Evolutionary and spatial dynamics underlying intra-tumor heterogeneity

Maya A. Lewinsohn

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:
Trevor Bedford, Chair
Alison Feder
Peter Nelson

Program Authorized to Offer Degree:
Genome Sciences

University of Washington

Abstract

Evolutionary and spatial dynamics underlying intra-tumor heterogeneity

Maya A. Lewinsohn

Chair of the Supervisory Committee:
Trevor Bedford
Department of Epidemiology

Intra-tumor heterogeneity (ITH) is a pervasive feature of solid tumors that complicates clinical prognosis and management. Yet, the evolutionary processes that explain ITH remain largely unresolved. Here, we demonstrate that phylodynamic models can provide valuable insight into these dynamics from multi-region sequencing data of resected tumors. We introduce a novel state-dependent birth-death and sequence evolution model that detects signals of boundary-driven growth in hepatocellular carcinoma. We further adapt a continuous trait evolution model to reconstruct joint spatial and genetic diversification in prostate cancer. Finally, we tackle the fundamental question of heritable versus non-heritable determinants of transcriptomic variability in prostate cancer and find unique genetic contributions to key cancer-related processes.

ACKNOWLEDGMENTS

This dissertation is the culmination of four years of work and learning. I am extremely grateful to my supervisor Trevor Bedford for his support and guidance throughout this time, and for giving me the opportunity to pursue a creative and inter-disciplinary project.

I am additionally very appreciative of Alison Feder for her guidance and contributions to this research, and in other parts of my PhD journey, especially for including me in the cancer evolution journal club. This became a key part of my learning and helped me stay scientifically engaged during the isolated periods of the pandemic. I would also like to thank the other members of my committee, Peter Nelson, Gavin Ha, Kelley Harris, and Ivana Bozic for their thoughtful feedback and time, and Mary-Claire King along with the MSTP leadership for providing encouragement and guidance in both my medical school and graduate school years.

This work would not have been possible without Nicola Müller for agreeing to work on a cancer project and his work in implementing SDevo. The collaboration with Colm Morrissey, and the University of Washington Genitourinary Research group was also essential for the success of these projects and included the extra efforts of Lori Kollath and Andrew Ustach in creating a non-standard study design. I would like to extend further thanks to Martine Roudier for her enthusiasm and help in pathological annotations of the prostate cancer specimen, Ilsa Coleman for her advice in sequencing and bioinfomatics, the Fred Hutch genomics team, and the Fred Hutch scientific computing team. I also depended greatly on the administrative support of

Carol Wallace, Lindsey Mwoga, Brian Giebel, and Sara Carlson .

Next, I would like to thank the members of the Bedford group, and my UW MSTP and Genome Sciences peers for creating a welcoming community and their generosity with sharing knowledge and advice.

Finally, I would like to thank my husband, Michael, for being my greatest supporter and teammate in this journey, along with my family and friends.

TABLE OF CONTENTS

Chapter

INTRODUCTION

1.1 Clinical impacts of intra-tumor heterogeneity

Intra-tumor heterogeneity (ITH) is a well-documented feature of cancer that plays an important role in cancer progression, metastasis, and resistance to chemotherapy (1–11). Intra-tumor heterogeneity can be seen both within one patient over time (temporal ITH) and across different locations (spatial ITH). Observations of ITH are not new, instead, genetic and phenotypic variation, including in morphology, karyotype, antigen expression, enzyme activity, and drug susceptibility, were noted decades ago (12–17). Variation in drug susceptibility is particularly notable, as although therapy is often aimed at targeting a large part of the population that may share a driver mutation or another key druggable characteristic (i.e. high growth rate, loss of DNA repair mechanisms), it rarely targets all cells equally. Even if single therapy may kill most of the detectable tumor cells, it can too often be much harder to cure an entire tumor due to variability in drug-susceptibility. Accumulating evidence also suggests that only a subset of the tumor population drives metastasis (7, 18), a main cause of cancer fatalities (19).

1.2 Evolutionary forces that contribute to ITH

Intra-tumor heterogeneity is a product of evolution within the tumor cellular population (20–22). Classically, tumors are thought to form due to the accumulation of a critical number of genetic mutations that allow cells to escape regulations on growth normally imposed on differentiated cells (23–25). As a tumor population expands, cells accumulate genetic variation, including single nucleotide variants (SNVs), insertions and deletions (INDELS), somatic

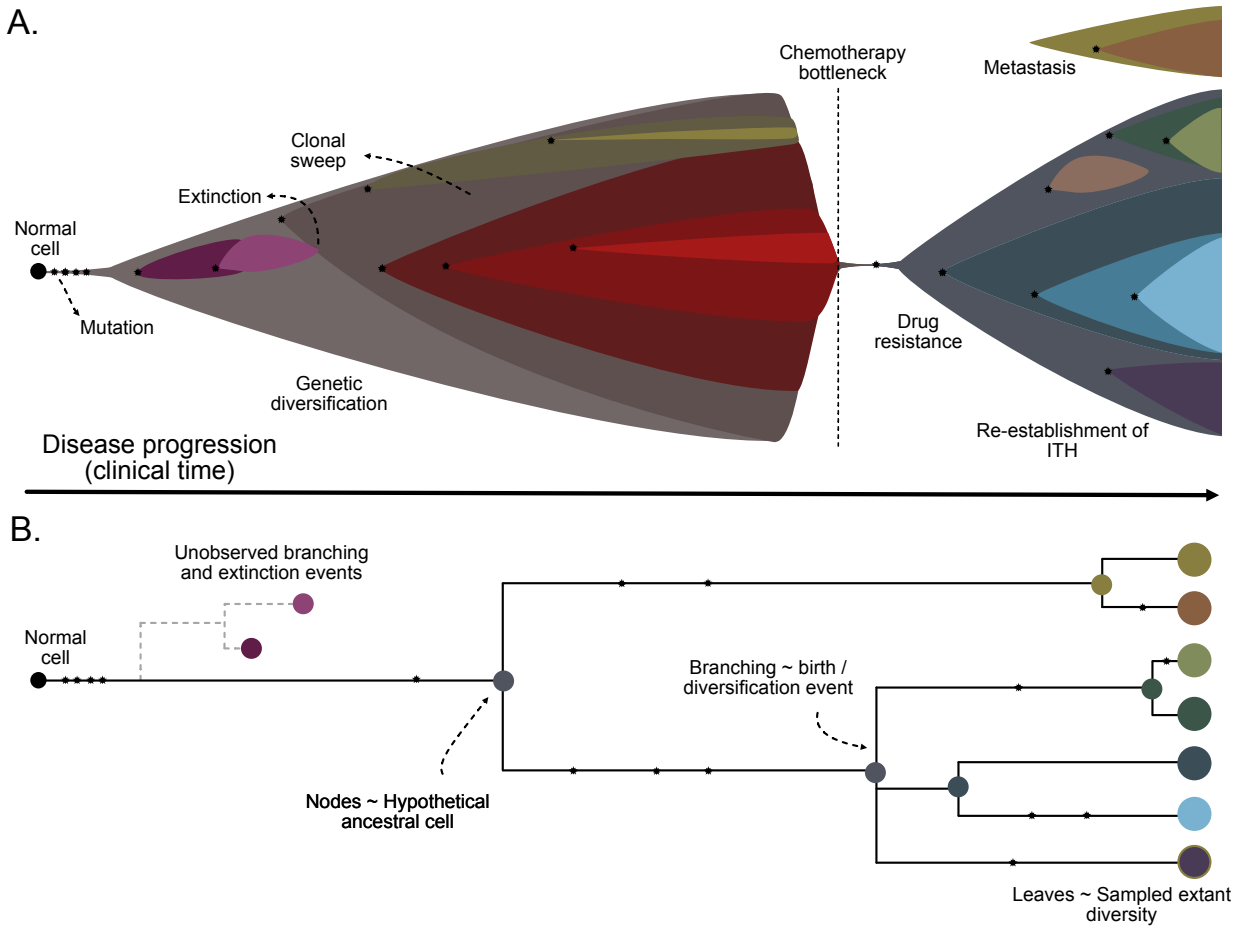

Figure 1.1: Evolutionary dynamics and reconstruction of intra-tumor heterogeneity. A. ITH develops through the gain of genetic aberrations (asterisks) and subsequent population dynamics of genetic subclones. Inversely, genetic diversity may be lost through stochastic extinction, clonal sweeps, or bottlenecks such as metastasis or chemotherapy. **B.** Phylogenetics can reconstruct a tumor's history from sampled extant genetic diversity (usually punch biopsies in multi-region sequencing studies). Nodes represent common ancestors of surviving lineages. Phylodynamics extends phylogenetic methods by incorporating models of sequence evolution and population dynamics, for example, unobserved branching and extinction events.

copy number alterations (SCNA), structural variation (SV), and whole-genome duplication (WGD) due to genetic instability and rapid division. These genetic aberrations initiate genetic subpopulations, or *subclones*, which contribute to *genetic* ITH as they expand in the population (26) (Figure 1.1). The forces that shape tumor evolution are largely overlapping with those classically investigated in species evolution and ecology, namely, replication, genetic (or otherwise heritable) variation, drift, selection, and environmental pressures (26). Within a tumor population fitness advantages conferred by driver mutations, immune evasion (27) or environment-specific adaptations may drive clonal success or sweeps. However, it is widely still debated to what extent selective forces act on tumor evolution and some studies have suggested that many tumors may evolve neutrally during primary growth (28, 29). Other examples of evolutionary-driven processes include chemotherapy bottlenecks, in which only a small, resistant subpopulation survives, and metastases, which are frequently seeded by a subset of the primary tumor population that have gained the ability to invade and survive in a new environment(18).

1.3 Spatial factors that influence subclonal success and ITH

Although genetic drivers are often linked to clonal success, solid tumors grow in physical environments and spatial factors may play a role in the evolutionary fate of tumor cells and the progression of the disease. These factors include variable access to resources in the microenvironment (18), immune invasion (30), microbiota (31) and anatomical constraints (32). Many evolutionary models have been proposed to explain observed spatial patterns of intra-tumor heterogeneity (33–37). For example, in branching evolution, there is continuous selection for driver mutations or other fitness advantages in the tumor cell population, but these fitness advantages are distributed across multiple subclones in the population in a way that leads to simultaneous subclonal expansions in different parts of the tumor (33, 38). Alternatively, the "big bang" model would propose most genetic diversity is established early

in tumor development and then expands neutrally (28). If genetic subclones are expanding under neutral evolution dynamics, it is also likely that spatial structure plays a critical role in determining cellular fitness. For example, boundary-driven growth, as described in Chapter 2, would explain subclonal distribution resulting from cells simply being in the right place at the right time (i.e. on the edge of the tumor). Spatial restrictions could further prevent clonal sweeps by restricting cancer cell competition to proximal clones (39) and therefore enhance diversity. On the other hand, increased cell turnover or even short-range migration could contribute to increased global sweeps and selection of cancer subclones even with spatial restrictions on cell growth (40).

1.4 The interplay between genetic and phenotypic ITH

As a whole, tumors progress by gaining attributes that are classically described as *hallmarks of cancer* (25). However, often cells do not necessarily uniformly gain these attributes, instead, these properties are gained through a myriad of different pathways and in different cell lineages (41, 42, 18). One unique attribute of cancer as an evolutionary system is that individuals (cells) can exhibit a high degree of phenotypic plasticity. During tumorigenesis, cancer cells often escape differentiation that normally constrains somatic cells to a particular morphology, behavior, and anatomical location. Cancer cells further often revert to undifferentiated stem cell-like phenotypes, enabling them to exploit early developmental pathways and access a wide range of cell states (43). Due to this property, it has been suggested that a large proportion of transcriptomic intra-tumor variability is not liked to heritable or genetic diversity (41, 44).

1.5 Phylogenetic and phylodynamics approaches as tools to study ITH

Phylogenies represent the evolutionary relationships between individuals in a population. Bayesian phylodynamics extend phylogenetic approaches by jointly modeling how popula-

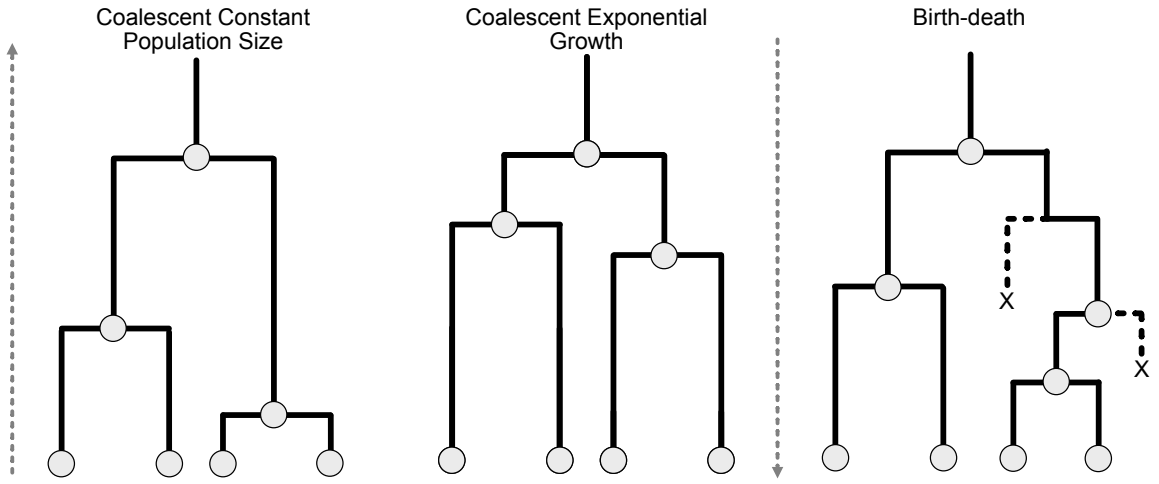

Figure 1.2: Examples of population models in phylodyamics. Coalescent models are originally based on an analysis of the Wright-Fisher model which assumes constant population size and uses backward inference (arrow). The expected coalescent times change if instead exponential population growth is assumed. Alternatively, birth-death models are a class of stochastic branching processes that include both observed and unobserved birth, death, and sampling events that give rise to the extant, sampled population.

tion and sequence evolution give rise to observed genetic diversity. Fundamentally, these approaches exploit well-established observations that different population processes lead to different trees, and incorporate these expectations into inference. For example, the Wright-Fisher model (45, 46), a classic population genetics model, assumes a constant population size and generation time, and random reproduction (no selection). The coalescent model (47) defines expectations derived from this model: for a given set of sampled extant individuals, when would they coalesce (when was their most recent common ancestor)? These expectations change if instead, the population is growing exponentially (Figure 1.2). Coalescent models are popular for their simplicity and usefulness in modeling deterministic population dynamics. Alternatively, birth-death models are another popular class of branching models that incorporate stochastic branching processes and, as the name implies, explain population diversity as a product of birth, death, and sampling events (48–51).

One of the strengths of phylodynamic methods is the ability to extend or vary the underlying models to incorporate additional information or complexity of a biological system. For example, both continuous and discrete traits of species or individuals can be inferred along a tree via trait evolution models (Figure 1.3). A core subset of discrete trait evolution models includes sequence evolution models which describe both the rate at which molecular sequences change (i.e. molecular clock rate, (52)) and how the sequence changes, including nucleotide or amino acid substitution models (53), and site-specific differences. However, other traits can be inferred in a similar manner. Importantly, these traits can in turn be tied to other evolutionary parameters, such as the rate of diversification and extinction in a birth-death model (54), or impose population structure in a coalescent model (55).

1.5.1 Phylodynamics in the study of cancer evolution

Phylodynamic models have been a broadly used and powerful tool in other fields, including virology (56–59) and ecology (60, 61), to elucidate underlying population processes from

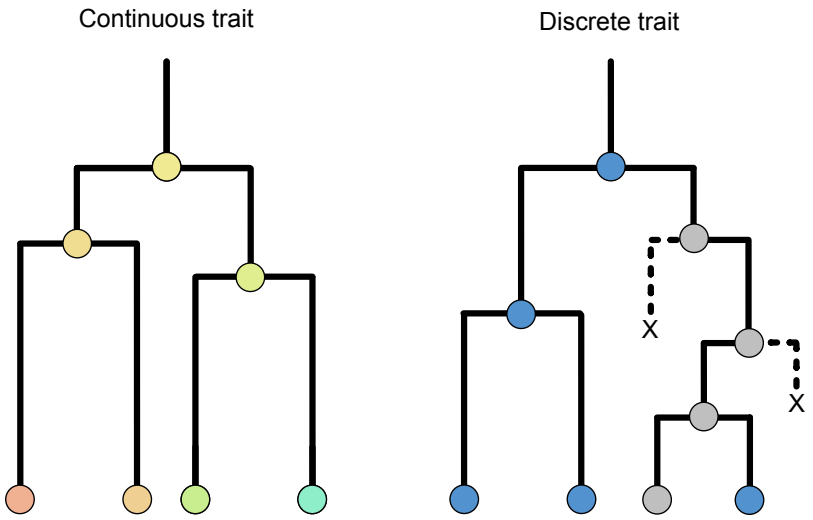

Figure 1.3: Trait evolution models. Trait evolution models are used in phylogenetics to describe how character states change along a tree. Continuous trait evolution models imply that traits pass through intermediates to reach the observed states. Discrete trait evolution models assume a rapid or immediate transition to a finite number of states. Trait evolution models are combined with other tree processes, for example, an exponential coalescent model (left) or a birth-death branching process (right).

genetic data. However, the high cost of sequencing, the limited resolution of genetic data, and the lack of cancer-specific process models have historically hindered analogous applications to cancer evolution.

Despite these past limitations, spatially resolved datasets and maps of intra-tumor genetic diversity are increasingly available (62, 32). Phylogenies are now routinely used to describe genetic ITH (63), reconstruct the evolutionary relationships between sampled tumor populations, and infer origin of metastases (64–66, 9) (Figure 1.1B). However, phylodynamic methods have still only been sparsely introduced to model evolutionary dynamics that could explain or interpret these tree structures (9) and most existing phylodynamic methods have not been adapted or applied to cancer processes. There is enormous potential to extend insights gained from existing and emerging phylogenetic studies through phylodynamic approaches.

