Supplemental Information for:

Title:

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

Authors:

Xiao Fu[1*], Yue Zhao[2,3,4,5*], Jose I. Lopez[6], Andrew Rowan[2], Lewis Au[7,16], Annika Fendler[7], Steve Hazell[8], Hang Xu[2], Stuart Horswell[9], Scott Shepherd[7,16], Lavinia Spain[7], Fiona Byrne[7], Gordon Stamp[10], Tim O'Brien[11], David Nicol[12], Marcellus Augustine[2], Ashish Chandra[13], Sarah Rudman[14], Antonia Toncheva[15], Andrew Furness[7,16], Lisa Pickering[16], Matthew Orton[17], Simon Doran[17], Dow-Mu Koh[17], Christina Messiou[17], Derfel ap Dafydd[17], Santosh Kumar[17], James Larkin[16], Charles Swanton[2,3,18], Erik Sahai[19#], Kevin Litchfield[2,3#], Samra Turajlic[7,16#] on behalf of the TRACERx Renal Consortium, Paul A Bates[1#]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Supplemental Note 1. Alternative implementation of fitness advantage conferred by drivers.

For an exploratory purpose, we evaluated the an alternative implementation of fitness advantage conferred by drivers (**Supplemental Figure 2a**, see Methods). Specifically, we focused on conditions of driver acquisition probability highlighted for Surface Growth model ($p_{driver} = 2 \times 10^{-4}$) and for Volume Growth model ($p_{driver} = 1 \times 10^{-3}$) in the Main Text (Main Figure 2). Furthermore, three scenarios were considered to reflect increasing amount of growth probabilities endowed by drivers on average (**Supplemental Figure 2b**). Consistent with the implementation of fitness advantage highlighted in the main text, for all of these scenarios, Volume Growth (**Supplemental Figure 2c**), even with a greater p_{driver} employed, appears to result in less extensive subclonal diversification that Surface Growth (**Supplemental Figure 2d**). Additionally, in the Volume Growth model, clonal proportions across repeated simulations with the second scenrio (i.e., $\min(s_k) = 0.015, \Delta s_k = 0.005$) are similar to those highlighted in the Main Text (**panel (ii) in Main Figure 2c**), suggesting a minimal impact of the implementation of fitness advantage on the clonal evolution. In constrast, in the Surface Growth model, repeated simulations with the second scenario displayed a greater extent of subclonal diversification (i.e., smaller size of largest subclone "subclone 1") than those highlighted in the Main Text (**panel (iii) in Main Figure 2c**), suggesting that Surface Growth mode is more sensitive to the implementation of fitness advantage conferred by drivers.

Supplemental Note 2. Whole-tumour frequency of parallel evolution in *in silico* tumours.

Due to the varying times of birth and growth advantages, subclones that acquire distinct parallel mutations in the same gene should grow to different sizes. To examine the overall prevalence, the number of parallel mutations in each driver gene was recorded and the average frequency was calculated among *in silico* tumours. As expected, we observed that the frequency of parallel mutations in the same driver gene decayed as a function of the detection limit, or resolution of detection (**Supplemental Figure 3a-b**). Moreover, parallel mutations in either *BAP1* or *PBRM1* were less frequent than those in other driver genes with measurements taken at low detection limits but more frequent at high detection limits. This trend reflected the formation of more established subclone(s) harbouring a parallel mutation in *BAP1* or *PBRM1* and, as a consequence, the remaining proportion of the tumour mass had a smaller capacity for accommodating secondary parallel mutations.

Supplemental Note 3. Parallel evolution in ccRCCs

Due to the limits of spatial sampling and sequencing, parallel gene mutations were detected in 28 out of 66 tumours. A total of 71 of 114 parallel events in 18 tumours were alterations in known ccRCC drivers, including *ARID1A*, *BAP1*, *KDM5C*, *PBRM1*, *PTEN*, *SETD2*, and

VHL. Among these drivers, parallel evolution of alterations in *PBRM1*, *SETD2*, and *BAP1* were most frequent, in 6, 5, and 4 tumours, respectively. Consistent with previous observations (Turajlic et al 2018), multiple parallel instances of *PBRM1* mutations co-existed, with one well established event spanning more than 10 regions and additional instances of *PBRM1* mutation displaying less expansion; one representative case was highlighted in the main text ("G_K520").

Supplemental Note 4. Evolutionary replay *in silico*

An *in-silico* tumour under Surface Growth was prepared for evolutionary replay (**Supplemental Figure 5a**). Using historical states of this tumour as a common starting state, 50 new *in-silico* tumours were grown (**Supplemental Figure 5b**). While re-grown tumours starting from earlier historical states displayed markedly divergent patterns of subclones in the end (**panel (i) in Supplemental Figure 5c**), those grown from historical states later than the emergence of the budding structure appeared very similar to the original pattern of subclones (**panel (ii-iv) in Supplemental Figure 5c**). Quantitatively, this was evidenced by a decreasing divergence in the eventual Shannon diversity after evolutionary replay as a function of the size of the starting tumour state (**Supplemental Figure 5d**). Similarly, this trend of decreasing divergence was also noted in in-silico tumours under Surface Growth (**Supplemental Figure 5e**) or under Volume Growth (**Supplemental Figure 5f**) with a greater probability of driver acquisition. This finding suggested that budding structures in a tumour under surface growth could indicate future evolutionary trajectories.

Supplemental Note 5. Scaling between clonal diversity and sampling area

As another perspective of characteristing the extent of subclonal diversification, we ask how the extent of clonal expansion of a genomic alteration could relate to continuing subclonal diversification subsequent to its acquisition (**Supplemental Figure 6a**). Specifically, we examined in the *in silico* setting whether a quantitative relationship existed between the number of subclones ("N") within the area spanned by parallel events of *PBRM1* or *BAP1* and the size of that area ("A"), inspired by the taxa-area relationship underlying the macroevolution of species in various ecosystems (Crawley and Harral 2001, Horner-Devine et al 2004, Zhou et al 2008). Interestingly, the N-A relationship depicted two regimes. For parallel mutational events with limited clonal expansion (up to 100 mm^2), the number of subclones increased sub-linearly with the size of the area (i.e., $N \sim A^z$ with $z < 1$) (**Supplemental Figure 6b**), reflecting the stochasticity in clonal diversification upon growth when A is small. By contrast, for parallel mutational events under greater extent of clonal expansion, the number of subclones increased nearly linearly with the size of the area (i.e., $N \sim A$), regardless of growth types (**Supplemental Figure 6b**). Moreover, for a range of driver acquisition probabilities, the Surface Growth model depicted a larger scaling exponent that the Volume Growth model, suggesting a greater extent of subclonal diversification as a function of an increase in the area (**Supplemental Figure 6c**).

To explore the N-A relationship in the TRACERx Renal study, the number of subclones was counted in the regions spanned by each of the 1665 mutational and SCNA events in 66 tumours. The number of subclones increased with the area sub-linearly (i.e., $N \sim A^z$ with $z < 1$), consistent with the scaling relationship at small values of A for the *in-silico* tumours (**Supplemental Figure 6d**). Moreover, when split according to the status of disease relapse, the subset of tumours with relapse depicted a slower increase in clonal diversity as a function of the area than the others (**Supplemental Figure 6d**), suggesting that the extent of subclonal diversification as a function of area might be a feature of differential clinial outcomes.

Supplemental Figure 1. Assumed growth advantages endowed by drivers.

(a) Ki67 immunohistochemistry (IHC) score in patient tumour (PT) regions where a particular driver is present.

(b) A table containing assumed levels of growth probabilities endowed by individual drivers.

Supplemental Figure 2. Alternative implementation of growth advantage endowed by drivers.

(a) Schematic figure of probabilistic growth of tumour voxels. The total growth probability of a tumour voxel is cumulatively contributed by all drivers harboured.

