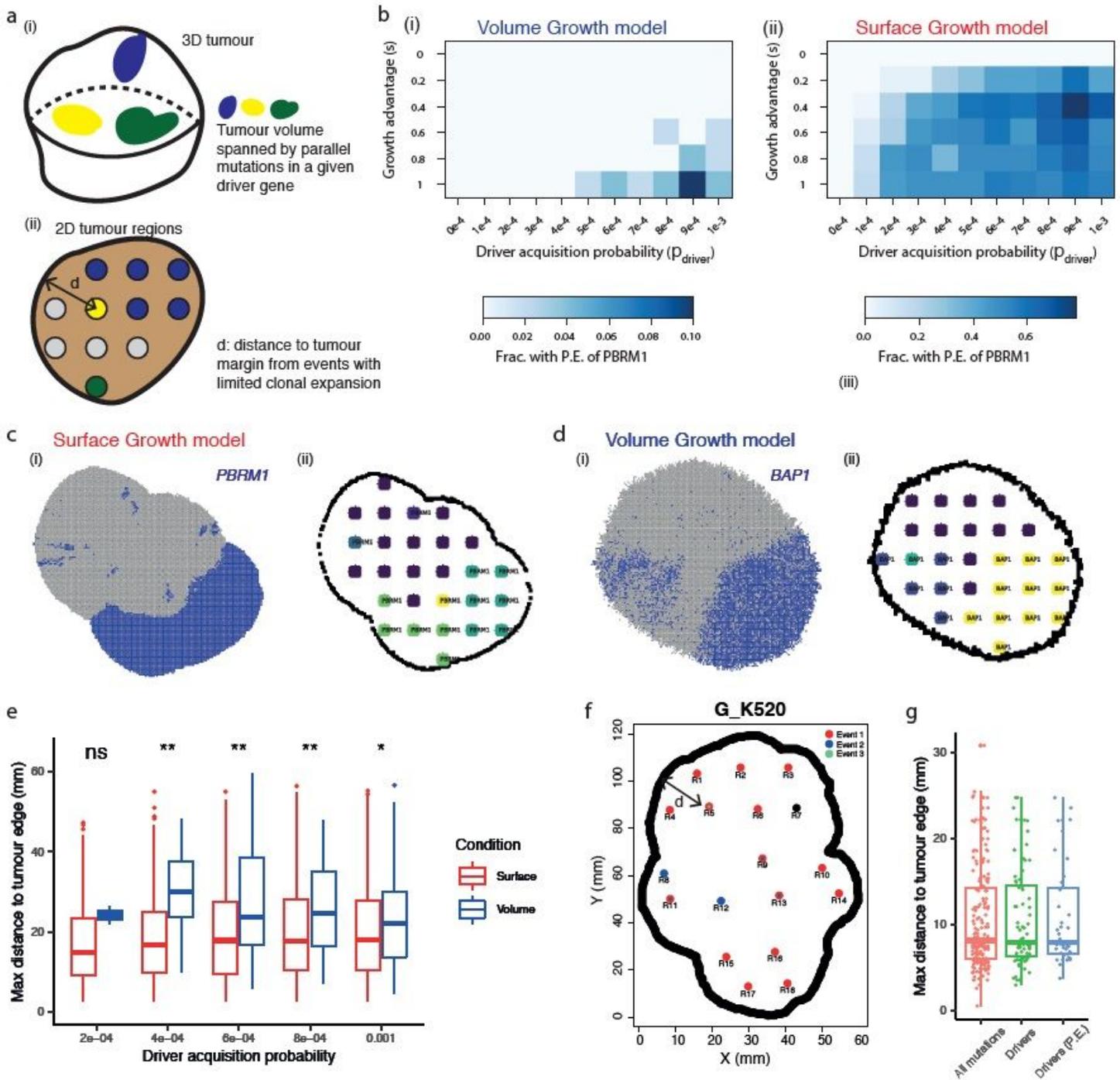


Figure 5

Frequency and spatial features of parallel evolution. (a) Schematic figure for measuring parallel evolution in the 3D tumour (i) and for analysing parallel mutational events within 2D tumour regions (ii). (b) Heatmap showing the fraction of in silico tumours that have parallel evolution in PBRM1 with respect to driver acquisition probability and growth advantage in the Volume Growth (i) and Surface Growth (ii) models. The fraction is calculated based on 50 in silico tumours per parameter condition. In silico tumours are counted, if there are two or more parallel mutational events in PBRM1 with a whole-tumour CCF of at least 0.05. (c-d) Spatial distribution of PBRM1 or BAP1 mutation (i) and its parallel events (ii) in