1.6 About this dissertation

In this dissertation, I describe a diverse set of phylodynamics and statistical approaches to study the evolutionary and spatial forces that contribute intra-tumor heterogeneity. First, in Chapter 2, I introduce a novel phylodynamic model that was developed in collaboration with Dr. Nicola Müller and Dr. Alison Feder which adapts a state-dependent birth-death model to cancer and detects boundary-driven growth in hepatocelluar carcinoma. In Chapter 3, I investigate the genetic ITH and spatiotemporal dynamics of a primary prostate by implementing a unique, spatially-resolved multi-region sequencing study design in collaboration with the University of Washington Genitourinary (GU) research group. To do so, I exploit a novel application of continuous trait evolution model to reveal the early spatial and genetic diversification of a primary prostate tumor. Finally, in Chapter 4, I use statistical approaches to quantify genetic and spatial contributions to transcriptomic intra-tumor heterogeneity in the same specimen.

Chapter

STATE-DEPENDENT EVOLUTIONARY MODELS REVEAL MODES OF SOLID TUMOR GROWTH

2.0.1 Introduction

Tumors develop and progress via an evolutionary and ecological process wherein cellular subpopulations expand and diversify. Over the course of tumor development, tumor cells acquire genetic mutations and new phenotypes that potentially help them compete for resources and adapt for success in their microenvironment. Understanding this process is critical to predicting clinically significant events such as if, how, and when cells metastasize or develop resistance to therapy.

Although tumor cell growth and success are often attributed to genetic and epigenetic aberrations, an additional important determinant of cell growth is physical location within the tumor. Position governs access to oxygen, nutrients, pro-growth signaling from the stroma, pH, cell-cell interactions, and degree of immune and drug exposure, all of which can affect cellular proliferation (67–73). Taken together, these effects may combine to create an environment in which cells on the boundary of a tumor have higher growth rates compared to those in the center (i.e., "boundary-driven growth").

Cancer biologists have long been interested in boundary-driven growth because it changes the evolutionary processes and genetic signatures of tumor progression. The evolutionary impact of boundary-driven growth has been explored via evolutionary theory (74, 75), microbial experiments (76–78), and decades of work involving cancer computational and mathematical models (67–70, 79, 28, 80–82, 36). Such investigations have revealed that boundary-driven growth blunts the efficacy of natural selection in selecting for beneficial (i.e, driver) mutations,

and purging slower growing (but potentially drug resistant) lineages (83). Boundary-driven growth should also enhance the effectiveness of adaptive therapy (39, 84) and cell-cell competition in the tumor interior. Further, such growth patterns should distort our expectations for the neutral variant allele frequency spectrum (85), which has been used as a null model for identifying signatures of natural selection (86), and it has been qualitatively suggested in tumor simulation studies that boundary-driven growth could be misinterpreted as selection on tumor trees (82). Therefore, establishing and incorporating these null expectations and models for boundary-driven tumor growth is essential in the context of the increasing interest in applying evolutionary theory to clinical disease, for example, in designing adaptive therapy (87), identifying driver events (88, 89), or estimating timings of metastases (90, 81).

An extensive history of clinical and experimental observations supports the importance of boundary-driven growth in tumor populations. These observations date back to the pioneering work of Thomlinson & Gray which first identified necrotic structures with surrounding boundaries of growing cells from histological sections (91) and subsequent cell staining approaches that found markers of cell division cluster preferentially on the tumor periphery (92, 93). Similar patterns have been noted in cultured tumor spheroids (24, 94) and organoids (95, 96). Since then, analysis of both clinical samples - via immunohistochemistry (97, 98), spatial transcriptomics (99–101) and genetic analysis (34, 102) - and experimental systems such as fluorescently-tracked xenografts (103, 73, 104, 105), have further supported spatial heterogeneity and preferential expansion on the tumor periphery in some tumors.

However, more recent studies have hinted at more complex modes of clinical tumor growth. Househam et al. (2022) (102) found that many colorectal tumors showed genetic patterns not consistent with boundary-driven growth, and a recent genetic analysis of renal cell carcinomas found the most recent common ancestors of metastatic lineages in the resected tumor interiors as opposed to the tumor boundaries (106). Additionally, experimental evidence suggests that although center-bound cells may experience oxygen and nutrient de-

privation, hypoxia-related signaling can be linked to stem-cell like tumor phenotypes with heightened survival and chemotherapy resistance (107, 108). These observations highlight that higher proliferation on the tumor edge is not necessarily synonymous with long-term lineage survival and progression (109).

A primary challenge in reconciling these conflicting observations is that clinical sequencing often only captures a limited snapshot of tumor diversity and growth. However, this sampled tumor diversity still offers a window into past population dynamics via phylogenetic and phylodynamic tools. *Phylogenetic* approaches, which reconstruct how cells within a tumor are related, have already proven useful in interrogating cancer evolution – for example, in determining the relative ordering of driver mutations (38, 110, 111), detecting parallel evolution of gene hits within a tumor (112, 113), and resolving whether metastases emerge early or late in tumor development (114, 115). In contrast, *phylodynamic* methods, which link shapes of phylogenetic trees to underlying population dynamics, have only rarely been used in cancer genomics (116), despite widespread application in other fields (117, 118).

Although phylodynamic approaches have high potential impact in cancer clinical settings, they are generally not adapted to study tumor biology or incorporate the complexities of cancer's spatial growth. To bridge this gap, we set out to develop a phylodynamic model suited for detecting boundary-driven growth in tumors. First, we quantify characteristic branching and genetic patterns in tumor trees simulated under boundary-driven growth and demonstrate that these patterns correspond to cellular lineages spending different amounts of time on the faster growing tumor edge versus in the tumor center. To fully exploit these patterns for inference, we develop a novel phylodynamic tool based on the multi-type birth-death process (49, 119, 120), in which cells have different birth and death rates on the tumor edge and center, and lineages can transition between states as the tumor grows. Crucially, we introduce an extension that links cell birth and mutation, and therefore incorporates rates of sequence evolution that depend on each cell lineage's inferred history

of spatial locations (i.e. spatial states). We provide this state-dependent evolution model (SDevo) as a package in the popular open source Bayesian software BEAST2 (121). We show that SDevo substantially improves our ability to infer boundary-driven growth dynamics in simulated tumors compared to non-cancer multi-type birth-death models, and validate this approach across a broad array of biological and sampling conditions, including those encompassing selection for driver mutations, 3-dimensional growth, and clinical sampling strategies. Finally, we apply SDevo to spatially-resolved multi-region sequencing data from hepatocellular carcinomas (34) and estimate that cells in the tumor boundary may have birth rates up to 3-6 times faster than those in the interior. More broadly, SDevo is a general tool for quantifying growth processes linked to any discrete state, and future investigations will expand beyond boundary-driven growth.

2.1 Results

2.1.1 Boundary-driven growth creates distinct tree structures

In order to characterize signatures of boundary-driven growth in tumor trees, we simulate spatially-constrained growth via a cellular agent-based model in a 2D lattice, following a rich literature of studying cancer dynamics via Eden models (122, 123, 79, 82). Simulated tumors grow from single cells over discrete time steps and gain mutations at cell division. Under spatially-constrained *boundary-driven growth*, a cell can only divide if there is an empty lattice spot in its Moore (8-cell) neighborhood, effectively tying its fitness to neighborhood density (Figure 2.10). Therefore, extant lineages closer to the tumor periphery have progressively higher mean birth rates than those in the center (Figure 2.1A). For comparison, we simulated a non-spatially constrained *unrestricted growth* model (Figure 2.1D), in which all cells can divide regardless of density and push their neighbors to create space.

We first investigated how such growth processes affect the shape and structure of cancer phylogenetic trees to identify detectable tree signals of boundary-driven growth. We consid-

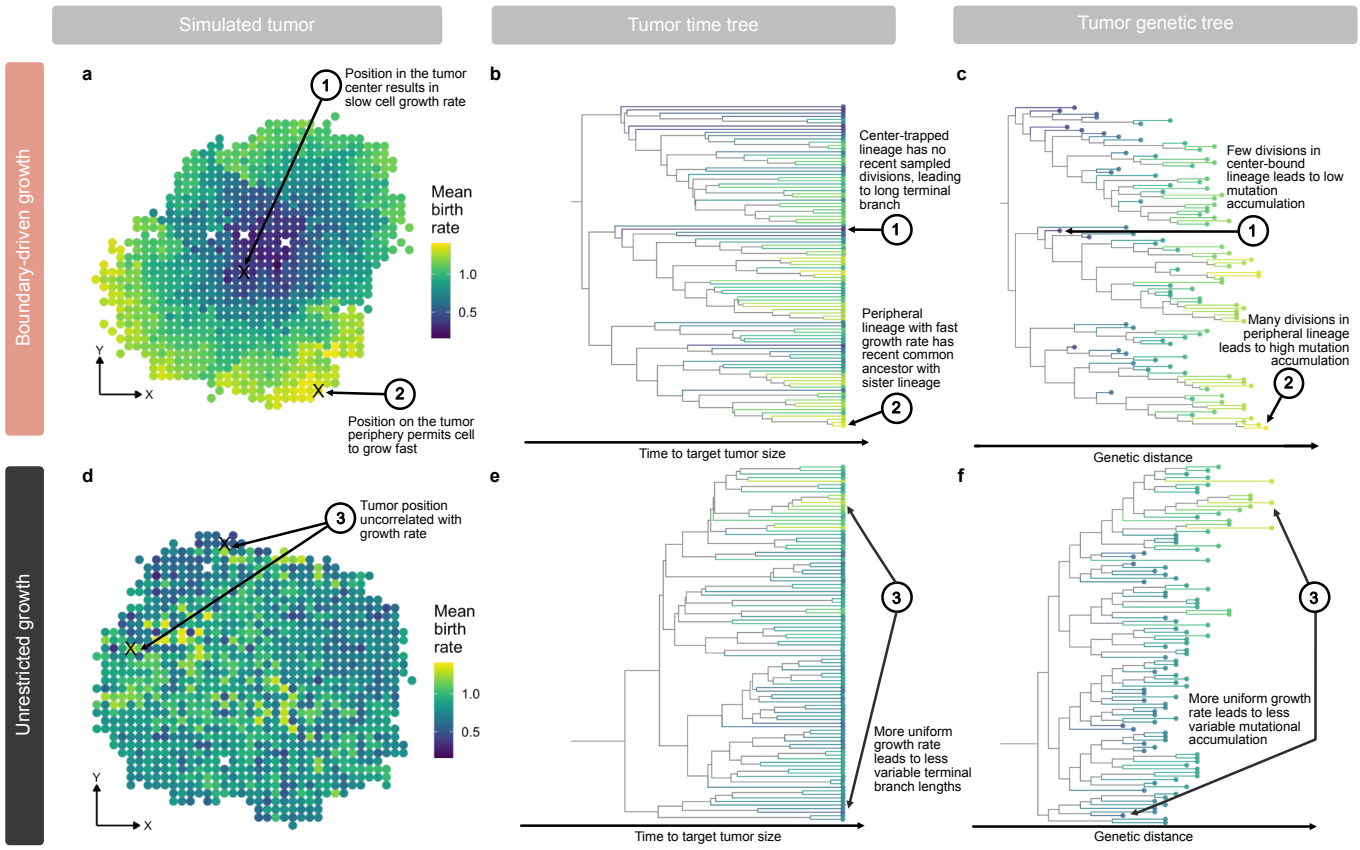

Figure 2.1: Boundary-driven growth causes characteristic tree patterns associated with asymmetrical division. Representative simulated tumors ($\alpha = 0.004$) showing variation in mean birth rate in 2D tumor space under **A.** boundary-driven growth via neighborhood-based spatial constraints, and **D.** unrestricted growth. Brighter colors represent higher birth rate (total number of divisions in cell lineage / simulation time) throughout all panels of the figure. **B.** Time tree of the representative boundary-driven growth tumor (subsampled to 100 cells for visualization) shows high variation in branching rates, leading to long terminal branches of center-trapped lineages. **C.** Genetic tree of the representative boundary-driven tumor (subsampled to the same 100 cells) shows ladder-like patterns due to mutation being tied to cell division. **E.** Time and **F.** genetic trees for the representative tumor under unrestricted growth (100 tips visualized) reveal less variation in branching rates and genetic distance.

ered two types of tree representations – 1) time trees (Figure 2.1B and E), where the branch lengths are in units of simulation time, and 2) genetic trees (Figure 2.1C and F), where the branch lengths are in units of number of mutations. We first compared the *time tree* of a tumor simulated under boundary-driven growth (Figure 2.1B) with one simulated with no spatial restrictions (Figure 2.1D). In the boundary-driven growth tree, we observed certain leaves (cells) with long terminal branches (i.e., cell 1) and other leaves with much shorter terminal branches (i.e., cell 2). These differential terminal branch lengths directly correspond to both mean lineage birth rate and spatial position within the tumor. Intuitively, lineages trapped in dense center neighborhoods (i.e., cell 1, Figure 2.1A and B) divide slowly and therefore exhibit longer times since diverging from another sampled cell. Conversely, lineages at the tumor boundary (i.e., cell 2) divide rapidly, and are therefore more likely to be recently related to another sampled cell. We quantified terminal branch lengths in the simulated tumor time trees and found that the asymmetries in birth rates due to spatial constraints result in an overall higher variance in terminal branch lengths under boundary-driven growth (Figure 2.2B) than under unrestricted growth (Figure 2.2C).

In Figures 2.1C and 2.1F, we reconstruct the *genetic trees* from the same boundary-driven and unrestricted tumor simulations. From this representation of the tumor trees, we observe that if mutation is linked to cellular division, then asymmetries in birth rates across tumor space logically correspond to varying rates of sequence evolution (Figure 2.2A).

This leads to repeated ladder-like patterns of genetic divergence that arise across multiple subclades of the boundary-driven growth tree in which fast-dividing cells on the tumor boundary accumulate more mutations than those in the interior (Figure 2.1C). These patterns are not observed in the unrestricted growth tree (Figure 2.1F). We quantified these patterns by measuring variance in mean clock rate (defined by total lineage mutations / simulation time) from extant cells in each simulation and demonstrate that clock rate is more variable across trees derived from boundary-driven growth than in trees simulated under the

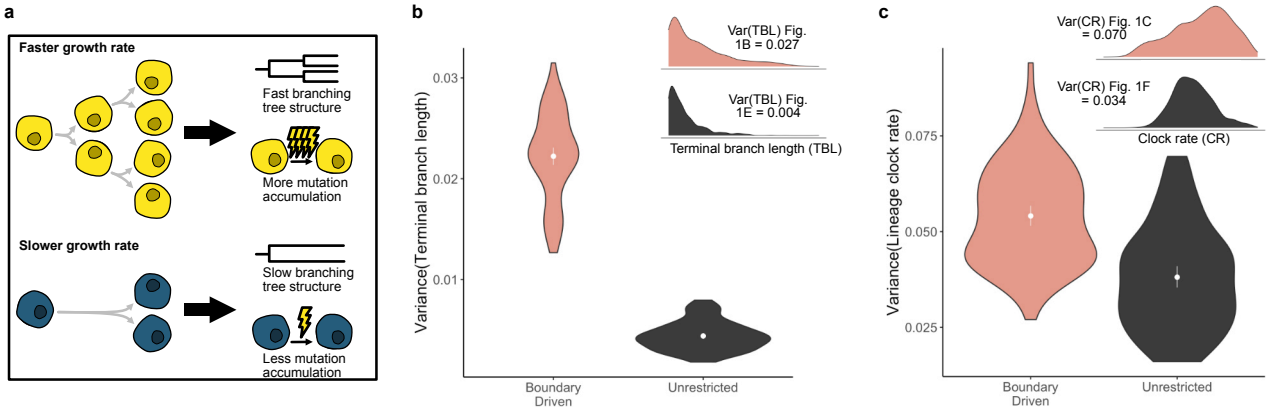

Figure 2.2: Quantifying signals of boundary-driven growth A. Cartoon schematic of the two signals of boundary-driven growth in trees left by asymmetric birth rates: variation in branching rates and variation in the number of mutations. Variance in **B.** terminal branch length (TBL) and **C.** clock rate (CR) in tumors under boundary-driven growth and unrestricted growth trees built from all extant tumor cells. Violin plots summarize statistics across 100 simulated tumors. Means (points) with 95% confidence intervals (error bars) are also indicated. Insets show distributions of TBL and CR signals for tumors plotted in Figure 2.1A and Figure 2.1B.

unrestricted growth model (Figure 2.2C).