(b) A table containing assumed growth probabilities added by each driver (p_{growth_k} for a driver k). Three scenarios were explored, indicated in the table by p_{growth_k} of the weakest driver, namely, min(p_{growth_k}), and the difference in p_{growth_k} between consecutive two drivers in their advantages, namely, Δp_{growth_k}.

(c-d) Whole-tumour cancer cell fraction (CCF) of parental and largest subclones in *in silico* tumours under Volume Growth (c) and Surface Growth (d), respectively, under the indicated parameter conditions. "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. N = 100 for each condition.

Supplemental Figure 3. Clonal diversity in the whole tumour under model conditions with small driver acquisition probabilities (p_{driver}).

(a-b) Whole-tumour CCF of parental and largest subclones in *in silico* tumours under Surface Growth (a) and Volume Growth (b), respectively, under the indicated parameter conditions. "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 $P \leq 0.0001$.

Supplemental Figure 4. Clonal diversity in the whole tumour under model conditions with zero or small relative growth advantages (s).

(a-b) Reproduced from **Main Figure 2a-b** for reference to parameter domains. (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 CCF of at least 0.05 are counted.

(c-d) Whole-tumour CCF of parental and largest subclones in *in silico* tumours under Surface Growth and Volume Growth, respectively, under the indicated parameter conditions. "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.

Supplemental Figure 5. Quantile-Quantile (Q-Q) plots of observed versus fitted distributions of microdiversity hotspots.

From left to right represent Surface Growth Model, Volume Growth Model and experimental data, with conditions indicated within figures. "S" and "V" in the figure reflect Surface Growth and Volume Growth, respectively. "p=2e-4" reflects a driver acquisition probability of 2e-4. In addition, the median fitted power law exponent k, as in $P(D \leq d) \sim d^k$, from bootstrapping (in **Main Figure 4f-g**) is indicated within figures.

Supplemental Figure 6. Parallel evolution.

(a-b) Average frequency of parallel mutations per driver gene among 100 cases in *in silico* tumours under Surface Growth (a) and Volume Growth (b), respectively.

(c) Birth time of a parallel event in *PBRM1* or *BAP* mutation against the maximum distance from the tumour margin in *in silico* tumours under Surface Growth.

(d) Birth time of a parallel event in *PBRM1* or *BAP* mutation in *in silico* tumours under Surface Growth, grouped by the number regions spanned.

(e-f) Same analysis as in (c-d) for in *in silico* tumours under Volume Growth.

Supplemental Figure 7. Time-course study on clonal diversity.

(a) The number of subclones as a function of the diameter of a 2D tumour slice in *in silico* tumours under Volume Growth with $p_{driver} = 2 \times 10^{-4}$ (i), $p_{driver} = 6 \times 10^{-4}$ (ii), and $p_{driver} = 1 \times 10^{-3}$ (iii), respectively. N = 50 for each condition. Colours reflect different repeat simulations.

(b) same as (a) for Surface Growth.

Supplemental Figure 8. Evolutionary replay *in silico*.

(a) Schematic figure for description of evolutionary replay.

(b) Spatial maps of subclones over time in a representative *in silico* tumour under Surface Growth with $p_{driver} = 2 \times 10^{-4}$. (i)-(iv) mark historical tumour states selected as starting points for evolutionary replay.

(c) Spatial maps of subclones at the end of two evolutionary replays that use (i)-(iv) in (b) as starting states.

(d) Shannon diversity index at the end of evolutionary replays that start with different historical tumour states. From left to right reflect increasing sizes at which historical tumour states were collected for evolutionary replay.

(e-f) Evolutionary replay performed based on *in silico* tumours under Surface Growth (e) and under Volume Growth with $p_{driver} = 6 \times 10^{-4}$, respectively.

Supplemental Figure 9. Scaling patterns of clonal diversity with sampling area.

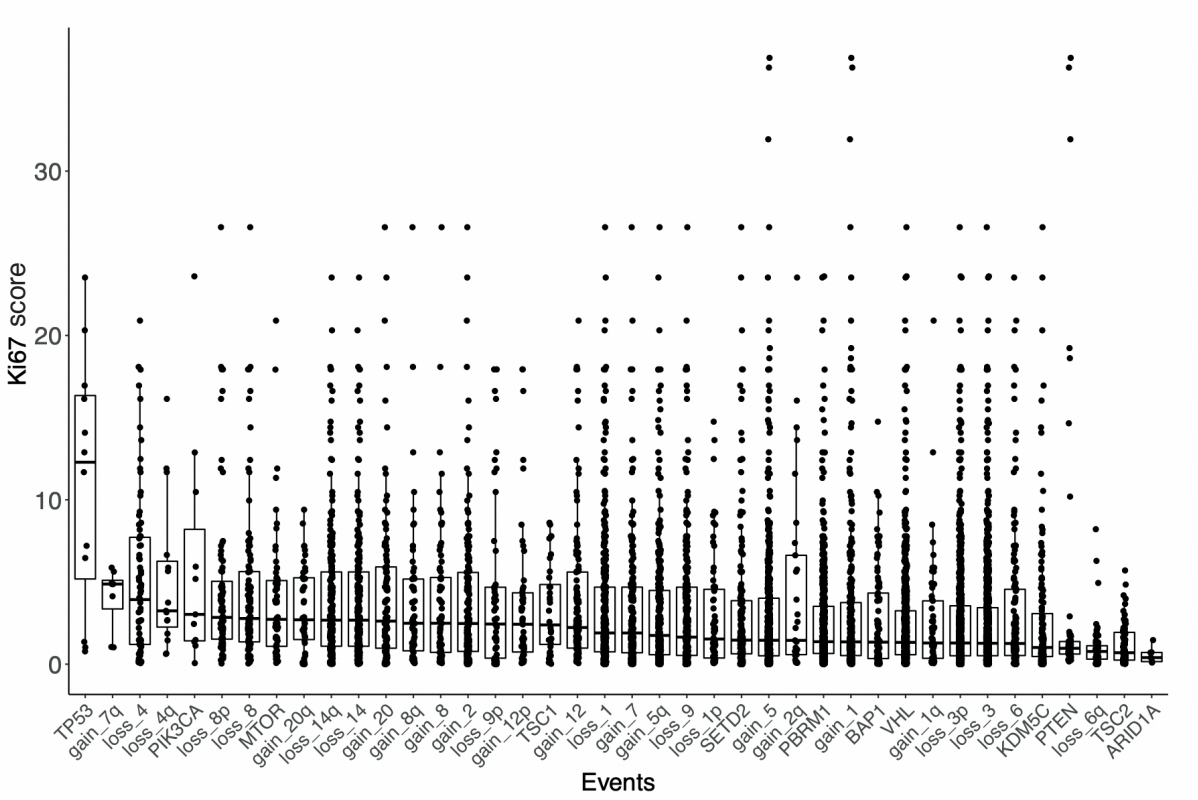
(a) Schematic figure for measuring the number of subclones in the area spanned by a mutational event within 2D tumour slice.

(b) The number of subclones in the area spanned by a mutational event of *PBRM1* or *BAP1* mutation against the size of that area (referred to as "N-A plot") in *in silico* tumours under Surface Growth (red) and Volume Growth (blue), respectively. Three sets of *in silico* tumours with different driver acquisition probabilities were shown. N = 100 tumours for each condition.

(c) Fitted power law exponent z, as in $N \sim A^z$, from the N-A plot in (b), for an interval of area from 1 to 100, as a function of driver acquisition probability.