a representative in silico tumour under Surface Growth (c) and Volume Growth (d), respectively. In panel (ii), different colours overlaid with "PBRM1" or "BAP1" reflect different parallel mutation events. (e) Maximum distances to the tumour edge from parallel mutations in PBRM1 or BAP1 that span up to two regions in in silico tumours under Volume Growth and Surface Growth with varying driver acquisition probabilities. $N = 100$ for each condition. (f) Maps of regions containing parallel mutations in PBRM1 in a representative case (G_K520) in the TRACERx Renal study. Distinct parallel mutations are coloured differently. For regions containing more than two parallel mutations, two colours are applied simultaneously. Doubleheaded arrow indicates a measurement of distance to tumour edge in this example. (g) Maximum distances from mutational events that span up to two regions in ccRCC tumours in the TRACERx Renal study. For comparison, all mutations ("All mutations"), mutations in driver genes ("Drivers"), and parallel mutations in driver genes ("Drivers (P.E.)") are shown as separate groups. Driver acquisition probabilities used in representative simulations are: $\frac{\$,\#-./\#}{\#} = 2 \times 1001$ in in silico tumours under Surface Growth and $\frac{\$,\#-./\#}{\#} = 1 \times 100!$ in in silico tumours under Volume Growth. Statistical annotations in (e) reflect two-sided Wilcoxon tests: "**" indicates $1 \leq 0.01$, "*" indicates $1 \leq 0.05$, and "ns" indicates no statistical significance.