2.1.2 Two-state birth-death process models boundary-driven growth

As tree structures differ between tumors simulated under boundary-driven and unrestricted spatial constraints, we sought a phylodynamic approach that could differentiate between these two growth modes. One such model is the multi-type birth-death model (49, 119, 120), which ties differential rates of birth, death, and sampling of lineages to multiple, discrete states. In our simulation studies, we observe that boundary-driven growth can be effectively simplified into two states. We find that the instantaneous cell birth rate under boundary-driven growth is elevated only in cells immediately adjacent to the tumor edge, but is uniformly low in all cells in the interior (Figure 2.3A). We can further decompose the tree patterns observed in Figure 2.1 into edge and center-linked dynamics. As shown in the

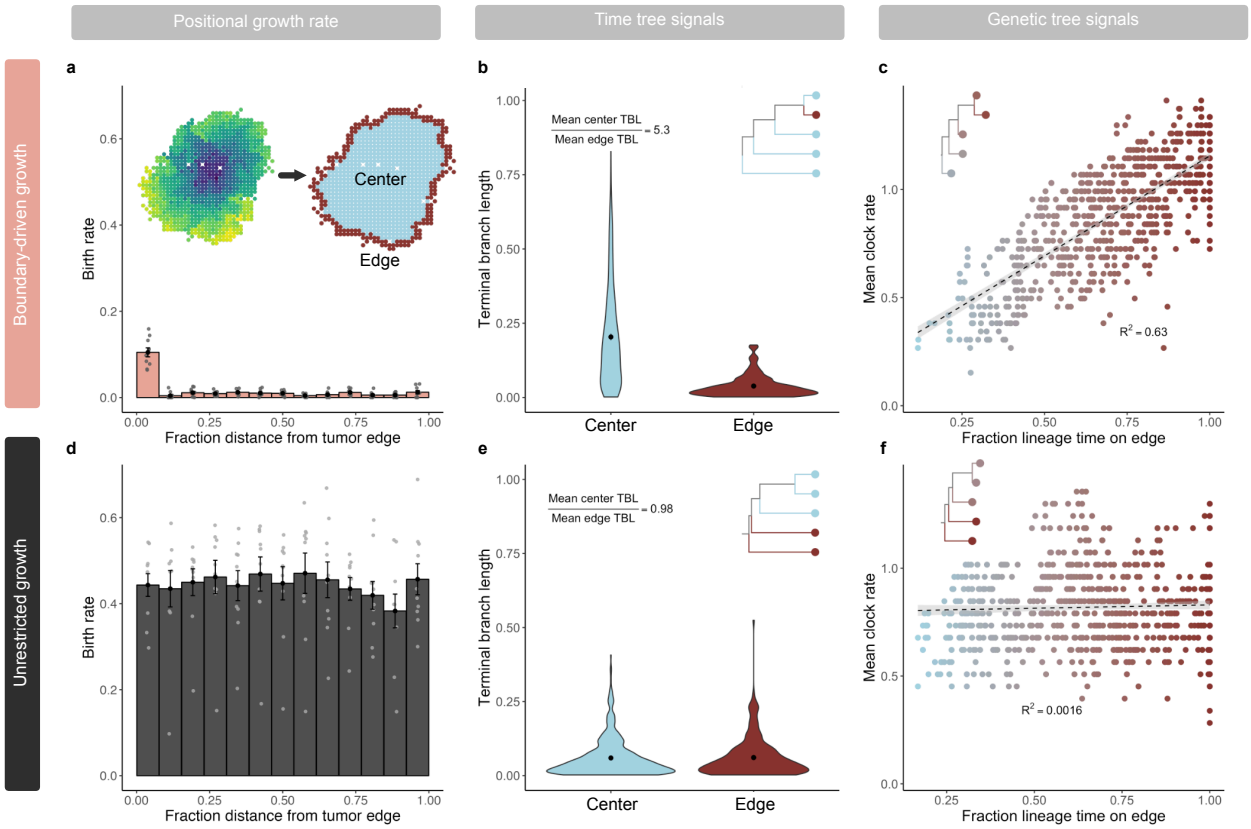

Figure 2.3: Asymmetries in cell birth rate and signals of boundary-driven growth in trees can be modeled by two-state dynamics. A. Histogram of instantaneous cell birth rate as a function of distance from the tumor edge (normalized by maximum distance). Rates are averages over 10 simulations under boundary-driven growth ($\alpha = 0.004$) with standard error bars. Points represent individual simulation means. **B.** Distributions of normalized terminal branch lengths (TBL) in a representative tumor under boundary-driven growth (Figure 2.1A) categorized by leaf edge or center state. **C.** Mean clock rate (total number of mutations / time) of cells in the example boundary-driven tumor versus the fraction of time a cell lineage spends on the tumor edge. Color gradient spans mostly center-associated lineages in blue to mostly edge-associated lineages in maroon. Dashed line is $y = x$. **D.** Histogram of instantaneous cell birth rate versus binned distance from tumor edge unrestricted growth simulations (average over 10 simulations with standard error bars, $\alpha = 0.004$). **E.** Distributions of terminal branch lengths for edge and center leaves in the representative unrestricted tumor (Figure 2.1D). **F.** Average lineage clock rates versus the fraction of time a lineage spends on the tumor edge in the example unrestricted tumor. Insets show representative phylogenetic patterns observed in time (**B** and **E**) and genetic (**C** and **F**) trees.

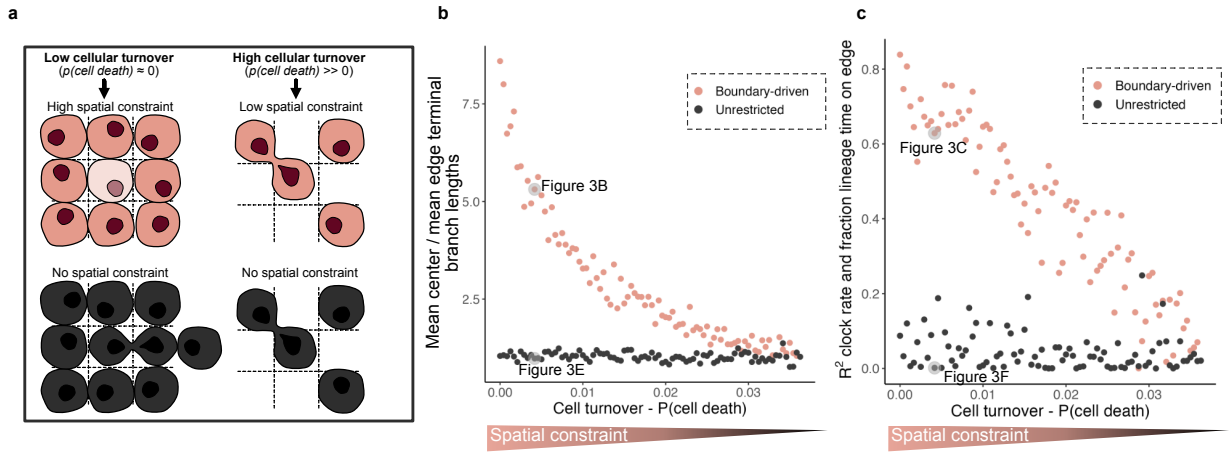

Figure 2.4: Asymmetries in cell birth rate and signals of boundary-driven growth in trees can be modeled by two-state dynamics (continued) A. Schematic comparing simulated spatial constraints under boundary-driven growth (coral) and unrestricted growth (black). **B.** Ratios of center-to-edge mean terminal branch lengths across simulations with decreasing spatial constraint (as modulated by cell death rate) under either boundary-driven or unrestricted growth modes. **C.** Correlations (measured via R^2) between the fraction of the lineage time spent on the edge and mean clock rate across the same range of spatial constraints and growth modes.

representative tumor from Figure 2.1A, all edge-associated cells have short terminal branch lengths. Whereas most of the variation in terminal branch length can be attributed to cells in the center and the mean terminal branch length of cells in the center is more than five times that of cells on the tumor edge (Figure 2.3B). If we trace the lineages of extant cells back to the root, the fraction of time cell lineages spend on the edge is highly correlated with the variation in mean clock rate observed in Figure 2.2 (Figure 2.3C, $R^2 = 0.63$). In other words, the most mutated cells have spent the majority of their lineage history on the tumor edge. Under unrestricted growth (Figure 2.3D), we observed no difference between edge and center terminal branch lengths (Figure 2.3E, ratio of center to edge mean terminal branch lengths = 0.98) and lineage time spent in the edge state is not correlated to clock rate (Figure 2.3F, $R^2 = 0.0016$).

To investigate the robustness of these patterns, we next simulated tumors under a wide range of cell turnover rates. Under boundary-driven growth, increasing cell turnover decreases spatial constraints and therefore lessens the growth advantage between edge and center states (Figure 2.10, Figure 2.4A). We measured the ratio of mean center to edge terminal branch lengths as in Figure 2.3B and E across these different effect sizes and found that this ratio is a consistent indicator of boundary-driven growth that decreases as spatial constraints are relaxed (Figure 2.4B). The correlation between fraction of lineage time spent on the edge and mean clock rate is also specific to the boundary-driven growth model and sensitive to effect size (Figure 2.4C). Therefore, we conclude that the patterns left by boundary-driven growth can be effectively approximated by a two-state birth-death model.

2.1.3 Phylodynamic models detect signals of spatial constraints

Two-state birth-death models incorporate how lineages divide, die, change states, and are sampled. In this class of models, birth events correspond to observed branching events on the tree and the rate of these branching events depends on an underlying type or state, in this case, the state of interest is the edge or center location within a tumor. Although existing phylodynamic models, such as BDMM'(120, 124) and BiSSE (49), permit asymmetrical division rates based on state, they do not link birth and mutation.

Therefore, although they are well-positioned to infer faster birth rates from branching structures, they cannot learn from differential rates of genetic divergence, a key hallmark of boundary-driven growth we observed in simulations. Additionally, branching patterns are prone to artificial inflation if more cells from a particular state are sampled in a clustered manner (125). Thus, existing models both do not incorporate all potential signals (i.e., clock rate differences) and, importantly, may be biased by sampling procedures in clinical tumor biopsies. To address these shortcomings, we introduce a **S**tate-**D**ependent sequence **evo**lution (SDevo) model to directly tie state-dependent birth rate to clock rate. This en-

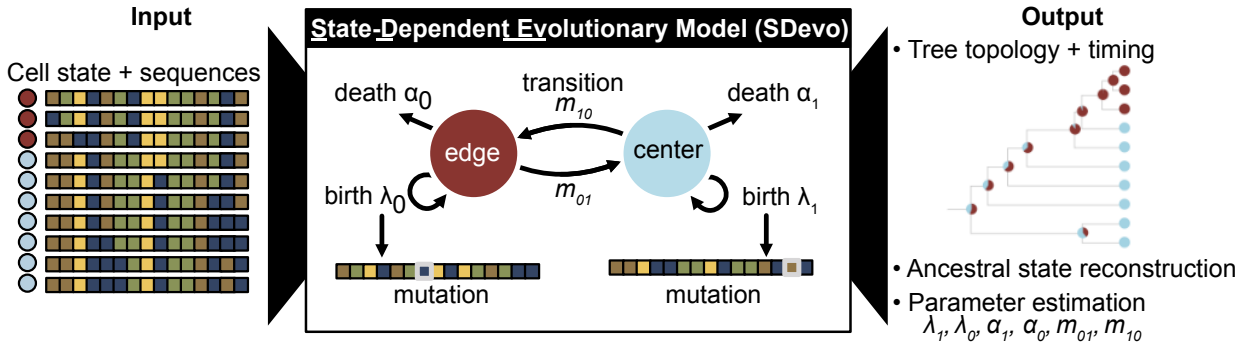

Figure 2.5: State-Dependent sequence evolution model (SDevo). Using input sequences and states (edge in maroon versus center in blue) of sampled tumor cells, SDevo reconstructs a tree with ancestral states (state probabilities represented by node pie charts) and estimates model parameters. SDevo links state-dependent clock rates to birth rates.

ables the model to learn from both clock rate and branching rate signals that arise from boundary-driven growth (Figure 2.5). SDevo accepts genetic sequences sampled from distinct spatial locations, along with a cell-state label (i.e., center versus edge). It generates posterior distributions of phylogenetic trees alongside joint estimates of phylodynamic model parameters (Figure 2.11). Inferred trees are time trees, which encompass the order and timing of cellular divergence events and include inferred internal node states, representing the location of unsampled ancestral cells. Model parameters include state-dependent birth and death rates, and the rate at which cells transition between states.

We first demonstrate the utility of SDevo on simulated tumors undergoing boundary-driven growth. From the genetic sequences and labeled cell-states for sampled cells isolated at a simulated tumor endpoint (Figure 2.6A inset), SDevo reconstructs the most likely relationship among sampled cells and the time at which those cells diverged (Figure 2.6A). The birth rates for edge and center-associated cells are inferred from the branching and mutational structure of sampled extant cells (leaves on the tree), permitting quantification of overall birth rate differences between the two spatial compartments (Figure 2.6B). SDevo

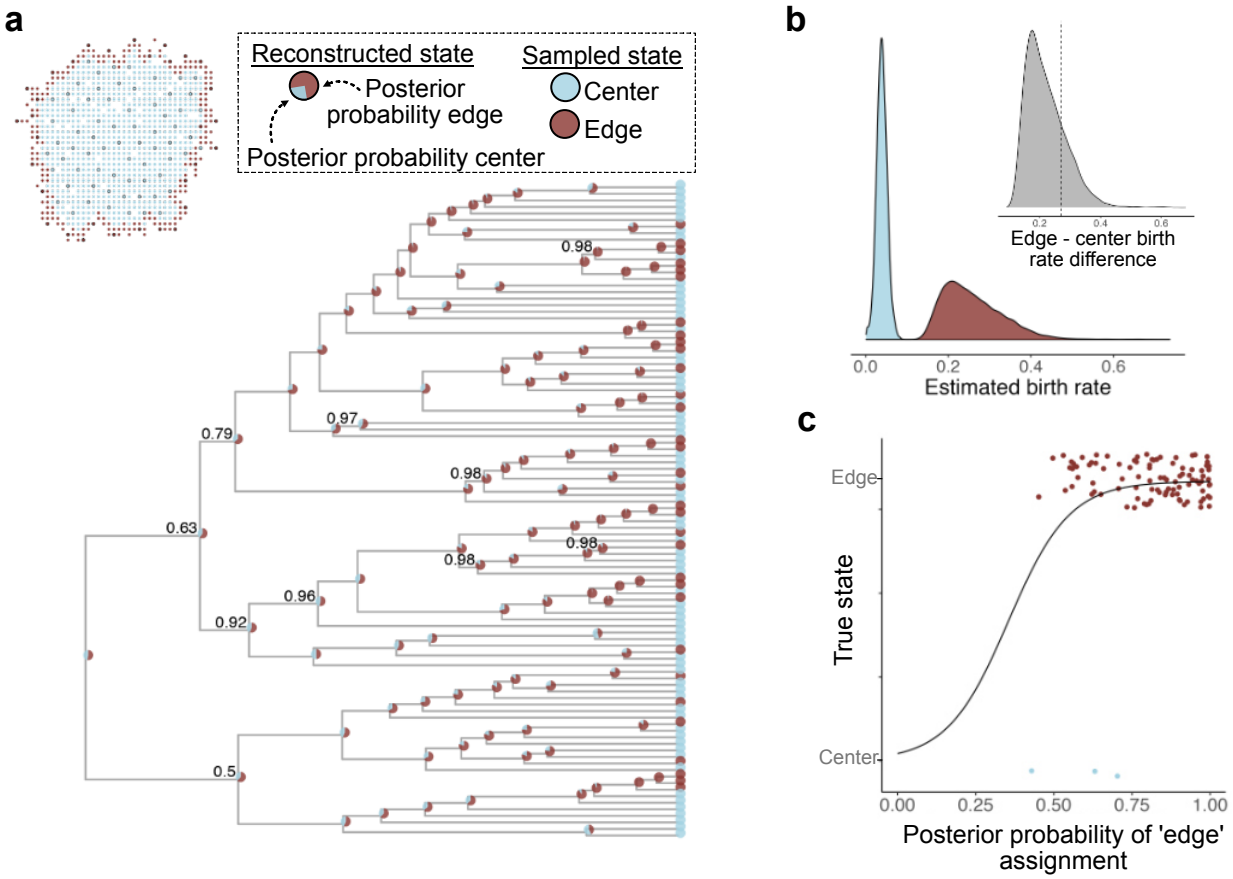

Figure 2.6: State-Dependent sequence evolution model (SDevo) estimates boundary-driven growth in simulated tumors. **A.** Reconstructed time tree estimated by SDevo on an example simulated tumor ($\alpha = 0.012$). At each internal node, the posterior probabilities for ancestral edge or center states are shown as a pie chart. Clade posterior support is indicated if less than 99%. The inset shows the sampling scheme for the tumor. **B.** Marginal posterior distributions of estimated edge and center birth rates, which are summarized by birth rate differences between edge and center cells (inset, dashed line indicates the true difference). **C.** Posterior probabilities of ancestral state reconstructions versus true state assignments.

correctly identifies that boundary-linked cells have a higher birth rate than center-linked cells (mean edge birth rate advantage = 0.22, 95% HPD = 0.12 - 0.35, true value = 0.27 in the representative simulation). SDevo additionally reconstructs the probability of each spatial state (center versus edge) for the ancestors of the sampled population (plotted as pie charts on the internal nodes of Figure 2.6A). These reconstructions suggest that the majority of ancestors divided on the tumor edge, consistent with the findings of Househam et al. (2022) (102) and our expectations of boundary-driven growth. SDevo's Bayesian approach further quantifies confidence in its ancestral reconstructions: ancestral cells with the highest posterior probability of existing on the tumor edge were indeed likely to have divided there (Figure 2.6C). On the other hand, cells with more uncertain ancestral reconstructions are less likely to have been on the tumor edge at division (Figure 2.6C). Finally, we applied SDevo to tumors simulated under a range of spatial constraints (see Methods). We find that at a moderate sample size ($n = 50$), SDevo is able to accurately quantify birth rate differences, whereas a two-state birth-death model without a state-dependent clock (BDMM' under a strict clock) fails (Figure 2.7A, Figure 2.12). We further observed that SDevo remains accurate for as few as 10 samples, whereas a strict clock model requires > 100 samples to reach close to the same accuracy (Figure 2.7B).