(d) The number of subclones in the set of regions spanned by a genomic alteration (i.e., mutation or SCNA) against the number of regions. The complete set of tumours are further split for analysis in subsets according to the status of disease relapse. Fitted power law exponents z, as in $N \sim A^z$, are indicated in the figure.

a

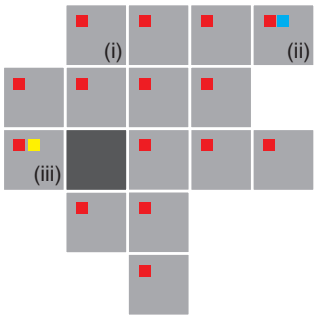


b

Level	Driver	Growth probability
1	*ARID1A, BAP1, KDM5C, MTOR, PBRM1, PIK3CA, PTEN, SETD2, TP53, TSC1, TSC2, VHL, gain 1q, loss 3p, loss 6q*	$p_{growth} = 0.25$
2	*gain 2q, gain 5q, gain 8q, gain 12p, loss 1p, loss 9p, loss 14q*	$p_{growth} = 0.25(1+s)$
3	*gain 7q, gain 20q, loss 4q, loss 8p*	$p_{growth} = 0.25(1+s)^2$

Supplemental Figure 1

a



$p_{growth} = 0.25 + \sum_k p_{growth\_k}$

BAP1
gain 20q
loss 4q

Tumour voxel (i): $p_{growth} = 0.25 + p_{growth\_BAP1}$

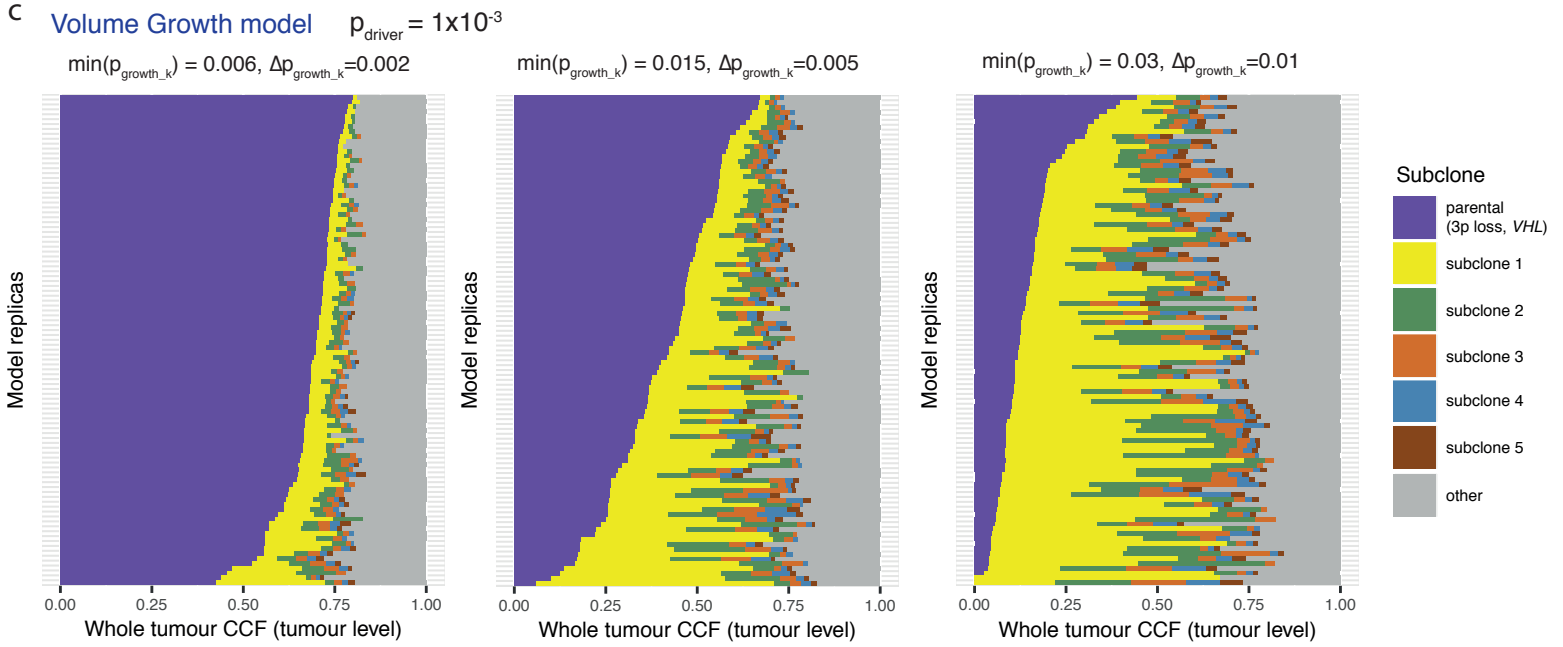
Tumour voxel (ii): $p_{growth} = 0.25 + p_{growth\_BAP1} + p_{growth\_loss\ 4q}$

Tumour voxel (iii): $p_{growth} = 0.25 + p_{growth\_BAP1} + p_{growth\_gain\ 20q}$