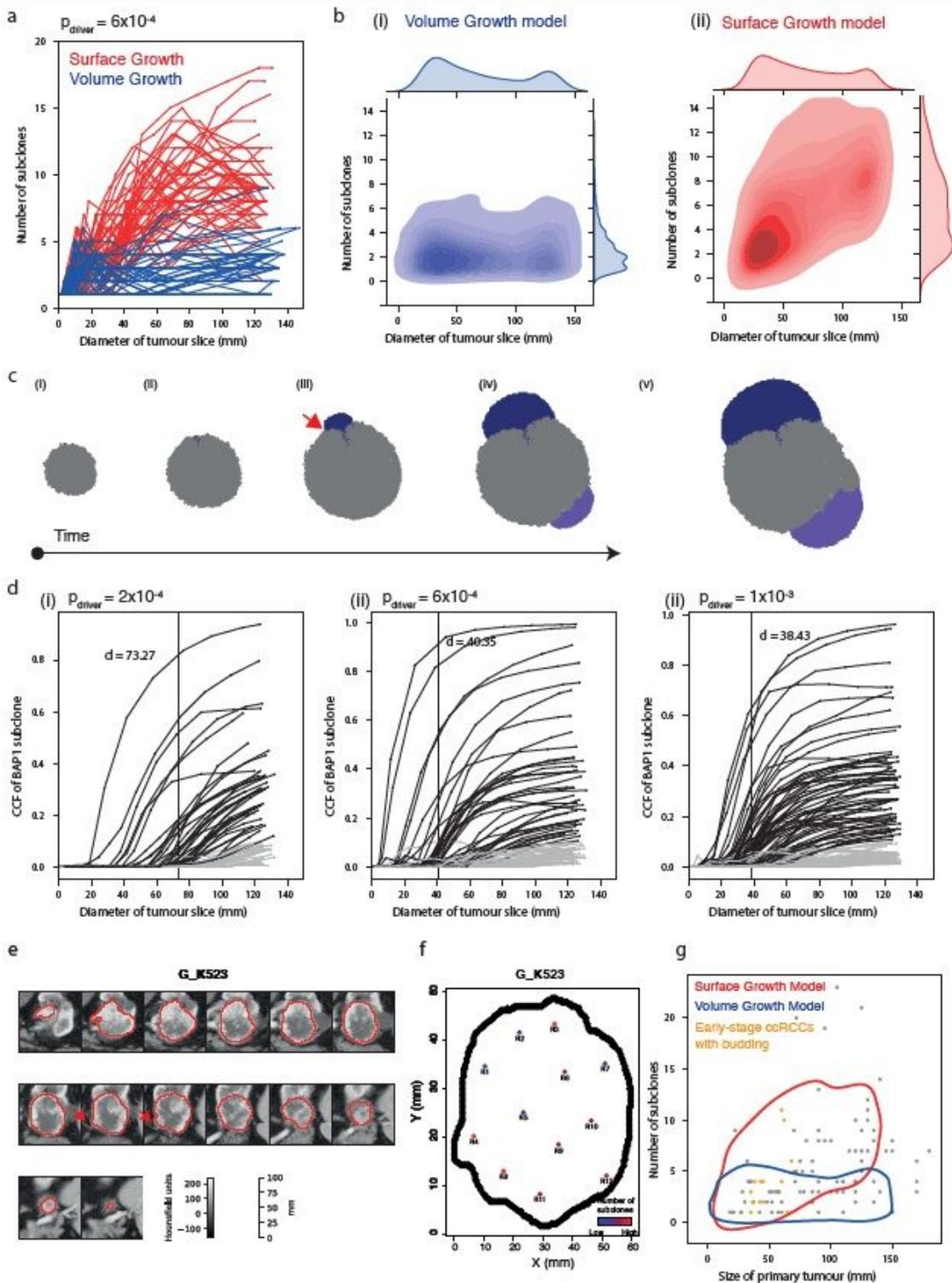


Figure 6

Early predictive indicators of evolutionary trajectories. (a) The number of subclones as a function of the diameter of a 2D tumour slice in in silico tumours under Surface Growth and under Volume Growth, respectively. $N = 50$ simulations with $\$,\#./\# = 6 \times 1001$ are shown for each condition. (b) Kernel density estimation (KDE) with respect to the number of subclones and the diameter of a 2D tumour slice in in silico tumours under Volume Growth (i) and under Surface Growth (ii). Each KDE plot is based on 250

simulations (50 per condition) under 5 conditions with $\$,\#-./\# = 2 \times 1001, 4 \times 1001, 6 \times 1001, 8 \times 1001, 1 \times 100!$. (c) The spatial patterns of parallel mutations in BAP1 over time in a representative in silico tumour under Surface Growth. The red arrow indicates a budding structure at early stage of subclonal expansion. (d) The CCFs of BAP1 clones within the 2D tumour slice as a function of the diameter of a 2D tumour slice in each of 50 in silico tumours. Only BAP1 clones with an ultimate CCF of 0.01 or higher are shown. For those with an ultimate CCF of 0.1 or higher, the diameters of tumour slices at which these clones first became detectable are recorded, with the median indicated by the dashed line. The others with an ultimate CCF of 0.1 or lower are shown in grey. (e) Axial image in the corticomedullary contrast phase of a representative case (G_K523) showing budding structure on the tumour surface (red arrow). Outlines in red were drawn giving volumetric tumour coverage by an oncologist (S.S.) and a radiologist (D.A.) (f) Maps of tumour regions with the number of subclones colour coded in a representative case (G_K523). Hues from red to purple to blue reflect decreasing number of subclones. “Low” reflect zero subclones, while “High” reflects the maximum number of subclones found in any region (i.e., 4 subclones in G_K523). (g) The number of subclones as a function of ultimate tumour size in the TRACERx Renal study, overlaid with kernel density estimation based on imulated data. Tumours with a size smaller than 7 cm and with radiologically evident budding structures on the tumour surface are highlighted (orange). Contours reflect 90% probability density based on in silico tumours under Surface Growth (red) and under Volume Growth (blue), respective in Figure 6b.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NEESubmissionSupplementalInfotextfigurescombinedXF.pdf](#)