2.1.4 SDevo is robust to variation in sampling and growth modes

To evaluate SDevo's strengths and limitations in clinical tumors, we sought to validate that SDevo detects boundary-driven growth under various sampling strategies. Whereas in the initial simulation studies, we maximized the distance between sampled cells (i.e. diversified sampling), we also implemented a random sampling scheme as might be present in single-cell studies (Figure 2.13A). Under random sampling, cells sampled close together provide minimal additional genetic information but may create spurious signatures of rapid branching. Despite this, SDevo successfully estimates edge-driven birth advantages from

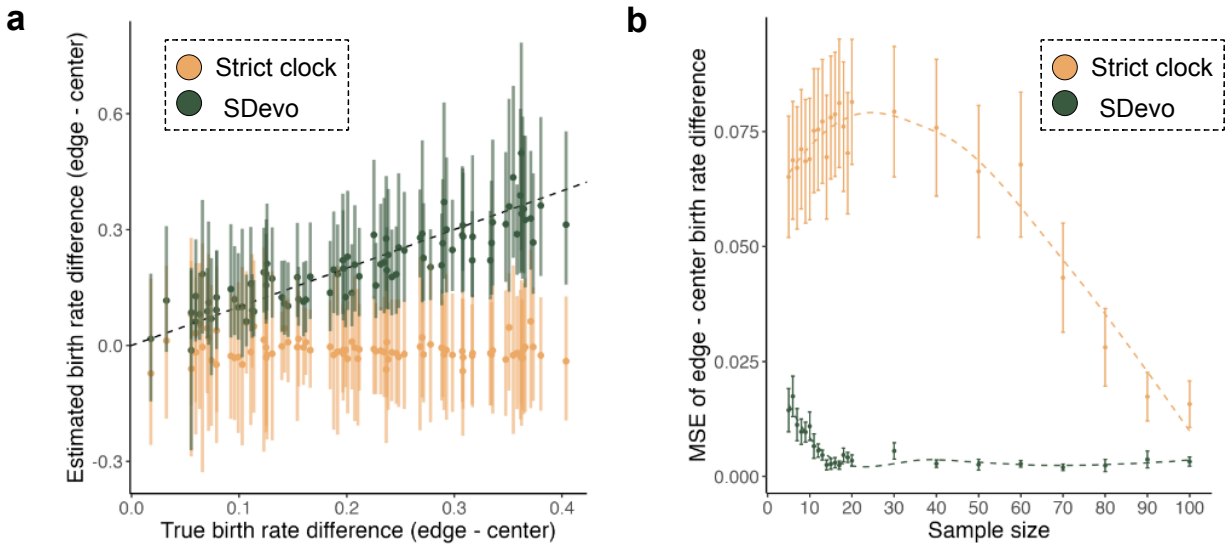

Figure 2.7: State-Dependent sequence evolution model (SDevo) estimates boundary-driven growth in simulated tumors (continued). **A.** SDevo (green) estimates of birth rate differences between edge and center samples across a variety of true per day birth rate differences (α varies between 0 and 0.036, $n = 50$) compared with estimates under a strict clock (gold). Points and bars represent mean and 95% HPD intervals, respectively. Dashed line is $y = x$. **B.** Mean squared error (MSE) of estimated birth rate differences in simulated tumors (α varies between 0 and 0.036) versus input number of cells sampled per tumor for SDevo (green) and strict clock sequence evolution (gold) models. Error bars represent the standard error of MSE and are summarized across 17 simulations per sample size.

randomly sampled cells (Figure 2.8A). In contrast, even with a large number of cells sampled ($n = 100$), the strict clock multi-type birth-death model often fails to detect the same birth rate differences (Figure 2.13B). We also assessed SDevo's robustness to punch biopsy sampling, in which a population of nearby cells is captured. We biopsy-sampled our simulated tumors, and only called mutations exceeding a 0.3 cellular fraction threshold within a punch (see Methods). We find that while punch-style sampling adds more random error due to variation in sampled diversity, especially in tumors with high turnover rates, SDevo largely still detects state-dependent birth rate effects (Figure 2.8B).

Next, we assessed SDevo's robustness to more complex growth by exploring an off-lattice model, a more flexible class of spatial models also employed to study tumor evolutionary dynamics (127–129). We simulated under a continuous space model of tumor growth implemented using the agent-based cellular engine PhysiCell (126). To mimic boundary-driven conditions, we linked division probability to mechanical pressure - cells crowded by their neighbors could not divide (see Methods, Figure 2.8C). As in the lattice-based simulations, higher cell turnover relaxes mechanical pressure, modulating spatial constraint. We first verified that SDevo continued to identify birth rate differences in these more complex simulations. We simulated 2D neutral growth, and found that SDevo sensitively detects an elevated birth rate at the tumor edge, even when birth rate differences were minimal (Figure 2.8D). However, SDevo slightly underestimates the birth rate differences at high death rates (i.e., low birth rate differences). We also confirmed that SDevo was robust to spatial division rate heterogeneity induced by increasing cell migration, as opposed to cell death (Figure 2.14A), and to a sigmoidal pressure threshold for cell proliferation (Figure 2.14C, D). We next simulated tumors grown in 3 dimensions and sampled across multiple z-slices, mimicking clinical sampling approaches. We determined that SDevo accurately reconstructs birth rate differences, albeit with wider posterior intervals (Figure 2.8E). We note that trees reconstructed from the 3D simulations tend to deviate more from expected edge-biased branching pat-

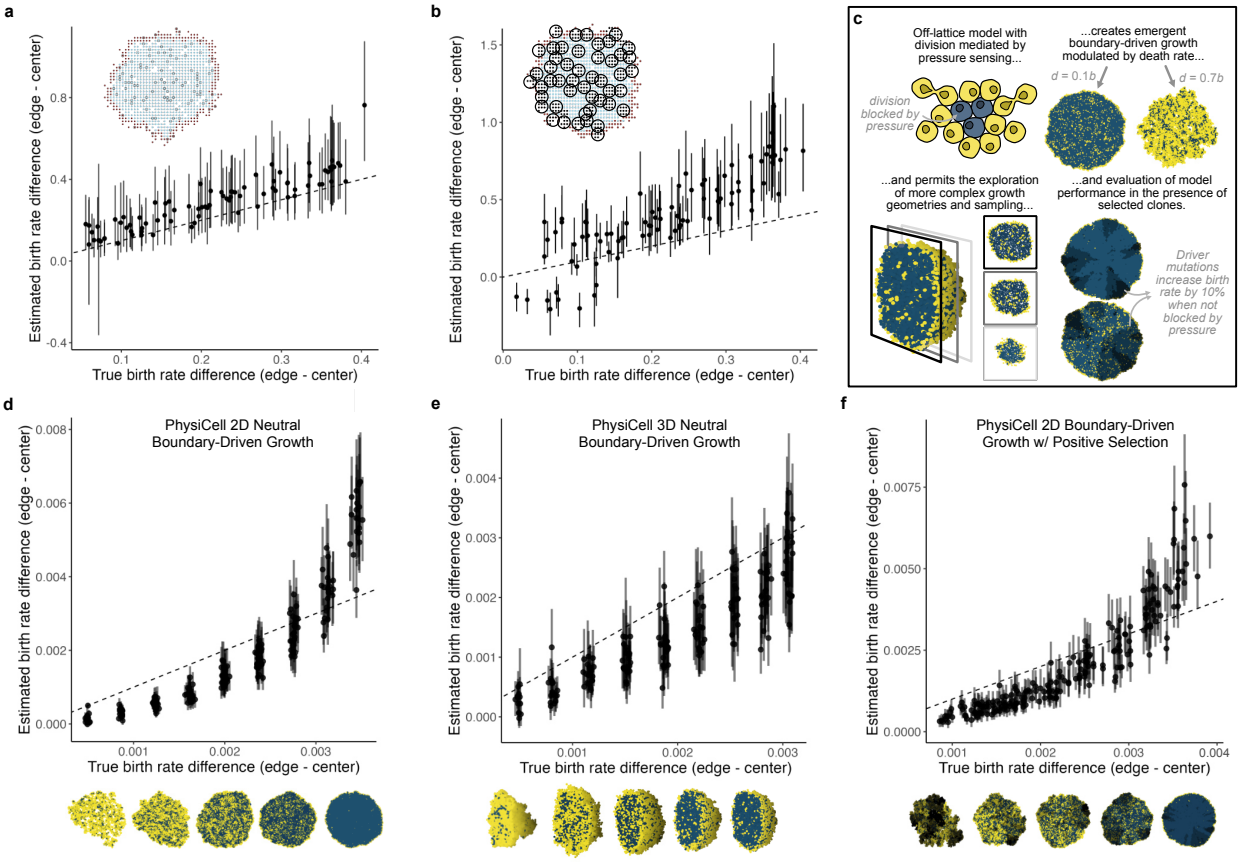

Figure 2.8: SDevo is robust to a variety of sampling approaches and growth modes. **A.** Mean and 95% HPD interval of estimated versus true edge - center per day birth rate differences with random sampling ($n = 100$ cells/tumor). **B.** Mean and 95% HPD interval of SDevo estimates of birth rate differences between edge and center when sequences are constructed from variants above 30% frequency in a simulated punch biopsy (inset, $n = 50$ punches per tumor). **C.** Schematic of continuous-space tumor growth simulation governed by biomechanics using PhysiCell (126). Only cells under low physical mechanical pressure from their neighbors (visualized in yellow as opposed to blue) can divide, generating boundary-driven growth. SDevo recovers growth rate differences generated through variable death rates under **D.** neutral 2D growth, **E.** neutral 3D growth, and **F.** 2D growth in the presence of strong driver mutations ($\mu_{driver} = 0.01$, $n = 100$ cells per tumor, driver fitness advantage = 10%, see Methods). Tumor snapshots below the x-axes show representative examples of growth dynamics under variable death rates (right to left: $d = 0, 0.2, 0.4, 0.6, 0.8$) and resultant pressure. In the case of the tumors under selection shown in **F.**, darker colors represent cells with driver mutations. Points and bars represent means and 95% HPD intervals, respectively, for all plots.

terns than those from the 2D simulations (Figure 2.15), reflecting more complicated growth dynamics and potential obfuscation via the sampling scheme. These observations further highlight the necessity of incorporating both branching and clock rate patterns to quantify boundary-driven growth in clinical scenarios.

Finally, we tested the extent to which SDevo detects boundary-driven growth dynamics when both spatially-determined and cell-intrinsic fitness differences influence growth, as the action of strong positive selection has been previously shown to distort the shape of tumor phylogenetic trees (82, 102, 34). We find that SDevo continues to detect differences in birth rates between center and periphery-associated cells even in the presence of strong selection (Figure 2.8F and 2.14B, see Methods). Notably, even as lineages with driver mutations expand, these cells are still subject to spatial constraints. As a result, similar patterns of branching and clock rate differences between center and periphery-associated cells re-emerge. However, we anticipate that if cell death is sufficiently high, a driver mutation could lead to rapid expansion of a center-bound lineage and mask signals of boundary-driven growth.

2.1.5 Boundary-driven growth in hepatocellular carcinomas

To quantify boundary-driven growth in a clinical tumor setting, we applied SDevo to multi-region sequencing data of two hepatacellular carcinoma (HCC) cancers published by Li et al. (34) (Figure 2.9). The authors sequenced two HCC tumors from a single patient, carried out 3-dimensional spatial micro-biopsy sampling followed by whole-genome sequencing (Figure 2.9A and E) and classified punches as "edge" or "center". The genetic maximum likelihood trees of each tumor (Figures 2.9B and F) qualitatively demonstrate an increased genetic divergence at edge punches. To apply SDevo, we created input pseudo-sequences for each punch using three independent 25,000 SNV random subsets of those identified in the original study. We assumed unidirectional transition from edge to center, in line with biological expectations of solid tumor growth, to constrain death and transition rate parameter space

(see Methods). SDevo jointly reconstructed tumor time trees along with the most likely ancestral internal node states. From these results we infer that while most ancestral cells divided on the tumor periphery, some population expansion occurred in the tumor center. We note that we do not use a predefined outgroup for this analysis, so there are slight differences in rooting for these time trees compared to the genetic maximum likelihood trees. SDevo found strong support for birth rate differences between edge and center in both tumors (Figure 2.9D and H). We estimated that cells on the edge have a mean 6.35x birth rate advantage over center cells in Tumor 1 (95% HPD = 4.53 - 8.32x), and a mean 2.83x birth rate advantage in Tumor 2 (95% HPD = 2.35 - 3.32x) summarized across all SNV subsets. To assess how sensitive these results were to differences in state classifications or punch heterogeneity, we also called alternate edge/center states based on a threshold of 10% of the tumor diameter (∼2mm and ∼1.5mm for Tumor 1 and Tumor 2, respectively) from the schematic boundary (Figure 2.16A and E). We found consistent results for Tumor 2, but observed that Tumor 1's alternate edge/center classifications showed more variable and reduced support for boundary-driven growth, which was not unexpected given that the alternate states updated the classification of previously center-assigned punches with less genetic divergence to edge (Figure 2.16D and H). We further found consistent results when removing a single punch from Tumor 1 (Figure 2.17), which may have captured multiple subclones (Figure 2.18).

Although we inferred a higher birth rate on the edge in these clinical tumors, the branching rate patterns in Tumors 1 and 2 qualitatively did not match our expectations from simulations. These branching patterns are potentially influenced by selection, as noted originally by Li et al. (2021), or by the non-uniform sampling scheme (Figures 2.9A and E). Likely due to these branching patterns, we find a strict clock model, which assumes independence of sequence evolution and cell division, did not detect boundary-driven growth. Instead, it estimated that center cells have a slightly higher birth rate (Figure 2.19). We note that

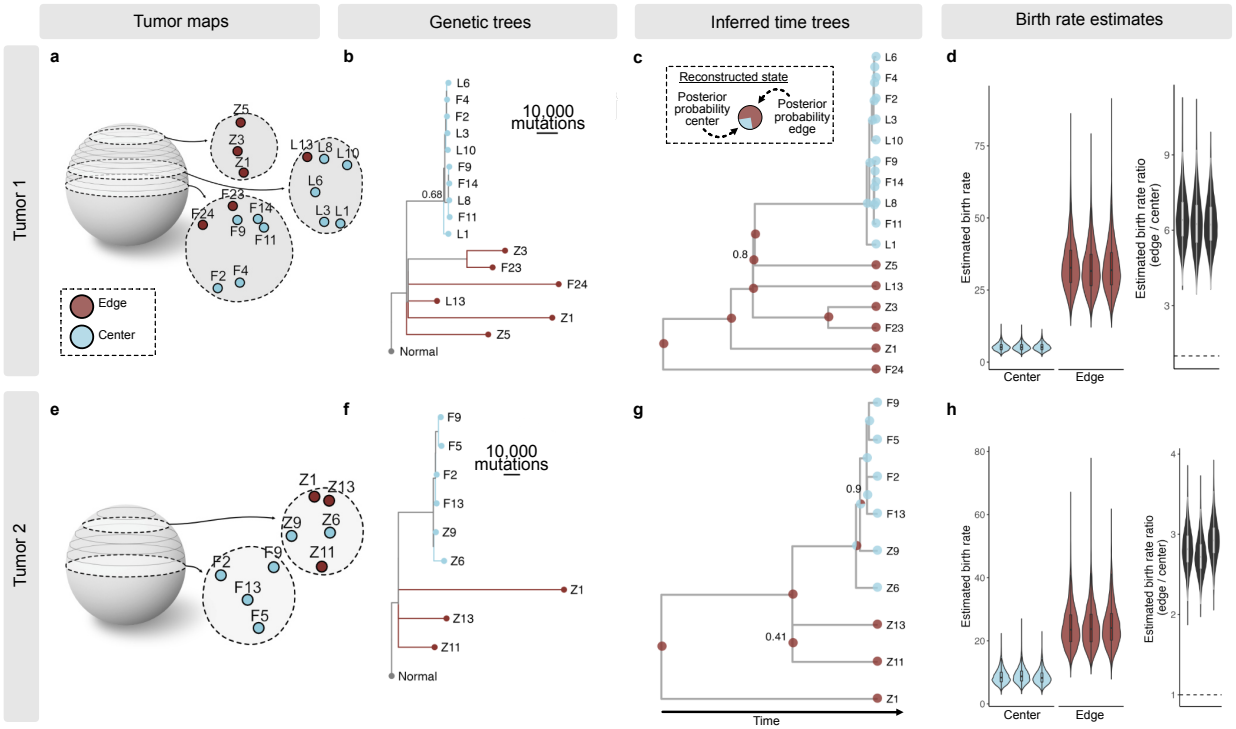

Figure 2.9: Quantification of boundary-driven growth in hepatocellular carcinomas. **A.** Multi-region 3D sampling map for Tumor 1 adapted from Li et al. (2021). Sampling locations are marked and labeled in z-slices, and center and edge classifications (taken from the original study) are shown in blue and maroon, respectively. **B.** Maximum likelihood genetic tree reconstructed from all variable sites. Node confidence is labeled if less than 99%. **C.** Tumor 1 phylogeny reconstructed from pseudo-sequences of one subset of variable sites. Tip colors indicate sampled punch state and pie charts on internal tree nodes represent posterior probabilities of ancestral state reconstructions. Clade supports are indicated at nodes if less than 99%. **D.** Marginal posterior distributions for edge (maroon) and center (blue) birth rates across three independent subsamples of variable sites, and estimated edge-to-center birth ratio (mean 6.35x). Posterior distributions summarize across three independent BEAST2 runs using all $n = 16$ punches. Dashed line marks ratio of 1. The same plots are shown for an additional HCC tumor (Tumor 2, $n = 9$) from the same dataset (**E-H**), estimating a mean 2.83x edge-to-center birth rate ratio. Tumor schematics are vertically inverted for ease of visualization. Violin plots show distribution density and the boxplots show IQR.

the sample sizes of Tumor 1 and Tumor 2 were well below the sample size requirements in simulations to detect boundary-driven growth with a strict clock model (Figure 2.5F). In addition, we found that incorporating a state-dependent sequence evolution model changed the estimated internal node timings (Figure 2.19C and F). Specifically, reconstructed center-bound nodes were estimated to have occurred more recently under a strict clock than under a state-dependent evolution model in which center cells would be expected to divide less frequently.

2.2 Discussion

Tumor evolutionary progression is a complex process driven by genetic, epigenetic, environmental, and immune factors. Quantitatively disentangling the contribution of spatial factors to tumor growth dynamics is an important component of both reconstructing tumor clinical histories and predicting future growth. Our understanding of spatial drivers of tumor growth has largely been informed by experimental models, as we have had limited ability to assay for these effects in clinical tumors. Here, we introduce SDevo, a new Bayesian phylodynamic model that learns differential cell birth rates of discrete classes (here, tumor periphery or center-associated). Although SDevo is general in scope and applicability, in this study, we demonstrate that it successfully infers birth rate differences between the tumor edge and center from multi-region sequencing data. We show that SDevo is relatively robust to sampling choices (i.e., punch biopsies and locations) and biological factors (i.e., cancer driver mutations and 3D versus 2D growth modes). We further find quantitative evidence for boundary-driven growth in clinically-derived hepatocellular carcinomas resected at a single time point.

Our assessment of boundary-driven growth in HCC quantitatively expands the observations of Li et al. (2021). The authors originally hypothesized that Tumor 1's tree structure matched a simulated scenario of boundary-driven growth followed by the expansion of a

selected clone in the center and that Tumor 2's tree structure matched dominant boundary-driven growth. The authors made these assessments by simulating tumors and comparing the distributions of clones and variant allele frequencies to the sequenced tumors. They further noted that genetic divergence was higher in punches collected from the tumor periphery.