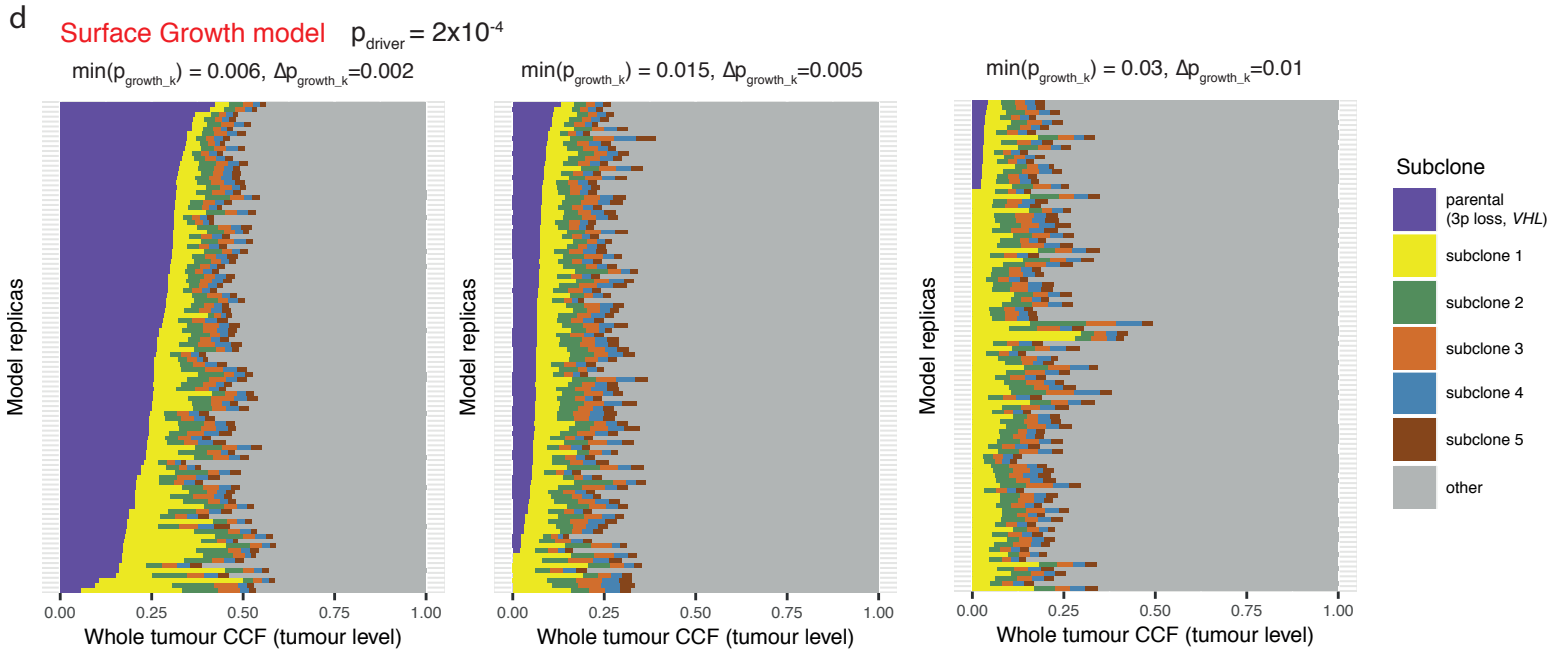
b

Driver k	Growth probability added by driver k (p_{growth_k})		
	$min(p_{growth\_k})$=0.006 Δp_{growth_k}=0.002	$min(p_{growth\_k})$=0.015 Δp_{growth_k}=0.005	$min(p_{growth\_k})$=0.03 Δp_{growth_k}=0.01
ARID1A	0.006	0.015	0.03
BAP1	0.022	0.055	0.11
KDM5C	0.014	0.035	0.07
MTOR	0.046	0.115	0.23
PBRM1	0.024	0.06	0.12
PIK3CA	0.05	0.125	0.25
PTEN	0.012	0.03	0.06
SETD2	0.028	0.07	0.14
TP53	0.056	0.14	0.28
TSC1	0.034	0.085	0.17
TSC2	0.008	0.02	0.04
VHL	n/a	n/a	n/a
gain 1q	0.018	0.045	0.09
gain 2q	0.026	0.065	0.13
gain 5q	0.032	0.08	0.16
gain 7q	0.054	0.135	0.27
gain 8q	0.04	0.1	0.2
gain 12p	0.036	0.09	0.18
gain 20q	0.044	0.11	0.22
loss 1p	0.03	0.075	0.15
loss 3p	n/a	n/a	n/a
loss 4q	0.052	0.13	0.26
loss 6q	0.01	0.025	0.05
loss 8p	0.048	0.12	0.24
loss 9p	0.038	0.095	0.19
loss 14q	0.042	0.105	0.21

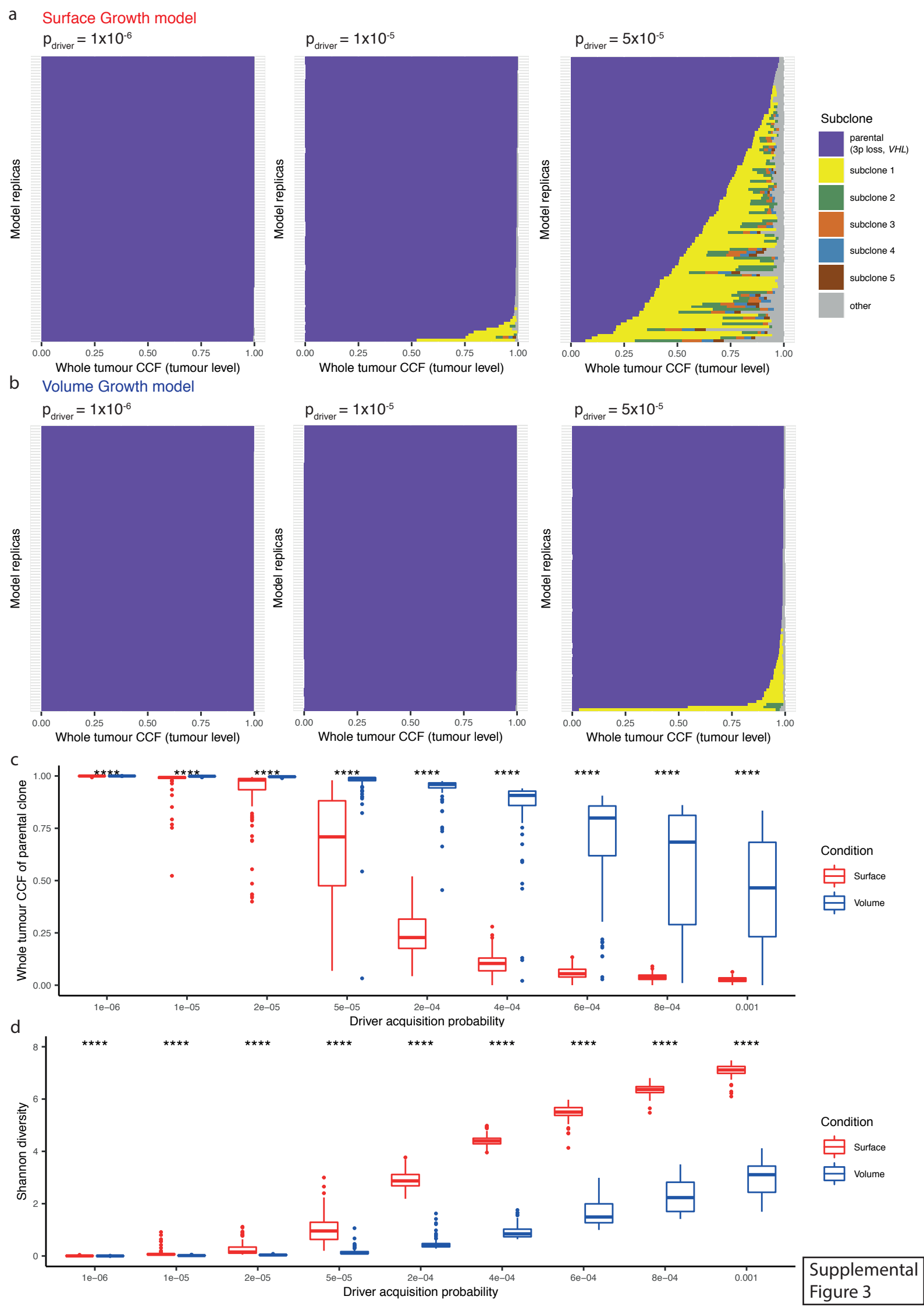
c



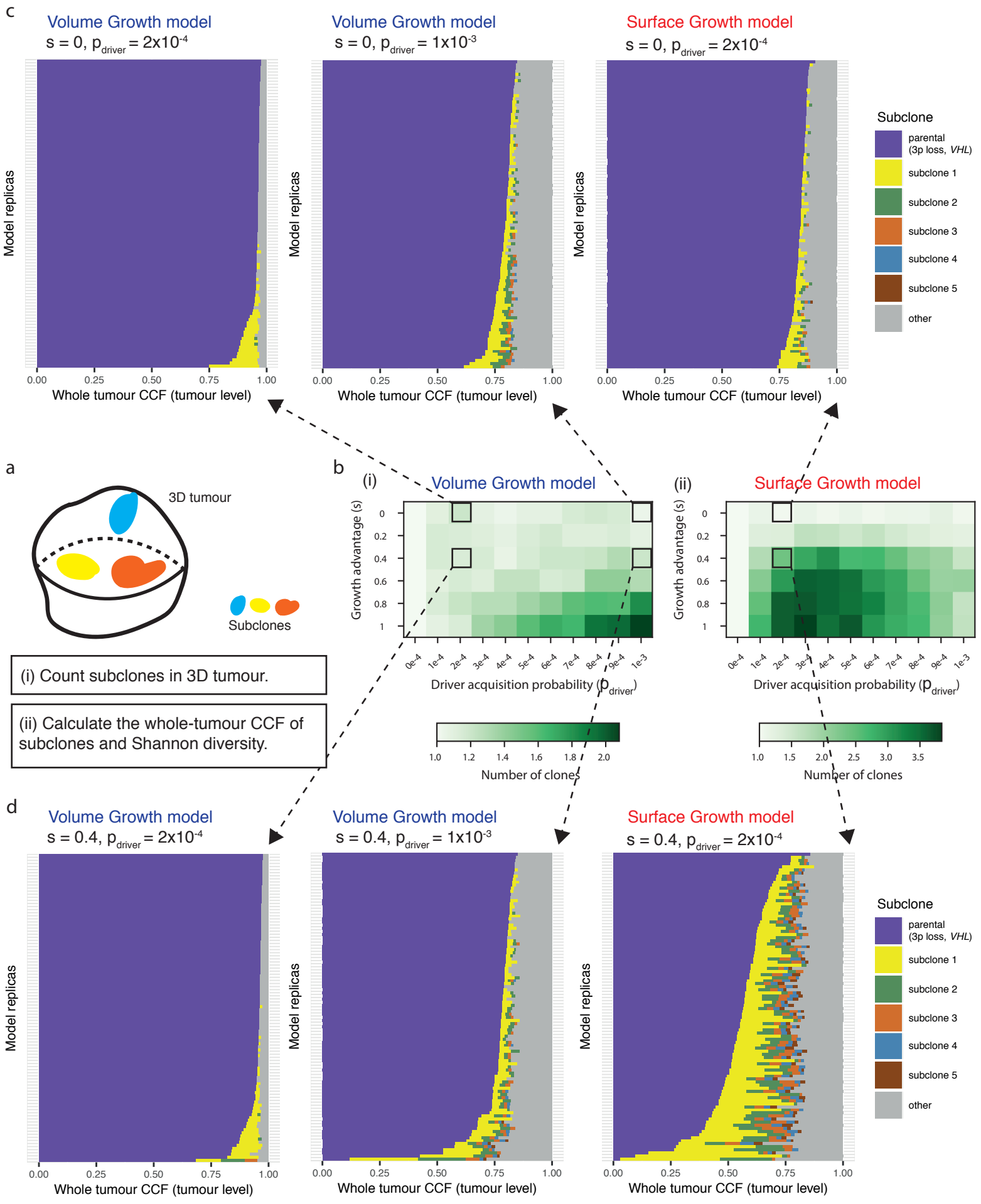
d



Supplemental Figure 2



Supplemental Figure 3

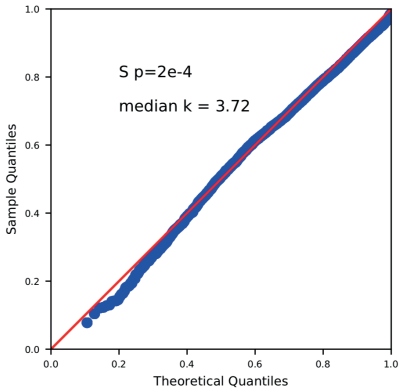


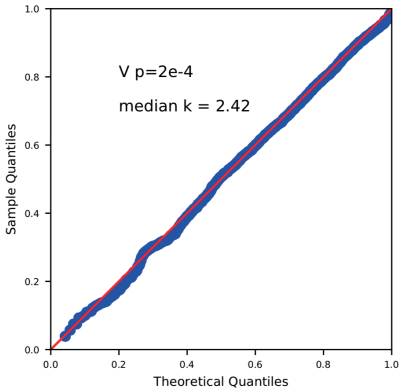
Supplemental Figure 4

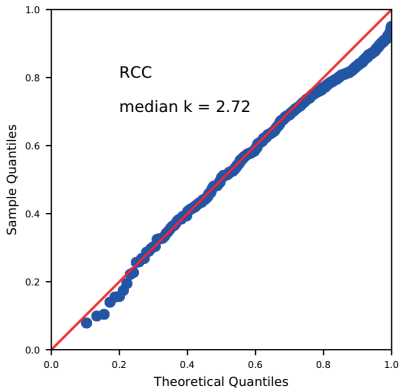
Surface Growth model