Our study quantifies these patterns by estimating these birth rate differences directly with joint inference of tree topology and sequence evolution. Notably, although small sample sizes, clustered sampling, and the hypothesized selection for an internal clone in Tumor 1 may have distorted the branching structure of the trees, SDevo is able to detect past boundary-driven growth from clock rate differences. By explicitly incorporating the mutational process, SDevo leverages data more effectively than models that only learn from state-dependent branching. This approach is particularly important when only a few areas of a tumor are sequenced. These findings, along with previous *in silico* evidence that selection changes the shapes of tumor trees (82, 130), highlight the importance of employing multiple tree patterns to quantify interacting modes of tumor growth. Although future work should more comprehensively profile how multiple spatial and non-spatial drivers of growth can impact observed tree patterns, our analysis of non-neutral tumors (Figures 2.8F and 2.14B) suggests that SDevo can detect boundary-driven growth in the presence of selection.

Quantifying the impact of spatial restrictions on clinical tumor growth informs how we understand, predict, and control cancer evolution. A robust literature has established that boundary-driven growth modulates the efficiency of positive and purifying selection (83, 131), alters overall growth rates (104, 132), and increases the efficacy of adaptive therapy (39, 84, 85). Spatial restrictions also change the expected distribution of genetic variation in solid tumors (80, 81, 79, 131), and impact how clinically informative biopsies should be collected (133). Although we find robust evidence for boundary-driven growth in HCC, its prevalence and strength likely vary by stage of tumor growth and tumor type (36). For example, increased vascularization, cellular migration, physical anatomical structures, or

tumors reaching a local carrying capacity could alter spatial growth restrictions. Further applications of SDevo to other tumor cases and types will enable us to explore the nuances of these growth phenomena.

Importantly, the utility of SDevo is not limited to understanding the impact of boundary-driven growth, but in fact, can be applied in any instance in which sequenced tumor samples can be classified into discrete, observable states. Immediately, SDevo could be extended to test other proposed tumor growth modes – for example, growth against a solid surface, such as bone in osteosarcoma, along a unidirectional invasive front (134), or in different glandular compartments (135). Because tumors can grow under a wide variety of anatomical constraints, integrating system-specific factors can help assign biologically-relevant environmental states for the application of SDevo (i.e., edge categorization may constitute those cells that have penetrated the basal layer as opposed to those that are most radially extreme). Even more broadly, SDevo could be applied to study the growth impacts of other environmental or cell-intrinsic factors, for instance, immune invasion, hypoxia, metastatic versus primary sites, or genetic features, by decomposing complex phenotypes into discrete states.

Phylodynamic approaches such as SDevo have major advantages compared to our current approaches for estimating evolutionary information from tumors, namely approximate Bayesian computation (ABC) (136) or other approaches that compare simulated and clinical tumors via summary statistics (36). To be clear, these approaches have yielded extensive insights into tumor evolution, including patterns under boundary-driven growth (28, 80, 82, 102). However, these approaches are computationally costly, requiring the generation of often tens or hundreds of thousands of simulated tumors, on which one must compute extensive summary statistics. In addition, ABC comes with technical challenges, including the necessary choice (and potential unavailability of) low dimensional sufficient summary statistics. Although Bayesian phylodynamics comes with its own technical chal-

lenges (i.e. identifiability, sensitivity to model assumptions, choice of priors, see Louca and Pennell (2020) and Louca et al. (2021) (137, 138)), it does not require tumor simulation. Furthermore, the generality of discrete traits affecting growth dynamics means it is easily adaptable to answer new questions. While both ABC and phylodynamics offer ways to understand clinically-derived samples, the full promise of phylodynamics has yet to be widely exploited.

Phylodynamic approaches to understanding tumor evolution offer additional benefits: 1) used in conjunction with well-calibrated molecular clocks, inferred trees can help estimate the timing of clinically-important events, such as the emergence of subclones or metastatic events. While these analyses have been employed in the context of uniform growth rates (139, 140, 116), the expansion of tree models to permit differential birth rates could improve timing accuracy. 2) Incorporating differential growth rates across a tree can lead to more accurate tree topologies, as has been demonstrated in influenza evolving in multiple host species (141). 3) Inferring ancestral states can elucidate population history and tumor evolutionary processes at time points that cannot be clinically sampled. Recently, Zhao et al. (2021) (106) analyzed the intra-tumor spatial and genetic architecture of renal cancers and concluded cells in the tumor center are more likely to seed metastasis. However, the study was limited to observing the *extant* position of these samples, whereas SDevo reconstructs these states at the time of clinical events (i.e. divergence of a metastatic clone). These three points suggest more broadly how tumor trees can be leveraged to gain new quantitative insights into tumor evolution, and demonstrate the broad utility of modeling evolutionary processes on trees.

Beyond its application to cancer evolution, SDevo is a novel phylodynamic model with broad usefulness to incorporate state-dependent clock rates into evolutionary inference. While the field of phylogenetics has developed a broad array of clock models, SDevo represents the first model in which clock rate is linked to population birth. SDevo could be particularly useful in microbial and viral populations where diversification and mutational

accumulation operate on similar timescales, and may be linked to underlying state variables (for example, location). We demonstrated that incorporating clock rate differences, instead of relying solely on tree diversification rates (as in BDMM' and other multi-state birth-death models (49, 120, 124)), can improve inference in cases where sampling may be non-uniform. This may be particularly important when sampling rates vary - for example, countries with variable rates of molecular surveillance for SARS-CoV-2. To facilitate broad application, SDevo is built as a package in the popular Bayesian phylogenetic platform BEAST2 (121). As with all phylodynamic models, identifiability represents a pervasive concern, but incorporating biological knowledge for determining priors can help constrain the model space. In our analysis of HCCs, we use information about cell transition and death rates to distinguish between multiple parameters that impact trees and estimation in interrelated ways.

Biological complexity within tumors can complicate SDevo's application and interpretation via spatially- or temporally-varying selection. First, strong selection can destroy or alter signals of boundary-driven growth (82, 34). For example, a hard bottleneck, as in the cases of surgery or chemotherapy, would likely temporarily destroy signals of boundary-driven growth. Such signals would likely also re-emerge were the tumor to regrow via boundary-driven growth. Second, gain of driver mutations will lead to cell-intrinsic fitness differences that may not correlate with spatial location. Third, disentangling boundary-driven dynamics from other environmental or cell-intrinsic factors could be especially difficult under time-varying selection. For example, angiogenesis could increase resources to center cells later in tumor growth (142) and complex cell-to-cell interactions may create frequency dependencies that further complicate observed spatial patterns (109, 84, 143). We have shown that SDevo can detect signals of boundary-driven growth even with driver-induced selection, but future work should further probe this robustness.

Although SDevo is a powerful tool, we note several important limitations that require further caution when applying it to data. First, SDevo assumes mutations occur at cell

division. If instead, most mutations emerge due to exogenous processes (144), birth-driven genetic divergence could be masked. While this might decrease SDevo's power, exogenous mutational processes distributed evenly across a tumor are unlikely to generate false positive signals of boundary-driven growth. Second, extensive cell mobility could weaken signatures of boundary-driven growth even if boundary-associated cells have birth rate advantages. Third, as we demonstrate in Figure 2.5, sample sizes must be sufficient to detect state-dependent effects. We maximize limited sample sizes by choosing priors that are biologically informed (for example, unidirectional state transitions), but larger sample sizes will enable inference with less informative priors. Data sets that meet this requirement are becoming rapidly available, so we anticipate phylodynamic models such as SDevo becoming increasingly powerful.

The expanded application of phylodynamics to cancer sequencing data relies both on developing methods to exploit single-cell sequencing data (145, 146), and understanding the relationship between sequenced multi-region punches and the many single cells that comprise them. As has been noted previously, multi-region sequence trees are not phylogenies (147), and punch-wide genetic composition does not necessarily capture all cellular genotypes (148). Although SDevo is fairly robust to our simulated punch-style sampling and we analyzed HCC data from small, largely homogeneous punch biopsies, best practices for applying phylodynamic models to trees of deconvoluted clones are an important area for future research.

Applying phylodynamic methods to tumor populations is in its infancy, but new methods that overcome the barriers of working with tumor data will help extend the applicability of these approaches (149, 145). Here, we demonstrate the utility of phylodynamic models in quantifying spatial factors driving cancer progression. As technologies enabling the widespread and high-throughput generation of tumor trees advance (130, 150), we expect adapted phylodynamic approaches such as SDevo to provide a rigorous analytical toolkit for

extracting quantitative insights from these data.

2.3 Methods

2.3.1 Tumor simulations

Eden model:

An agent-based model was implemented in Python3 which places simulated cells on a 2D lattice. Simulations are initiated with a single cell in the center of the lattice. At each time step (2 hours) cells have a probability of dying (α) and a probability of attempting division given survival (λ). Under boundary-driven growth, cells only successfully divide if there is an empty lattice spot in its Moore neighborhood. If multiple neighboring spaces are available then the cell randomly chooses the location for its daughter cell from open neighboring spaces. Under unrestricted growth, if a cell attempts division, its daughter cell will occupy an empty lattice spot in the Moore neighborhood if available, but if not, the cell will still divide and push cells in a random direction to make space. Overlapping cells are pushed in the same direction until a neighboring lattice spot is available, which the pushed cell will occupy. In both simulations, if a cell divides, each daughter cell can gain mutations with probability μ (per division per genome). Mutations are then drawn from a Jukes-Cantor model of sequence evolution and follow an infinite-sites assumption. Therefore, each time a mutation is gained, a site is added to all cells in the simulation. Simulations are stopped when the number of living cells is more than 1000. The ground truth birth rates are assessed at discrete time points in the simulation by recording the current state of each cell and the proportion of cells that have progeny in the next time step. True birth rates are considered to be the mean across all time steps weighted by the number of cells in each category. This method calculates effective birth and death rates on the edge and center given the simulated spatial constraints by calculating empirical division rates on the edge and center of cells through simulated time. Effective spatial constraints in the boundary-

driven model were controlled by changing the probability of cell death, where increased cell turnover allows center-trapped cells to divide more readily (Figure 2.10). To evaluate the accuracy of parameter estimation, we ran 1000-cell tumor simulations where the probability of cell death per time step, α, varied from 0 to 0.036, the probability of attempting division given survival, λ, was 0.04, and the rate of mutation per division was $\mu = 1$. Not accounting for spatial constraints, these birth and death parameters translate to an approximate 0.32 - 0.40 per day probability of division (per cell) and a range of 0 - 0.35 per day (per cell) probability of death. Although clinical tumors have large variability in rates of proliferation, death, and mutation, these parameters fit within this biological range (151–153, 79).

Eden tree statistics

Tree statistics in Figures 2.1, 2.2 and 2.3 were calculated from simulated tumor trees that include all extant cells. Normalized terminal branch lengths were calculated by dividing terminal branch lengths of tumor time trees by total simulation time. Clock rates were calculated by dividing the total number of mutations accumulated in each alive cell by simulation time. Edge and center states for terminal branch lengths are defined by cell location at the end of the simulation, where edge cells are defined by being the most extreme cell on either the X or Y spatial axis for each row and column, respectively, or within one cell of this boundary. The fraction of the lineage time spent on the edge is determined by averaging across all lineage node states weighted by time tree branch lengths.

Continuous space model

To probe the robustness of SDevo to more complex selective events and higher dimensions, we implemented an additional set of simulations in the physics-based cellular simulator, PhysiCell (126). Briefly, PhysiCell is an open-source, agent-based model implemented in C++ in which cell movement is governed by biomechanical interactions among cells. To

simulate boundary-driven growth, we created a PhysiCell instance in which cells are only able to divide when under low mechanical pressure, using the cell state variable, *simple_pressure*. As a result, similar to the Eden model, most cell division is restricted to the tumor periphery, or to cells with adjacent space created by the recent death of a neighboring cell. Cells initially divide at a rate we arbitrarily set to 1, except when above the pressure threshold, τ, in which case, they divide at rate 0. We also explored a sigmoidal relationship between pressure and birth, where the birth rate $b = 1 - (1 + exp(-5 * (pressure - \tau)))^{-1}$. Cells die at rate d, regardless of their pressure status. To simulate selection, during each cell division, a daughter cell can acquire a driver mutation conferring a 10% fitness advantage (154) with probability μ_{driver}, which acts multiplicatively (i.e. a cell with two drivers has a 21% faster growth rate than one with 0) (155). Tumors are grown to a final size of N extant cells, of which n are sampled. After the simulation, a Poisson-distributed number of neutral mutations is augmented to each cell division with $\lambda = \mu_{passenger}$. Using the continuous space model, we investigated all pairwise combinations of 2D and 3D, neutral and selective scenarios, and ran 25 tumor simulations for each combination of parameters ($\tau = 1$, $d = (0, 0.1, 0.2, ...0.8)$, $\mu_{pass} = 1$, $n = 100$), except for 3D selection, where we simulated $d = (0, 0.2, 0.6, 0.8)$ with 10 tumors each. For the 2D models, $N = 10,000$ and for the 3D model, $N = 15,000$. For the selective model, $\mu_{driver} = 0.01$ and for the neutral models, $\mu_{driver} = 0$. Note, we used a value of μ_{driver} well above expected rates of driver mutations ($\approx 10^{-5}$, (151)) to conservatively test SDevo in an extreme case of selection. To probe SDevo's performance when cellular constraint is reduced by migration instead of cell death, we performed 10 simulations at $d = 0.2$ where cells migrate at $0, 0.5, 1, 1.5$ or $2 \mu m/minute$ at an angle drawn from $[0, 2\pi]$ and updated on average each minute (all other parameters as above). To probe SDevo's robustness under a sigmoidal relationship between pressure and birth rate, we ran 10 simulations with $d = (0, 0.2, 0.4, 0.6)$, and all other parameters as above. One outlier in the 3D boundary-driven growth simulations was removed due to convergence on a local

optimum. Ground truth edge and center birth rates were determined by first classifying cells as within 10 microns (approximately 1 cell width) of the tumor periphery as edge, and those more than 10 microns from the edge as center. The average birth rate was computed separately within each of those classes over multiple discrete time points (10-40, depending on the overall rate of tumor growth) and combined by a weighted average according to the number of cells at each time point. Cells under too much pressure to divide at the sampled time ($simple\_pressure > \tau$) were calculated as having an instantaneous birth rate of 0.

2.3.2 Sampling procedures

2D simulations were sampled by maximizing the distance between sampled single cells in physical space (diversified sampling). This ensures that a sufficient number of edge and center classified cells were sampled and that sampled cells were not clustered. Bulk punch biopsy sampling was mimicked by choosing a center cell and a target of 8 cells immediately surrounding that was grouped into a single punch. Punches were iteratively drawn and shifted if they overlapped with a previously punched group of cells. Sampling ended when the target number of punches was reached (50 punches) or sampling was no longer possible without significant overlap. Punch sequences were generated using all mutations above a cellular fraction cutoff of 0.3. 3D sampling was approximated by taking 5 simulated slices through the tumor z-plane at 2/8ths, 3/8ths, 4/8ths, 5/8ths and 6/8ths of the range of the z values of a given tumor. Within each slice, cells were sampled to maximize the inter-cell distance, as described above, and the number of cells per slice was proportional to the number of cells in the slice relative to the number of cells across all slices.

2.3.3 Multi-type birth-death models of to boundary-driven growth

The birth-death process describes how lineages duplicate (birth), die (death), and are sampled (where samples are tips on a phylogenetic tree) (156). The multi-type birth-death model

extends this by considering birth, death and sampling to occur in different states (sometimes also referred to as different sub-populations, traits, or types) and how lineages jump between these states. The rates of birth, death and sampling vary depending on the state of a lineage. For the case of boundary-driven growth, we model a two-state process, with one state denoting cells in the center of the tumor and the other state denoting cells on the edge of the tumor.

2.3.4 Posterior Probability

To perform Bayesian inference, we define the posterior probability $P(T, \sigma, \theta|D)$ of the timed phylogenetic tree T, the evolutionary model and parameters (σ), and the population model and parameters θ, given the data, D. This posterior probability is typically expressed as:

$$P(T, \sigma, \theta|D) = \frac{P(D|\sigma, T)P(T|\theta)P(\sigma)P(\theta)}{P(D)} \tag{2.1}$$

In the case of the state-dependent multi-type birth-death model, we cannot assume the tree likelihood $(D|\sigma, T)$ and the tree prior $P(T|\theta)$ to be independent, as the rate of evolution directly depends on the population model. In other words, how fast evolution happens on a lineage depends directly on the state of that lineage. We therefore define $\mathcal{H}$ as a mapped state transition history, that contains a random mapping of state change events given a set of parameters θ of the multi-type birth-death model. We then define the tree likelihood as $P(D|\sigma, \theta, T, \mathcal{H})$. Additionally, we say that instead of computing $P(T|\theta)$ directly, we only compute the tree prior for one realization of the state transition history, i.e. $P(T, \mathcal{H}|\theta)$. The posterior probability then becomes:

$$P(T, \mathcal{H}, \sigma, \theta|D) = \frac{P(D|\sigma, \theta, T, \mathcal{H})P(T, \mathcal{H}|\theta)P(\sigma)P(\theta)}{P(D)}. \tag{2.2}$$

Performing MCMC inference to characterize this posterior probability distribution would require integrating over all transition histories $\mathcal{H}$ using MCMC. This is overall incredibly slow

and limits the application of the method. Instead, we formally integrate over all possible histories $\mathcal{H}$, to get the following posterior probability:

$$P(T, \sigma, \theta|D) = \frac{\int_{\mathcal{H}} P(D|\sigma, T, \mathcal{H}) \int_{\mathcal{H}} P(T, \mathcal{H}|\theta) P(\sigma) P(\theta)}{P(D)}. \tag{2.3}$$

$P(T|\theta) = \int_{\mathcal{H}} P(T, \mathcal{H}|\theta)$ is computed as described in Kühnert et al. (2016) (120), which is achieved by treating the states of lineages probabilistically instead of discretely.

Lastly, we set $\int^{\mathcal{H}} P(D|\sigma, T, \mathcal{H}) = E[P(D|\sigma, \theta, T, \mathcal{H}) = P(D|\sigma, \theta, T, E[\mathcal{H}])$, with $E[\mathcal{H}]$ being the expected/average state transition history, which contains, for each lineage i in the phylogeny, its expected time spent each state s. This leaves us with:

$$P(T, \sigma, \theta|D) = \frac{P(D|\sigma, \theta, T, E[\mathcal{H}]) P(T|\theta) P(\sigma) P(\theta)}{P(D, S)}. \tag{2.4}$$

2.3.5 Modeling birth-dependent evolution

In order to model different rates of evolution for different states, we first compute the expected time each lineage in the phylogenetic tree T spent in each state. To do so, we use a stochastic mapping approach related to those described in Nielsen (2002) and Huelsenbeck et al. (2003) (157, 158). We first compute the probability $g_s^{i,b}$ of each lineage i in the phylogenetic tree being in any possible state s over time t from the tips to the root as described in Kühnert et al. (2016) (120). These state probabilities are conditional only on events that occurred more recently than t and therefore not on all events in the phylogeny. During this backwards propagation, we keep track of the time-dependent transition matrix $Q(t)^i$ that describes the rate of probability flow between any two states at time t due to state transitions or birth events between states. As a result, once we reach the root, $g_s^{i,b}$ contains all events in the phylogeny and is therefore equal to $g_s^{i,f}$, i.e. the forward probability of lineage i being in state s.