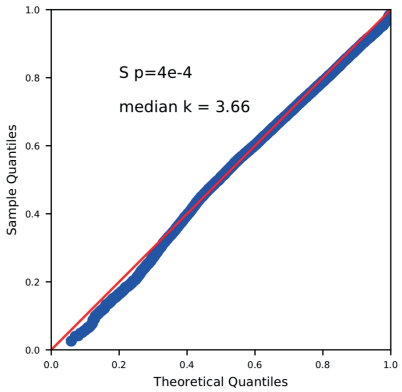
Volume Growth model

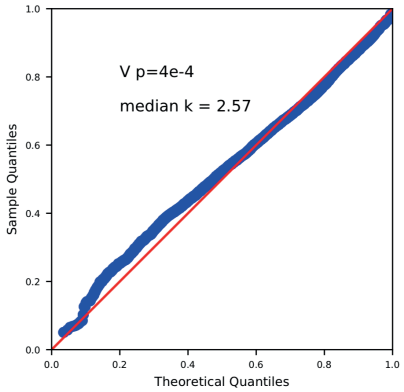
RCC

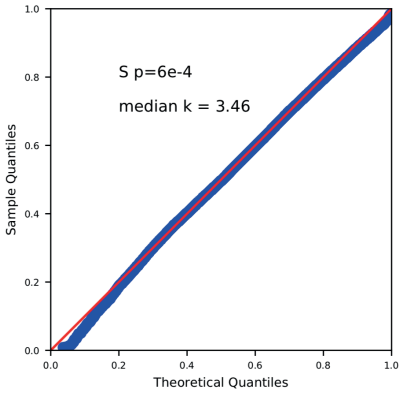


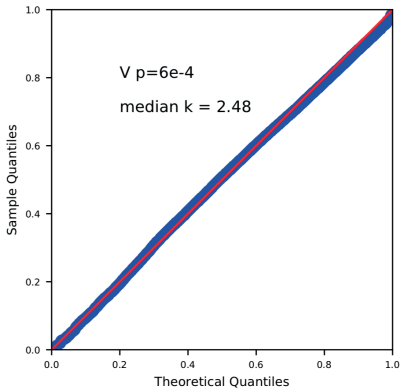


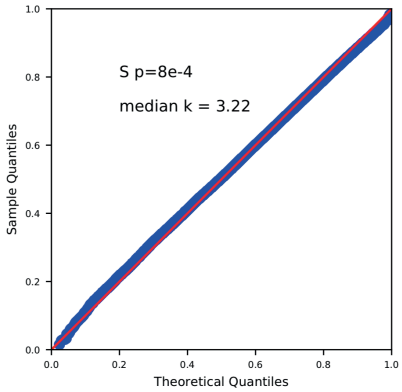


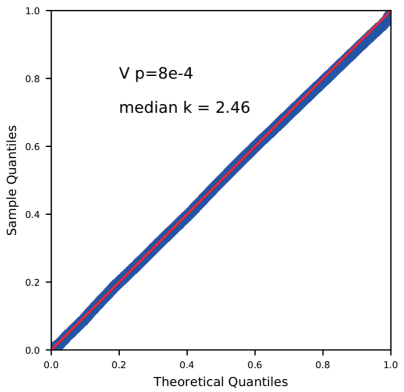


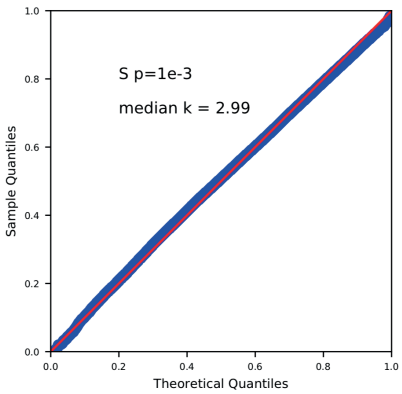