Following Stolz et al. (2022) (159), we first define q_{ab}^i as:

$\mathrm{q}_{ab}^i = \mu_{ab} \frac{g_b(t)_a^i}{g_b(t)_b^i}$ with μ_{ab} being the rate of state change due to state transitions or cross-birth events.

We then compute the probabilities of any lineage being in any possible state conditional on all events in the phylogeny $g_s^{i,f}$ forwards in time as: $\mathrm{dg}_s^{i,f} \frac{}{dt=\sum_{a=1}^{states}\left(q_{as}^i g_a^{i,f} - q_{sa}^i g_s^{i,f}\right).}$

By keeping track of the forward probabilities $g_s^{i,f}$ on each lineage, we can then compute the expected time t_s^i that lineage i spends in any of the possible states s. The values for t_s^i make up the entry for $E[\mathcal{H}]$ in the posterior distribution posterioreq. We then say that c_s is the rate of evolution, that is the clock rate, of a lineage in state s. Next, we compute the average rate of evolution on branch i, c^i as

$\mathrm{c}^i = \sum_s^{states} t_s^i * c_s.$

At each replication, an error in copying the genetic material of a cell can occur. These errors tend to be more likely in cancer cells, where cellular control mechanisms are often faulty. Phylogenetic methods typically assume the evolutionary processes to be independent of population processes, such as cell replication. To model mutations happening at birth events, we assume that the birth rate b_s in state s and the clock rate in state s are proportional such that $c_1 = c_{avg} b_1, c_2 = c_{avg} b_2, ..., c_n = c_{avg} b_n$.

2.3.6 Implementation

We implemented the multi-type birth-death model with state-dependent clock rates as an addition to the Bayesian phylogenetics software BEAST2. SDevo depends on BDMM-prime v0.0.30 (https://github.com/tgvaughan/BDMM-Prime) to compute the tree prior $P(T|\theta)$ and is built in Beast v2.6.6. To model mutations occurring at cell division, we set the relative rate of evolution in the different compartments (edge and center) to be proportional to the birth rates in these compartments. The implementation itself does not explicitly require this assumption and the relative rates of evolution can also be treated as a distinct parameter in the inference. All SDevo analyses were performed using SDevo v0.0.2. SDevo can be installed through the interface BEAUti.

2.3.7 Validation

To validate the implementation, we perform a well-calibrated simulation study. In it, we simulate phylogenetic trees under a two-state birth-death model in which we assume the rate of evolution to be proportional to the birth rate in either compartment. We randomly sample the birth, death, and transition rates from the prior distribution, while fixing the sampling rate to 0.001 and then simulate a phylogenetic tree using MASTER (160). We then simulate genetic sequences on top of the phylogenetic trees using different rates of evolution depending on the lineage's compartment. Next, we infer the birth, death, and transition rates from the genetic sequences and show that the 95% highest posterior density (HPD) interval covers the truth in 95% of the 100 runs (see Figure 2.11).

2.3.8 SDevo application to simulated tumors

We applied SDevo to outputs of the Eden and PhysiCell simulations generated as described above. For each simulated tumor we calculated clock rate (*mutations/tree length/sequence length*) and edge and center sampling rates (*sampled / alive cells*). We set exponential priors on birth, death, and transition rates. Full parameterization can be reproduced from XML templates. MCMC chains were run to convergence. We used only chains that had a minimum effective sample size (ESS) for birth rate parameters greater than 200 for analysis. We also excluded rare ($n < 5$) cases that converged to local optima. We summarized the output posterior distributions by mean and 95% HPD intervals. We further inferred maximum clade credibility (MCC) trees with median heights using BEAST 2.6.2 TreeAnnotator (121). TreeAnnotator also gives posterior state probabilities for each MCC internal node.

2.3.9 SDevo application to hepatocellular carcinoma tumors

To apply SDevo to the hepatocellular carcinoma data, we labeled punches based on edge/center state labels as published by Li et al. (2021) (34), Table S8 (reproduced in Fig-

ures 2.9A and 2.9E). For alternate states (Figures 2.16A and 2.16E), we labeled punches as edge if they were located within approximately 10% (∼2mm for Tumor 1 and ∼1.5mm for Tumor 2) of the tumor diameter from the schematic boundaries. Slices were reported to be from tumor hemispheres. Assuming a 0.2mm slice thickness, we estimated that slices Tumor 1Z and Tumor 2Z fell within the boundary region. The original amplicon genotyping panel artificially increases the apparent diversity within some clones relative to others, so to avoid incorporating this bias into the model, we used only whole-genome sequenced punches. Li et al. (2021) identified a large number of SNVs (254,268 for Tumor 1 and 142,032 for Tumor 2). To reduce computational requirements and improve convergence, we generated input pseudo-sequences by randomly subsampling 25,000 variable sites. We summarized results across three independent subsamples for each tumor. We called presence or absence of a variant at each site based on a VAF cutoff of 0.05. Variant allele frequency histograms displayed single-peaked distributions characteristic of a single major clone per sample, with the exception of tumor sample T1L13 (Figure 2.18). To ensure Tumor 1 results were not driven by over-counting mutations across multiple subclones of T1L13, we repeated the analysis excluding this sample and found quantitatively similar results (Figure 2.17).

We use a GTR+Γ_4 site model, a fixed clock rate of 0.3 (units are arbitrary as we only use sites which are variable relative to healthy cells), and estimate sampling proportion (uniform prior). We use log-normal priors for birth (mean=20, S=0.5) and death rates (mean=15, S=0.5). We used an exponential prior for the edge-to-center transition rate (mean=1). Note that these units are also arbitrary and are not calibrated to clinical time. In applying SDevo to these tumors, we constrain the parameter space in several ways to adapt to having relatively few samples, only a single observed time point, and unknown sampling proportion. 1) We assume unidirectional transition so that cells can only move from edge to center but not vice versa. As we only have a few observed state transition events, the transition rates would otherwise be relatively poorly informed. 2) We set priors on mean birth, death,

and transition rates across the two states. Birth and death priors are identical across both states, while transition rates priors are asymmetrical to inform unidirectional transition and enable convergence in a complex parameter space. Full parameterization can be found in the XML template. We combined posterior estimates across three independent runs for each tumor. We inferred MCC trees with ancestral state reconstructions with TreeAnnonator. In addition to the SDevo-inferred trees and parameters, we also generated maximum likelihood trees using FastTree (161) and Augur (162) under a Jukes-Cantor model for each tumor using all reported variable sites. Homoplastic sites contributed to lower support for one node in the maximum likelihood tree of Tumor 1 (Figure 2.9B) and we masked homoplastic sites to enable convergence in Tumor 1 and Tumor 2 SDevo inferences. Homoplastic sites represented $< 1\%$ (Tumor 1) or 6% (Tumor 2) of all sites across all tumor samples. In Tumor 2, more than 2/3rds of homoplasies were between two edge-associated punches (Z1 and Z13) potentially pointing to subclonal mixing, which is supported by their proximal spatial locations. The remainder of homoplasies in Tumor 2 and all of the homoplasies in T1 were evenly distributed across punches. As a result, the removal of homoplasies did not act to bias branch lengths across the tree, with the exception of T2Z1 and T2Z13. As these punches are on the edge of the tumor, this masking should a priori result in lower estimated birth rates on the edge and thus conservatively bias the results towards a more equal birth rate between edge and center.

2.3.10 Data availability

Data required to reproduce analyses are available at https://github.com/blab/spatial-tumor-phylodynamics, including variant allele frequencies and input BEAST2 XML files. Raw sequencing data are publicly available (GSA-Human: HRA000188) as published by Li et al. (2021) (34). SNVs used for the hepatocellular carcinoma analysis are provided in a de-identified format on GitHub. Please cite Li et al. (2021) if using these data.

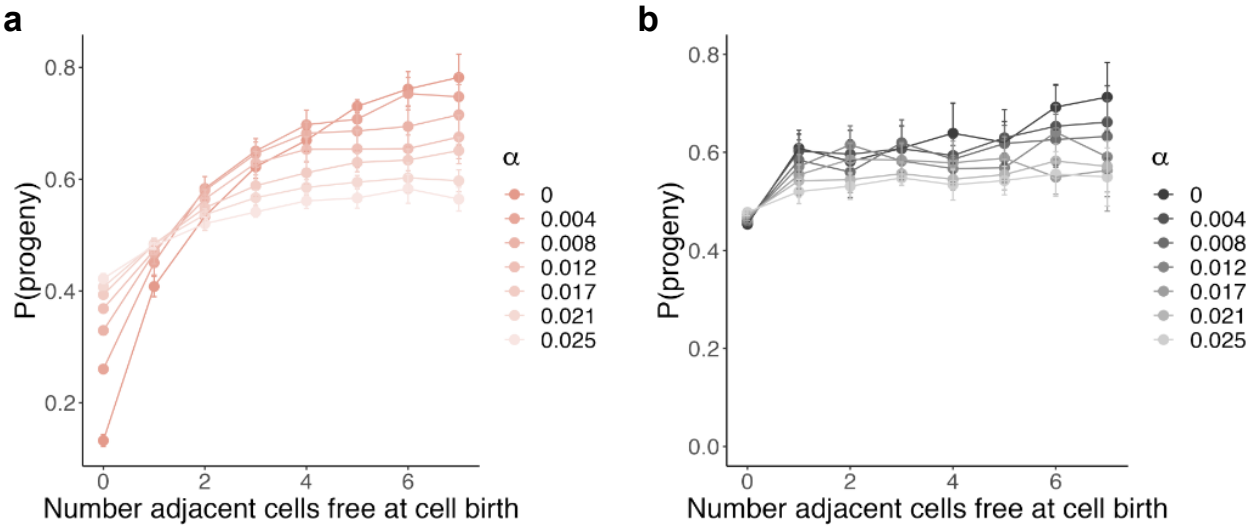

Figure 2.10: Cellular density creates fitness differences in expanding lattice-based simulations. **A.** Fitness, here approximated by the probability a cell has a daughter cell in the population (P(progeny)) versus the number of adjacent free cells at birth under boundary-driven growth. Spatial impacts on cell fitness are relaxed with increasing cell death rate α (color tint). Means and standard error bars are summarized across 10 simulated tumors per death rate. **B.** Under unrestricted growth, most cells are born into a dense neighborhood (free cells = 0), but fitness is not impacted by spatial location. Means and standard error bars are summarized across 10 simulated tumors per death rate.

2.3.11 Code availability

Custom scripts were used for simulation studies and data analyses. All code to generate figures is publicly available. Scripts to replicate analyses and figures are available at https://github.com/blab/spatial-tumor-phylodynamics, including a local R package *tumortree* (https://github.com/blab/spatial-tumor-phylodynamics/tree/main/tumortree), which can be installed to build trees from the simulation outputs. SDevo is built as a BEAST2 package and can be installed from https://github.com/nicfel/SDevo. The source code to run spatially-constrained PhysiCell simulations and generate trees can be found here: https://github.com/federlab/PhysiCellTrees. All other packages used for analysis and visualization are also open source (163, 164).

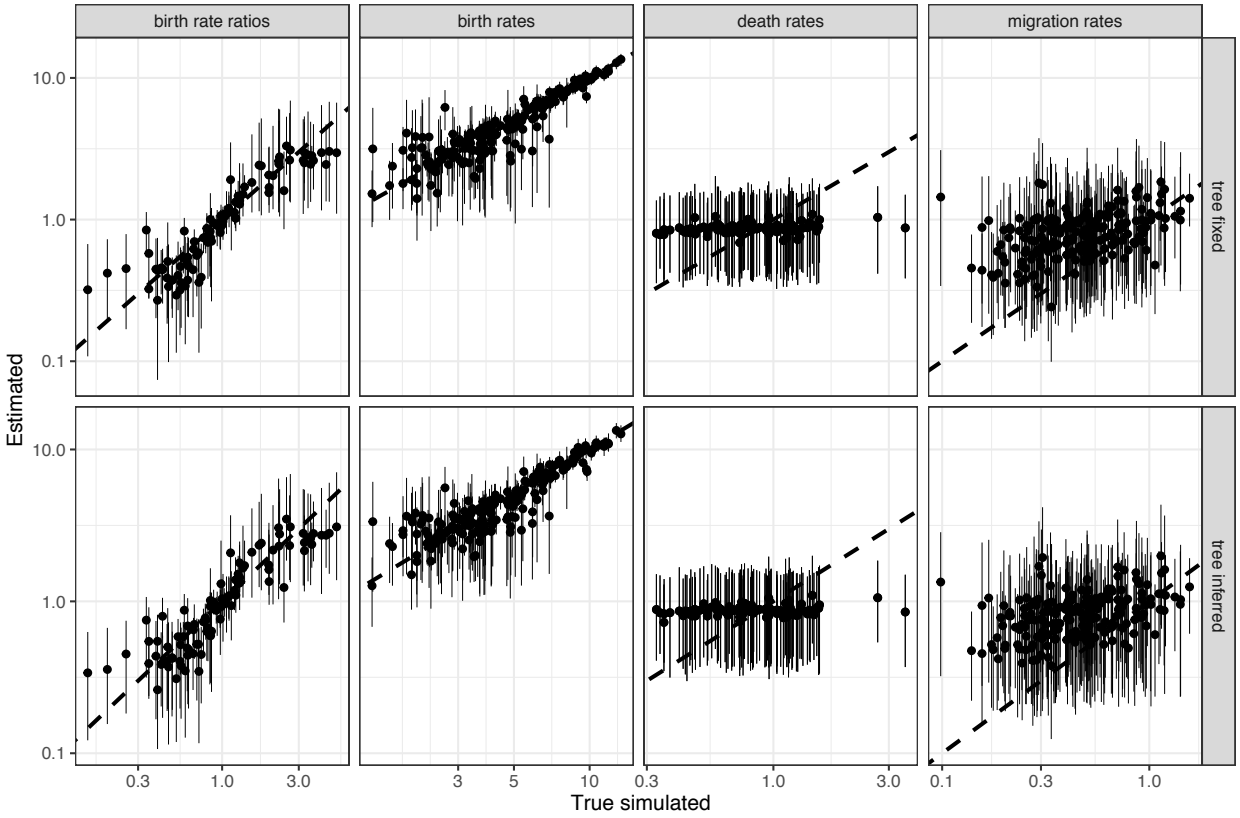

Figure 2.11: Simulation study to validate SDevo implementation. Birth, death, and transition rates, and ratios of state-dependent birth rates estimated by SDevo versus true population parameters of phylogenetic trees simulated under a two-state birth-death model (see Methods). Medians (points) and 95% HPD intervals (bars) of estimated values are plotted for each parameter (columns) while either fixing or jointly inferring the tree topology (rows). Trees were simulated under a birth-death with migration model with rates drawn from log-normal distributions for birth (mean=5, S=0.5), death (mean=1, S=0.5), and migration (mean=0.5, S=0.5). The sampling proportions were also drawn from a log-normal distribution (mean=0.15, S=0.5). This process leads to trees with between 32 and 72 tips.

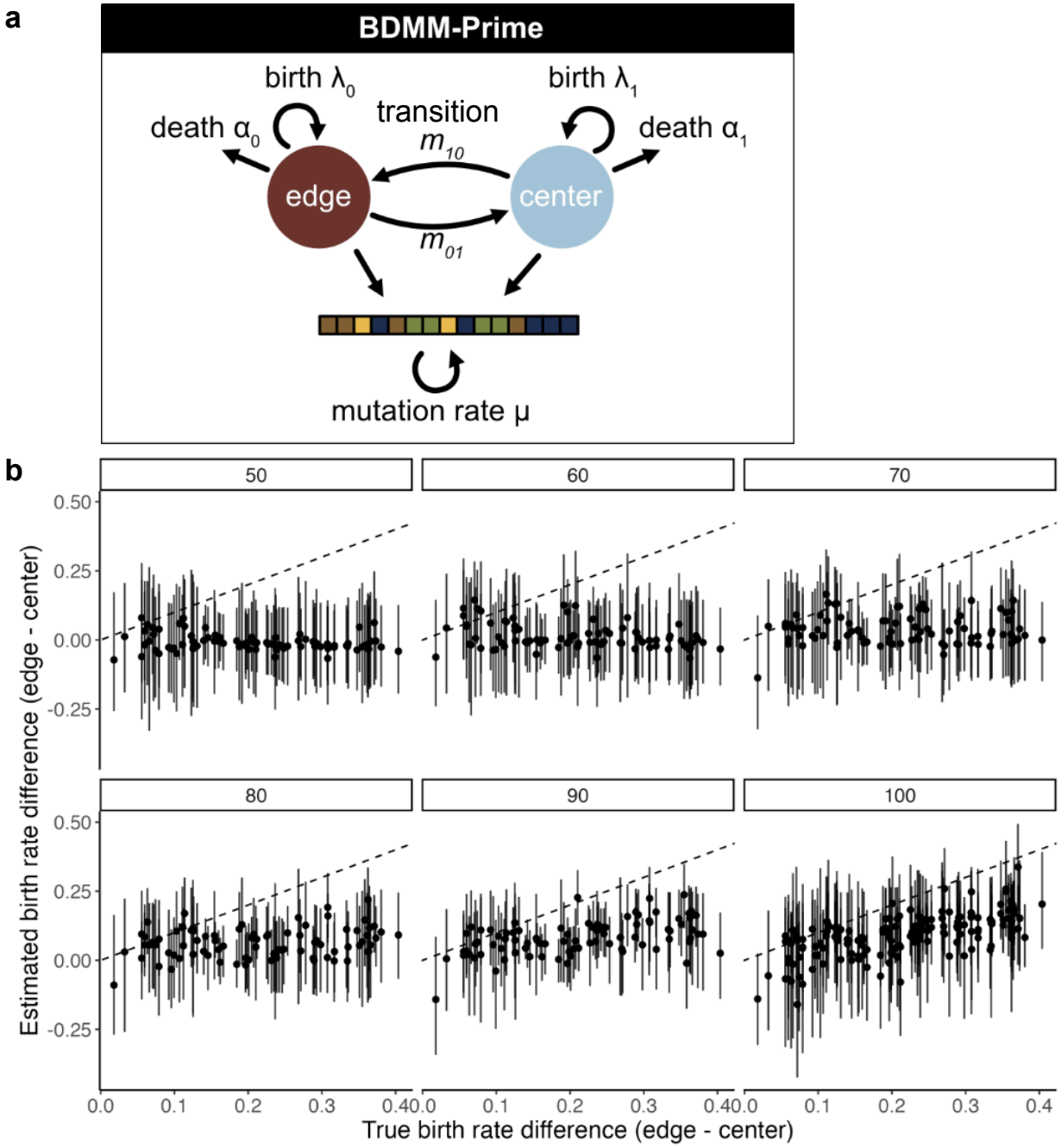

Figure 2.12: Multi-state diversification models without state-dependent clocks do not sensitively detect growth rate differences in simulated tumors. **A.** Schematic of BDMM-Prime, which does not link state-dependent effects on division to sequence evolution. **B.** True versus estimated means (points) and 95% HPD intervals (bars) of birth rate (per day) differences between the edge and center of simulated boundary-driven tumors over a range of sample sizes ($n = 50, 60, 70, 80, 90, 100$, see headers). Dashed line is $y = x$.