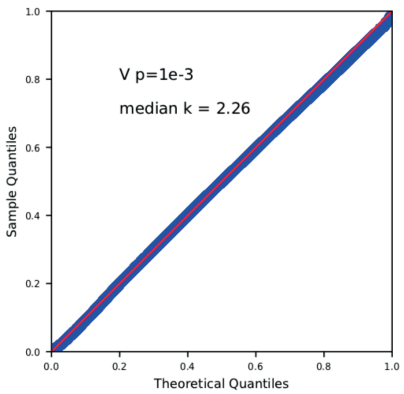




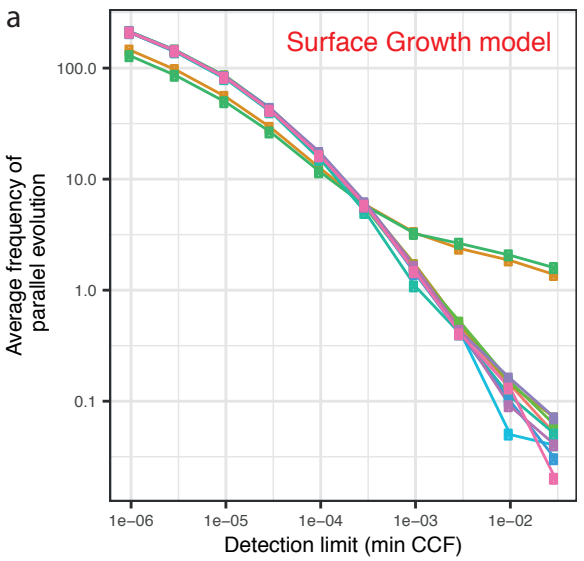


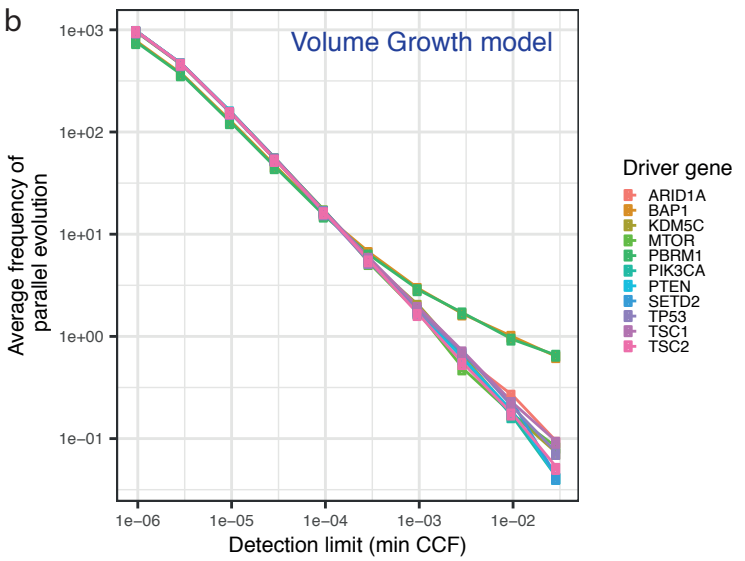


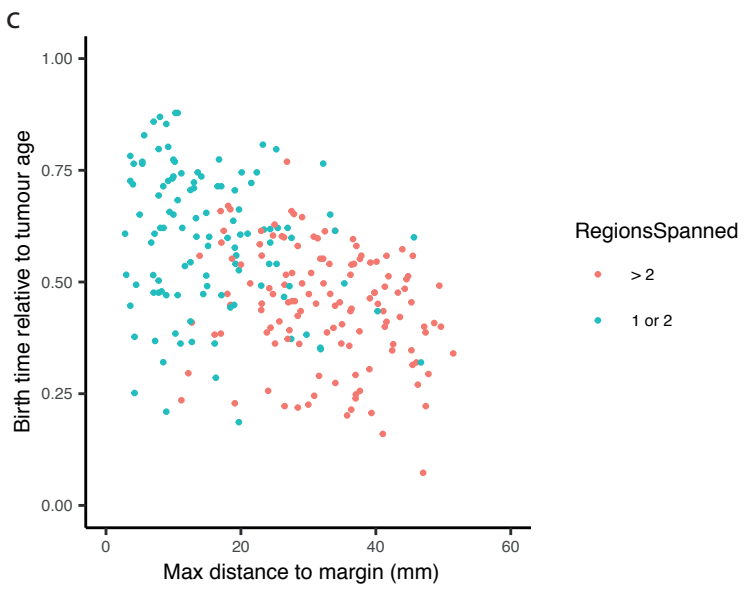


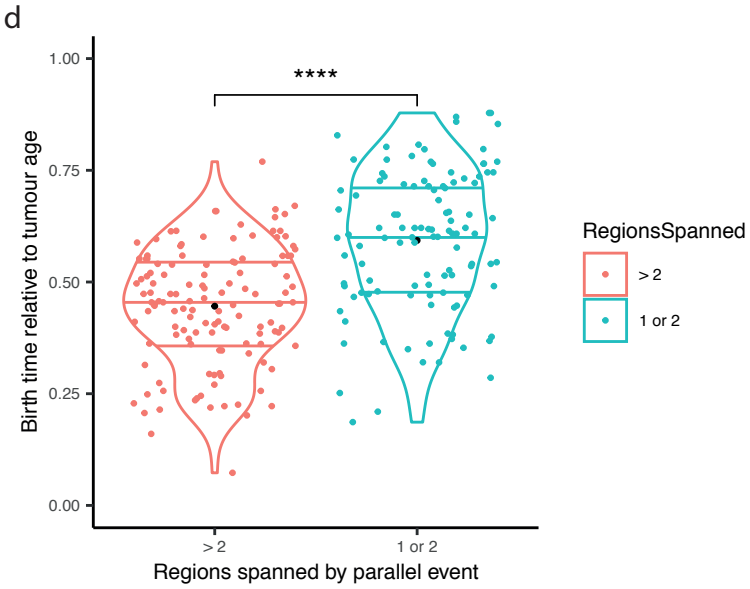


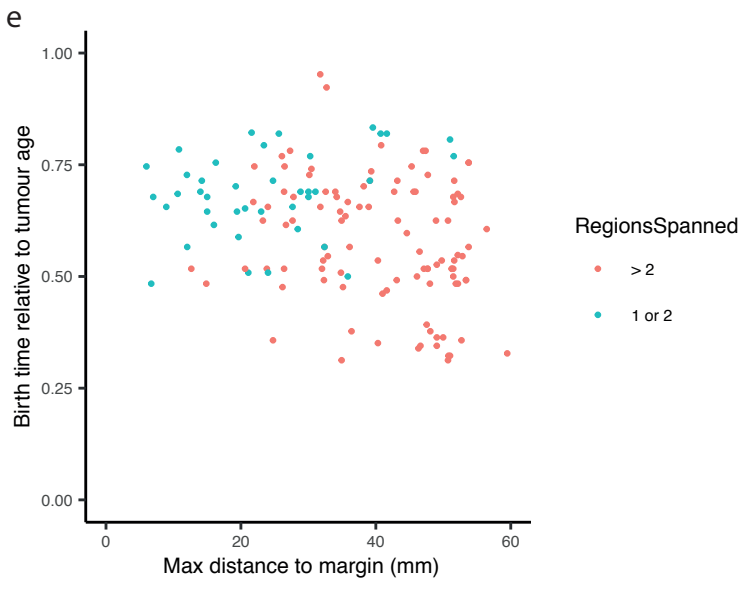
Supplemental Figure 5