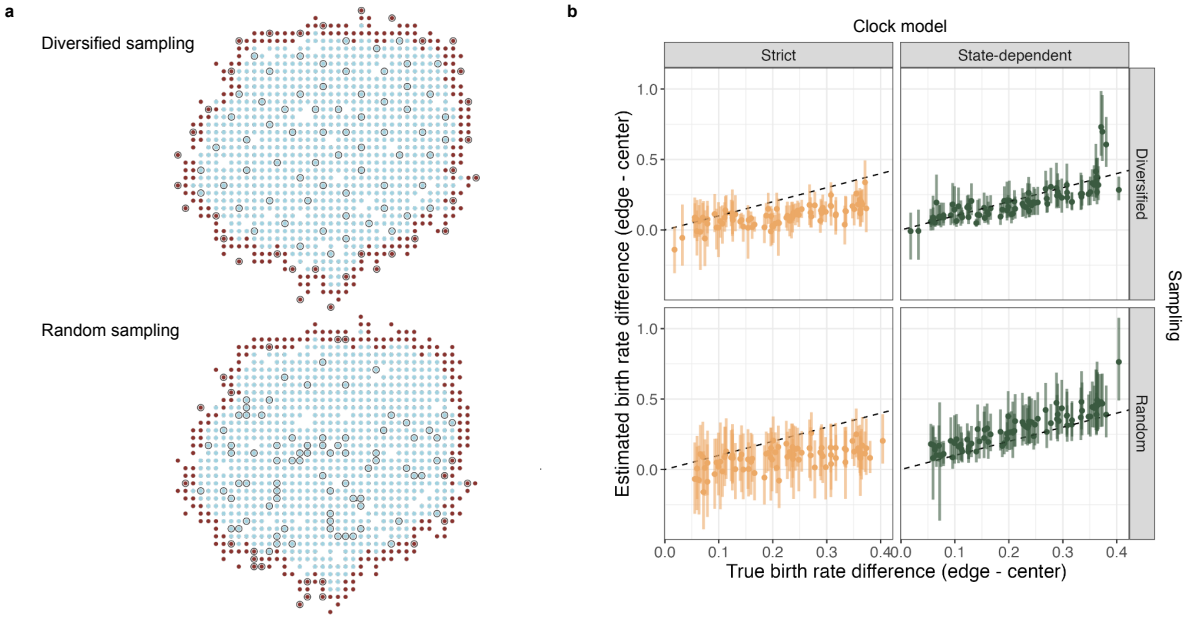

Figure 2.13: SDevo improves birth rate estimation with more variable (random) sampling over a strict clock model. A. Example 2D tumors under either diversified or random sampling schemes. Cells are colored by edge (maroon) or center (blue). Grey-highlighted cells are sampled. Diversified sampling maximizes the physical distance between sampled cells. **B.** Estimated means (points) and 95% HPD intervals (bars) of per day birth rate differences between the edge and center of simulated boundary-driven tumors based on $n = 100$ sampled cells versus true state-dependent effects (α varies between 0 and 0.036). We compare SDevo (green) with strict clock model (gold) for either diversified or random sampling (rows).

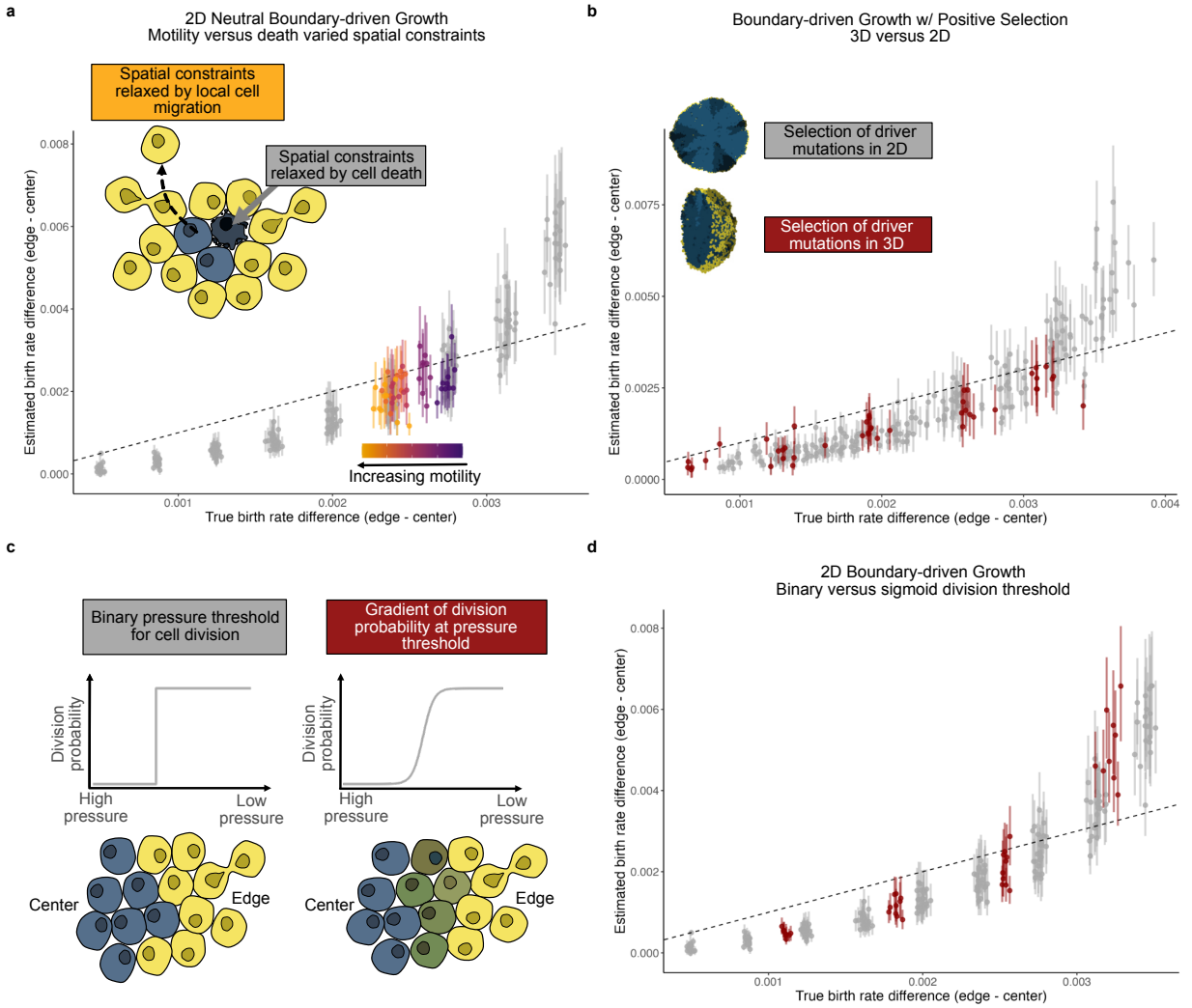

Figure 2.14: Investigation of extended modes of spatial tumor growth in PhysiCell simulations. **A.** Mean (points) and 95% HPD intervals (bars) of birth rate differences estimated by SDevo when spatial constraints are relaxed by increasing cell motility (purple to orange gradient) compared to when spatial constraints are relaxed by cell death (grey).$N = 100$ per tumor for all simulations. The x-axis is the effective true birth rate difference in both scenarios. **B.** True versus SDevo-estimated birth rate differences in simulations with both boundary-driven growth and positive selection of driver mutations ($n = 100$ sampled cells per tumor). We compare simulations in 2D (grey) and 3D (red). **C.** Schematic of simulated relationships between cell pressure and division probability for either a binary (left) or sigmoidal (right) gradient in PhysiCell simulations. D. True versus estimated birth rate differences of simulated tumors ($n = 100$ sampled cells per tumor) with either a binary (grey) or sigmoidal (red) pressure threshold. For **A, B, and D.** points and bars represent the mean and 95% HPD interval of each estimate.

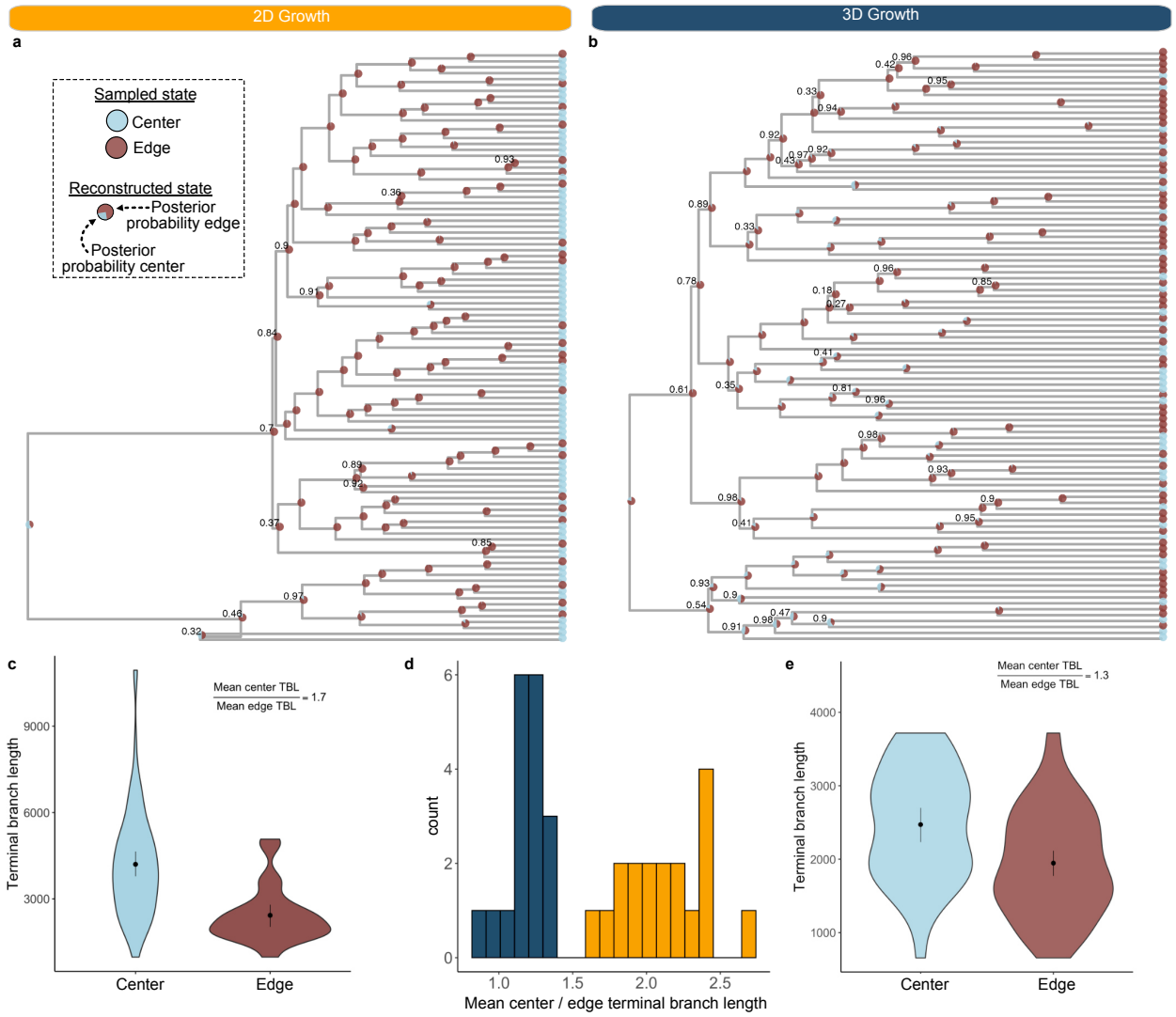

Figure 2.15: Complex growth and sampling in 3D tumors lead to more variable branching patterns. A. Example inferred phylogeny of 2D PhysiCell tumor with reconstructed ancestral edge and center states ($d = 0.1$). Node pie charts represent posterior support for each state. 100 cells were sampled to maximize the distance between cells (diversified sampling). **B.** Example inferred phylogeny of 3D PhysiCell tumor with reconstructed ancestral edge and center states ($d = 0.1$). Cells were sampled to maximize distance in 2D space across z-slices of the simulated tumor as described in Materials & Methods. For both trees, posterior node support is indicated if less than 99%. **C, E.** Comparisons of inferred terminal branch lengths between cells sampled on the edge and center of the example 2D and 3D tumors. Points and error bars represent means and bootstrapped confidence intervals for $n = 75$ center and $n = 25$ edge terminal branches. **D.** Distribution of the relative ratio of center-to-edge mean terminal branch lengths across multiple simulations with equivalent spatial constraints. Asymmetric branching between edge and center states is observed more often in 2D (gold) than 3D (navy) tumors. Points and error bars represent means and bootstrapped confidence intervals for $n = 38$ center and $n = 62$ edge terminal branches, respectively.

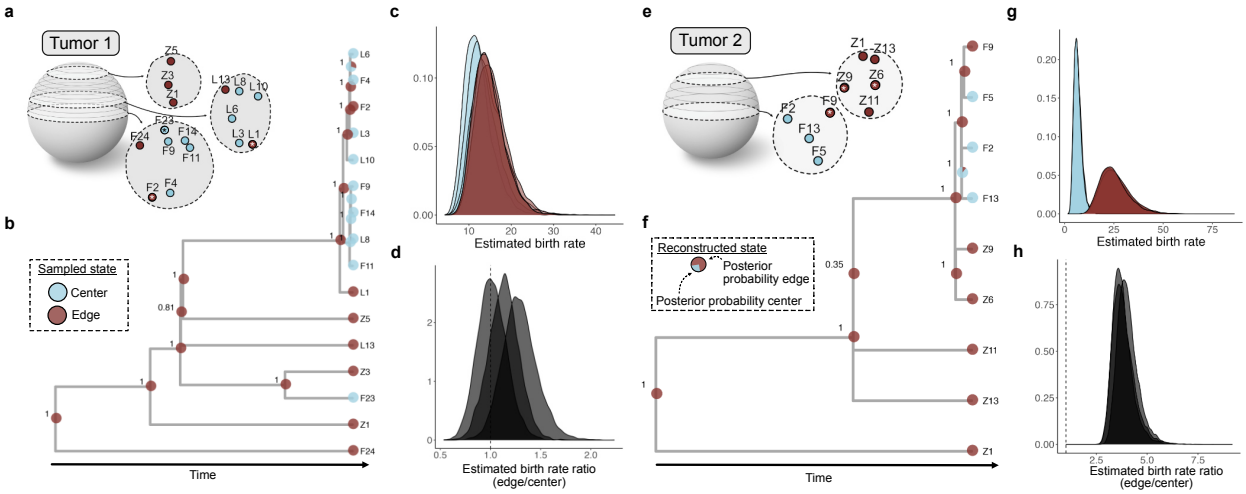

Figure 2.16: Detection of boundary-driven growth in hepatocellular carcinoma with variation in edge/center state calling. We called an alternate set of states based on a distance of $< 10\%$ of each tumor diameter from the edge instead of published edge/center labels. **A.** Multi-region sampling map for Tumor 1 adapted from Li et al., (2021) (40) with alternate state labels. Asterisk indicates a change in state from published states (Figure 5A). **B.** Inferred tumor phylogeny and reconstructed ancestral spatial states for a single SNV subset. Clade posterior supports are indicated at nodes. **C.** Marginal posterior distributions for edge (maroon) and center (blue) birth rates estimated from the Tumor 1 WGS data across three independent SNV subsets. **D.** Posterior distribution of edge/center birth rate ratio. Dashed line indicates ratio of 1. We estimate a mean 1.15x higher birth rate on the edge compared to center. **E.** Multi-region sampling map with alternate states for Tumor 2 reproduced from Li et al., (2021) (40). **F.** Tumor 2 tree and ancestral edge/center states inferred from the sampled populations. **G.** Marginal posterior distributions for edge and center estimated birth rates and **H.** edge/center ratio. We estimate a mean 3.89x higher birth rate on the edge versus center based on the alternate state calls.