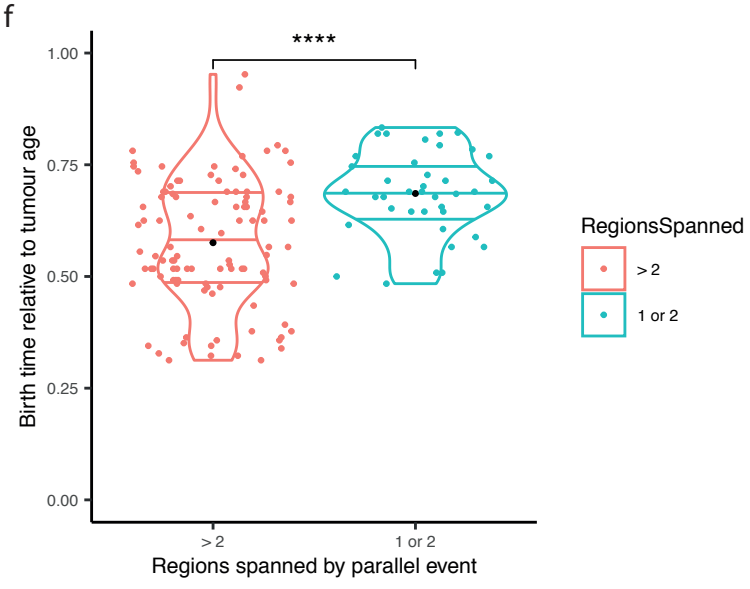




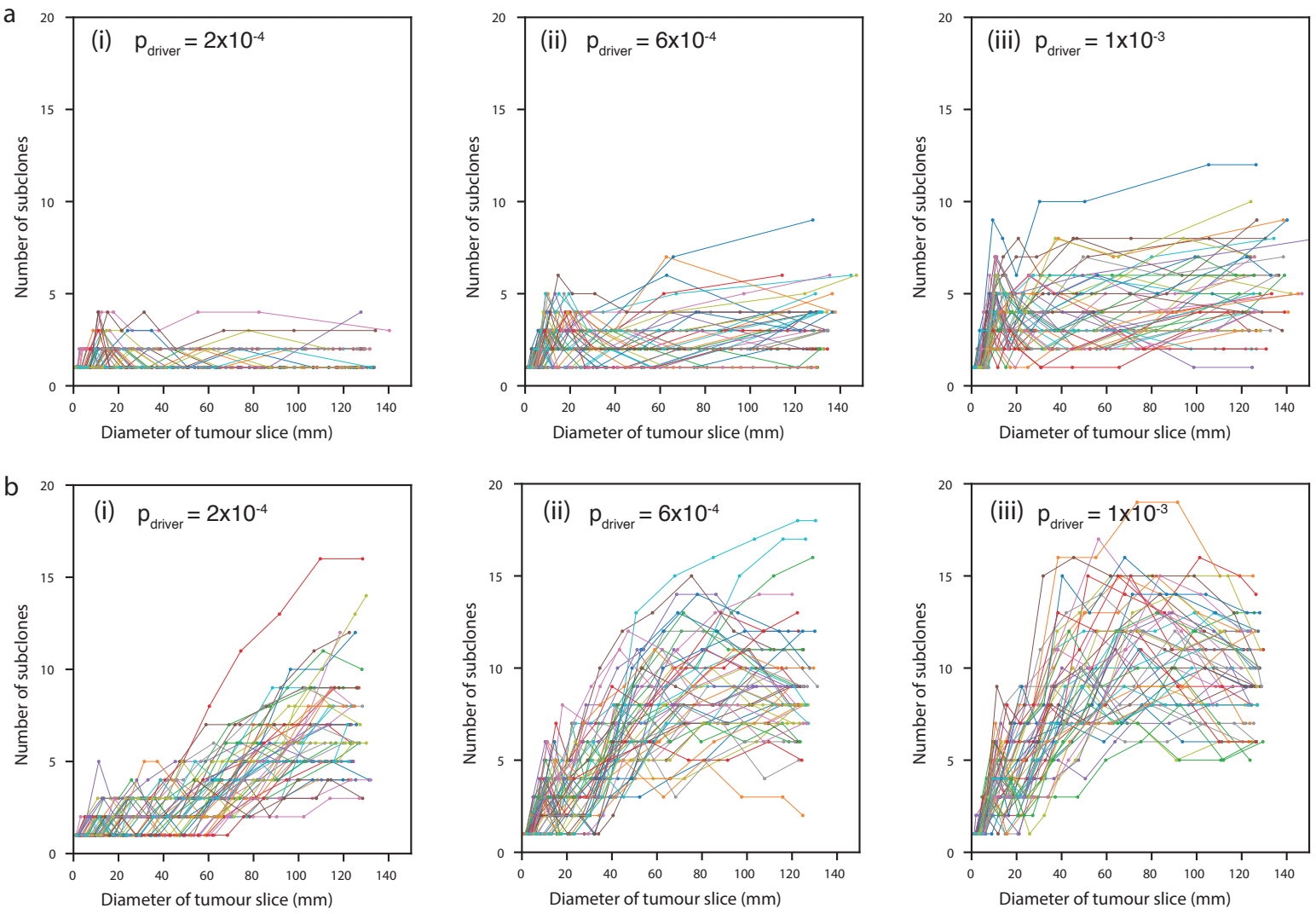




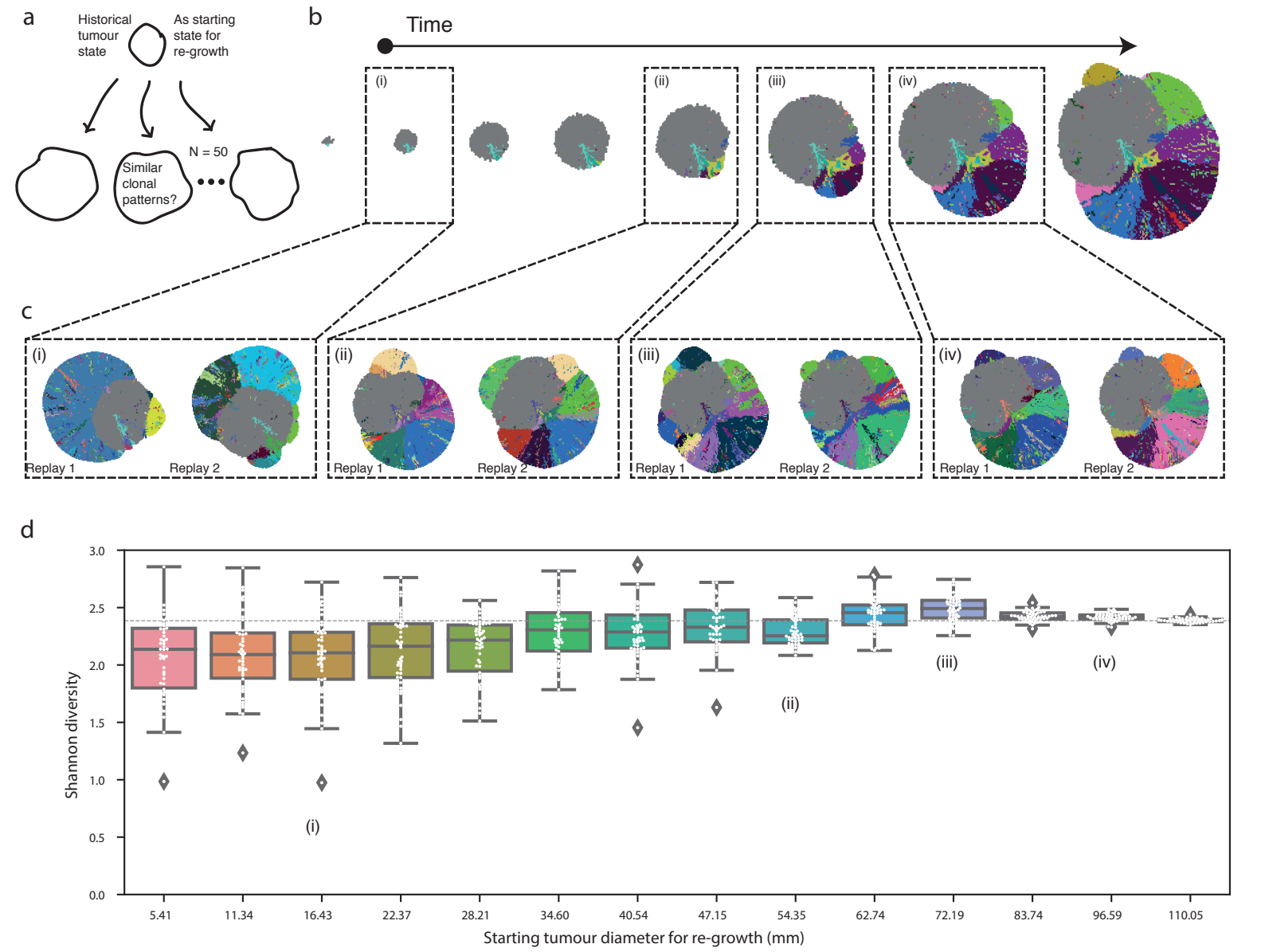


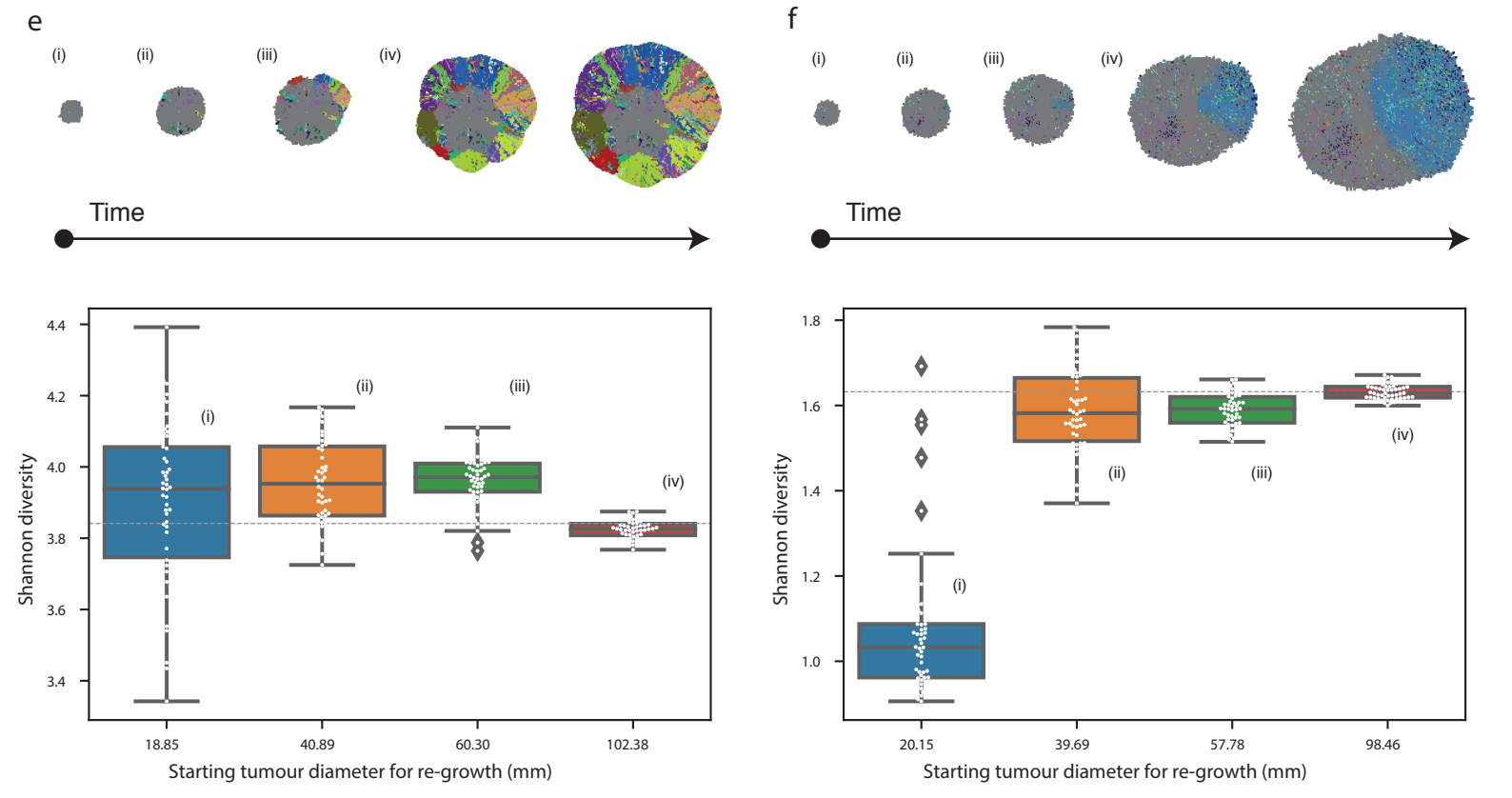


Supplemental Figure 6

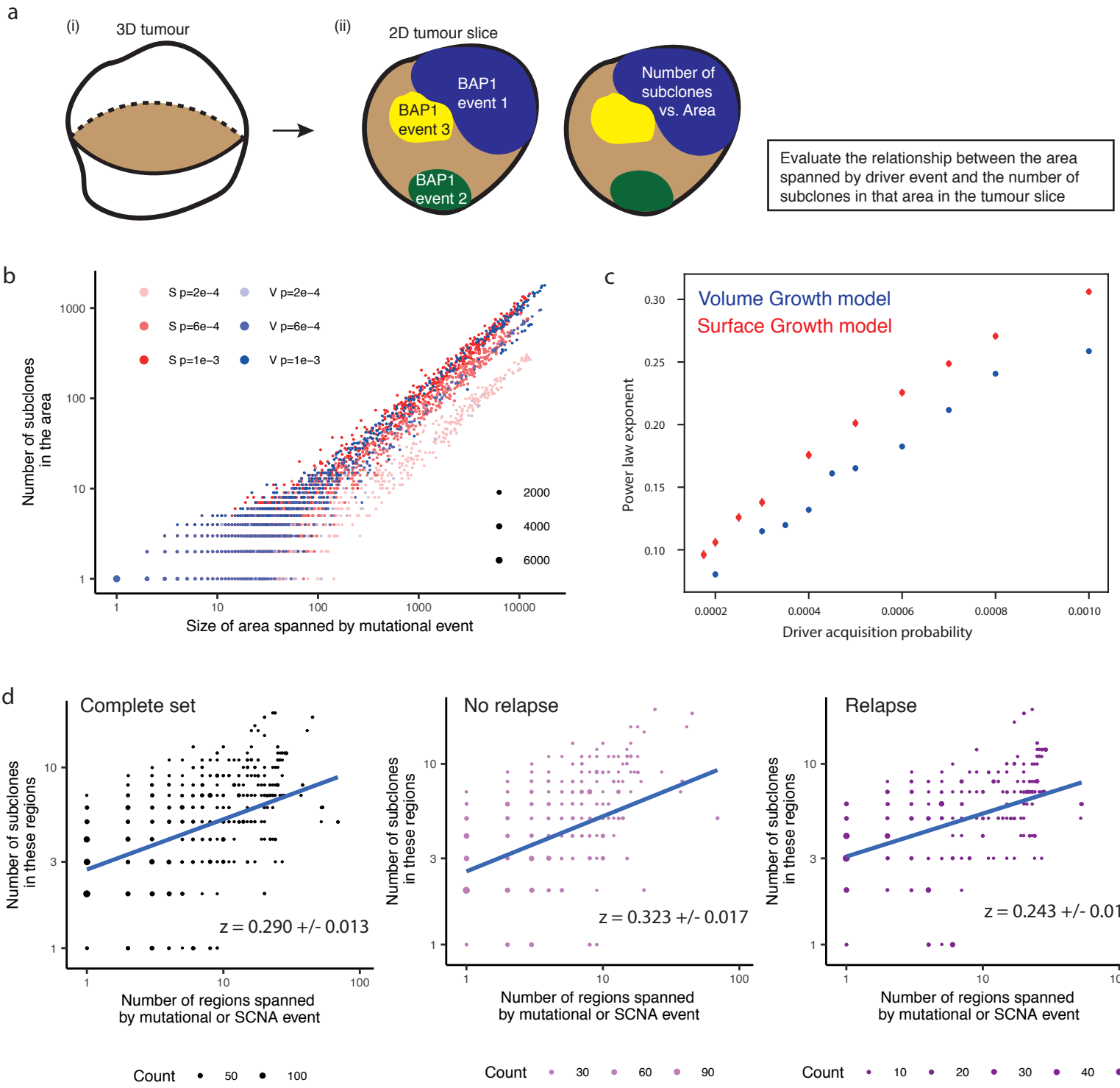


Supplemental Figure 7





Supplemental Figure 8



Supplemental Figure 9