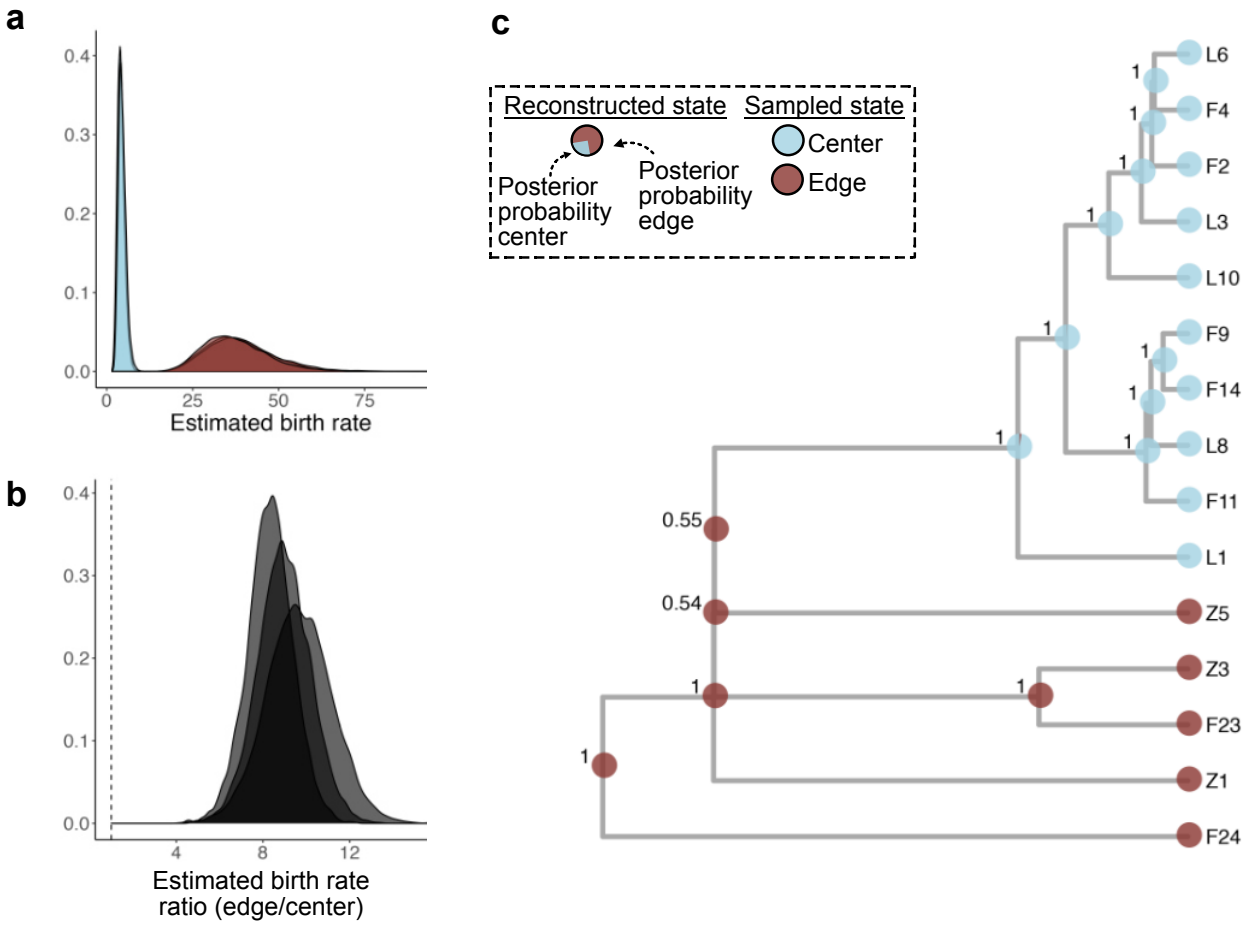

Figure 2.17: SDevo analysis excluding heterogeneous punch T1L13 also estimates birth rate differences between center and edge samples. To ensure our results are not driven by an edge-associated sample (T1L13), which potentially contains multiple subclones, we repeated the analysis of Tumor 1 without this punch. **A.** Marginal posterior distributions for edge (maroon) and center (blue) birth rates estimated from the Tumor 1 WGS data excluding T1L13 inferred for three SNV subsets. **B.** We estimate a 8.99x higher birth rate on the edge compared to center (95% HPD 6.37-11.91x, summarized across the same three independent inferences). **C.** Inferred MCC tumor phylogeny and reconstructed ancestral spatial states. Posterior clade support is indicated at each node.

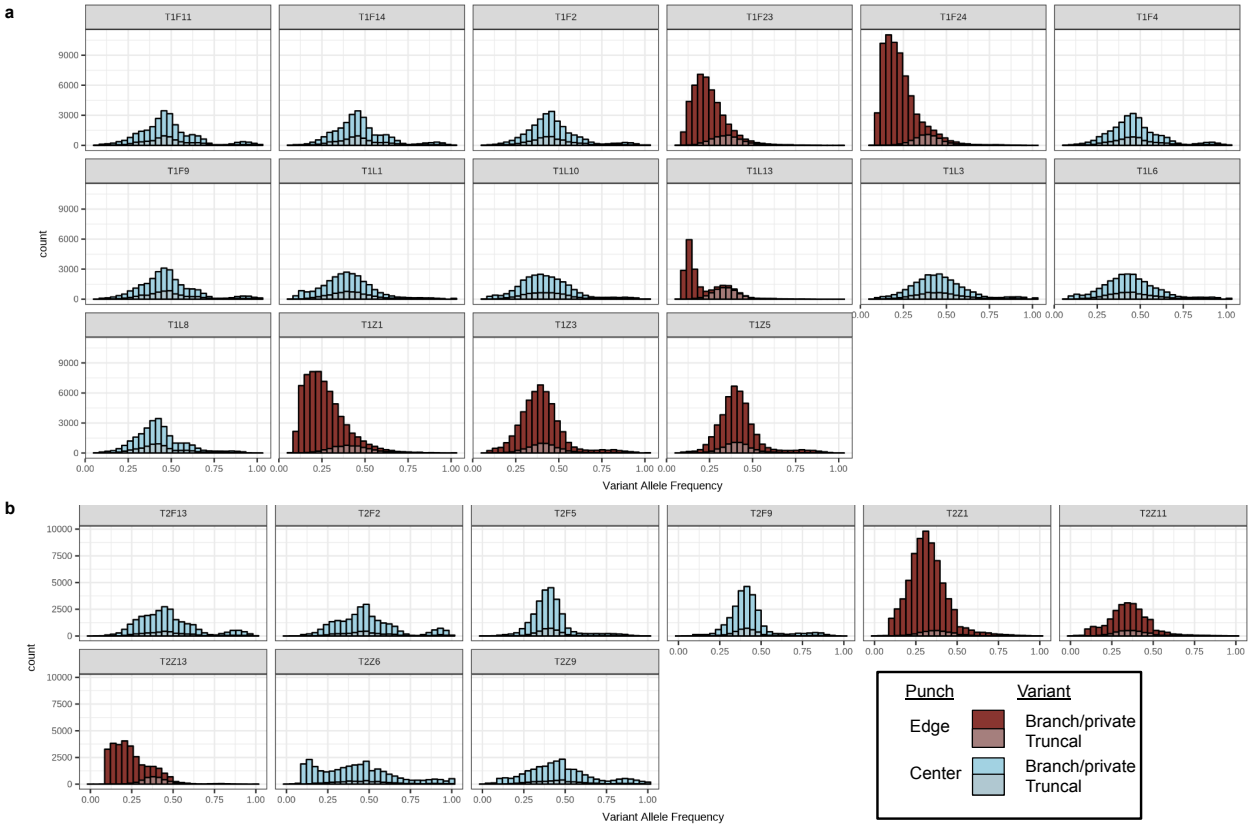

Figure 2.18: Variant allele frequency (VAF) histograms reveal punches are largely clonal. Variant allele frequencies for all non-truncal (opaque) and truncal (transparent) mutations observed in tumor punches from Tumor 1 (**A**) and Tumor 2 (**B**) reveal that punches contain only a single high-frequency clone, with the exception of T1L13. Punches are colored by their edge (maroon) or center (blue) status. State labels correspond to Li et al., (2021), Table S8.

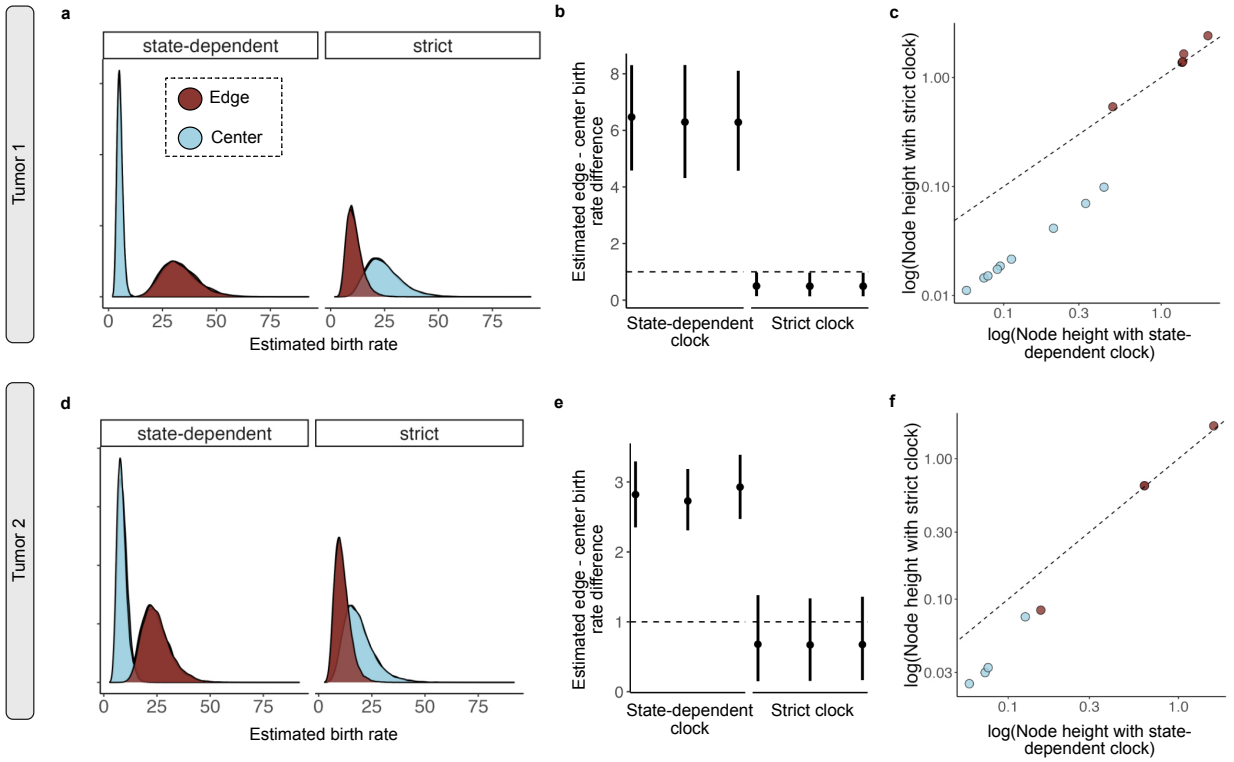

Figure 2.19: SDevo infers boundary-driven growth in HCC tumors where a strict clock fails and changes inferred node timings. We compared estimates of birth rate differences between edge and center under a state-dependent birth-death model (BDMM') using both our novel state-linked sequence evolution model or a strict clock (state-independent) sequence evolution model. For Tumor 1 **(A)**, Tumor 2 **(D)**, posteriors of edge and center birth rate estimates for each sequence evolution model are shown in maroon and blue, respectively. Means and 95% HPD intervals (points and bars) for the inferred birth rate ratios between edge and center states for Tumor 1 **(B)** and Tumor 2 **(E)**. Posteriors are inferred across three independent SNV subsets. Dashed lines indicate ratio of 1. Note, power analyses on simulated tumors (Figure 3 and 2.12) suggest that the strict clock model should be under-powered and sensitive to sampling variation at these sample sizes (Tumor 1: $n = 16$, Tumor 2: $n = 9$). **C. and F.** Scatterplots show ancestral node heights inferred under strict clock versus heights inferred by SDevo colored by most probable ancestral state. Nodes are compared based on matching a subset of tips.

Chapter

RECONSTRUCTING SPATIOTEMPORAL DYNAMICS OF PRIMARY PROSTATE CANCER

3.1 Introduction

Prostate cancer (PCa) is the second leading cause of death for males in the United States and has a high prevalence, as 1 in 9 will be diagnosed with prostate cancer in their lifetime (165). Yet, the clinical progression of the disease is highly variable. Where many patients may have slow-growing tumors over decades that never metastasize, others have highly aggressive disease that is fatal within a couple years (166). There is an urgent need to understand the underlying determinants of prostate tumor growth and clinical phenotype to determine why some tumors progress and others do not.

Evaluating and predicting the clinical course of prostate cancer has long been complicated by its heterogeneous nature (167–169). Like many other solid tumors, prostate cancer exhibits spatial genetic diversity (170, 171, 63, 172–175), and although intra-tumor variability is still considered to be lower as compared to inter-patient variability (176), it likely still has important clinical consequences. At the time of diagnosis, patients frequently have multiple foci within the same prostate with different, non-overlapping driver mutations or structural rearrangements (177, 178, 172, 175), and subclonal variation within the same loci (170, 171). Multi-clonality of driver mutations is associated with an increased rate of clinical progression (179) and multiple evolutionary approaches have observed that expansion of subclonal lineages occurs prior to metastasis(7, 65). There is further evidence that this heterogeneous primary tumor population likely undergoes further selection for cells that metastasize or develop drug resistance. For example, Woodcock and colleagues estimated that two primary

lineages co-existed in space and time in a single PCa tumor, but, only one of these lineages led to a detectable metastasis (7). These observations highlight the value of elucidating the mechanisms that give rise to ITH and determine subclonal fate. Recent successes in incorporating evolutionary principles to delay drug resistance in prostate cancer treatment have further emphasized the need to understand evolutionary mechanisms that underpin the development of ITH in order to understand cancer progression (180, 181)

Despite the pervasiveness of ITH in clinical disease, mechanisms that give rise to spatial ITH in primary prostate cancer remain largely unknown. From studies in other cancer types, it is likely many evolutionary factors and processes could contribute to spatial ITH (discussed in Chapter 1). To ultimately understand and predict PCa progression, cancer evolution must be understood within the context of both genetic and spatial heterogeneity. Existing major obstacles to understanding the interaction between these two factors include that; (1) PCa grows unobserved for years before detection, (2) current sampling is insufficient to capture an accurate picture of genetic population structure and, (3) studies linking genetic and RNA expression often do not account for spatial growth dynamics. Furthermore, while many studies have approached this question by analyzing a large number of tumors (44, 7, 88), they summarize effects across many different environmental contexts.

In Chapters 3 and 4 we aim to characterize underlying factors contributing to spatial, genetic, and transcriptomic heterogeneity through extensive multi-region analysis of a single prostate tumor with detailed spatial resolution. In this chapter, we apply coalescent and continuous trait evolution phylodynamic models to reconstruct the biogeographic history of the tumor and demonstrate that sampled genetic structure can be explained by early divergence and diffusion of subclonal populations across tumor space.

3.2 Results

In order to sample intra-tumor heterogeneity, we undertook multi-region punch-biopsy sampling as described in Figure 3.1A across two transverse sections of a cryo-preserved primary prostate tumor. This tumor was chosen for morphological diversity (mixed grade) and sufficient spatial distribution. For each punch, we carried out whole-exome sequencing (WES) and RNAseq on the bulk population of tumor cells (Figure 3.5).

3.2.1 Characterization of intra-tumor spatial genetic heterogeneity

From the WES data, we characterized somatic copy number aberrations (SCNAs) and called somatic SNVs (Figure 3.1B and C) based on the sequences of adjacent normal glands. As is common in prostate cancer (171), we inferred extensive SCNAs. Several structural variants identified in this tumor are known to be recurrently mutated in prostate cancer, including deletions in 8p, 13q, 1p, 22, and 10q, and amplification of 8q (182). High-level amplification of the 8q chromosomal arm is especially notable as it contains *c-myc* (8q24). We also noted that homozygous deletions of 18q and 22q would lead to loss of tumor suppressors SMAD4 (183) and CHEK2, respectively. Other potentially affected oncogenes or tumor suppressors include RB1 (13q heterozygous deletion) (184), DCC (18q homozygous deletion) (185). Finally, we inferred an amplification in 11q13, which also commonly occurs in cancer (186). Notably, all detected alterations were universally present in all biopsies, indicating that they arose early in tumor development and supporting a clonal origin of the sampled regions as opposed to multi-focal disease (63). Taken together, our analysis of SCNAs suggests early loss of tumor suppressors paired with the amplification of C-MYC and 11q13 due to large-scale copy number variation are plausible drivers of disease (187). In addition, extensive chromosomal arm aneuploidy may be clinically relevant as it is implicated in contributing to more aggressive disease (188).

Although subclonal SCNA was not detected, SNV inference revealed that punch biop-

sies were comprised of subclonal genetic populations as indicated by SNVs with lower allele frequencies (Figure 3.7). In this analysis, we identified a truncal, nonsense ATM mutation (S478*), which is especially notable as we also predicted a loss of heterozygosity of part of 11q where ATM is located. We further detected truncal non-synonymous variants in GPR158 (p.T812M), RAB12 (p.Y218C), TST (p.V271M), CD1D (p.M1I), and AHRR (p.I134M). Other shared variants with predicted coding consequences include, PROSER3 (p.P187S), CYFIP1 (p.N829S), CDH1 (p.154fs), ATP12A (p.V479M), LEXM (p.R117W), RNF112 (p.A218T), SRSF5 (p.R95L), UPF1 (p.A201G) and TRADD (p.D24H). Although some of these genes have known roles in cancer (189–193), we expect that most SNVs identified are not driver mutations, but instead are used as markers of genetic lineages within the tumor. Therefore, we can use variability in these lineages within a population to describe intra-tumor genetic variability (Figure 3.1D).

We next aimed to describe the tumor population structure through 1) the deconvolution of bulk populations into unique genetic subclones, and 2) mapping their spatial distribution. We applied PyClone-VI (194) and Pairtree (195), computational tools that cluster SNVs based on covariance within the sampled populations and infer a tumor clone tree, respectively. We observe that most genetic subclones are present across the sampled punches but with widely varying population frequencies (Figure 3.2A-B). By mapping estimated subclonal fractions to each sampled punch location, we observed clustering of population genetic structure in space (Figure 3.2C) where nearby punches generally were more likely to be close genetically (Figure 3.2D).

3.2.2 Early genetic and spatial diversification of subclones explain population structure

We investigated the underlying dynamics that gave rise to the sampled population structure through phylodynamic modeling. Phylodynamic models incorporate sequence mutational processes and population growth to explain the sampled genetic structure. Here, we apply

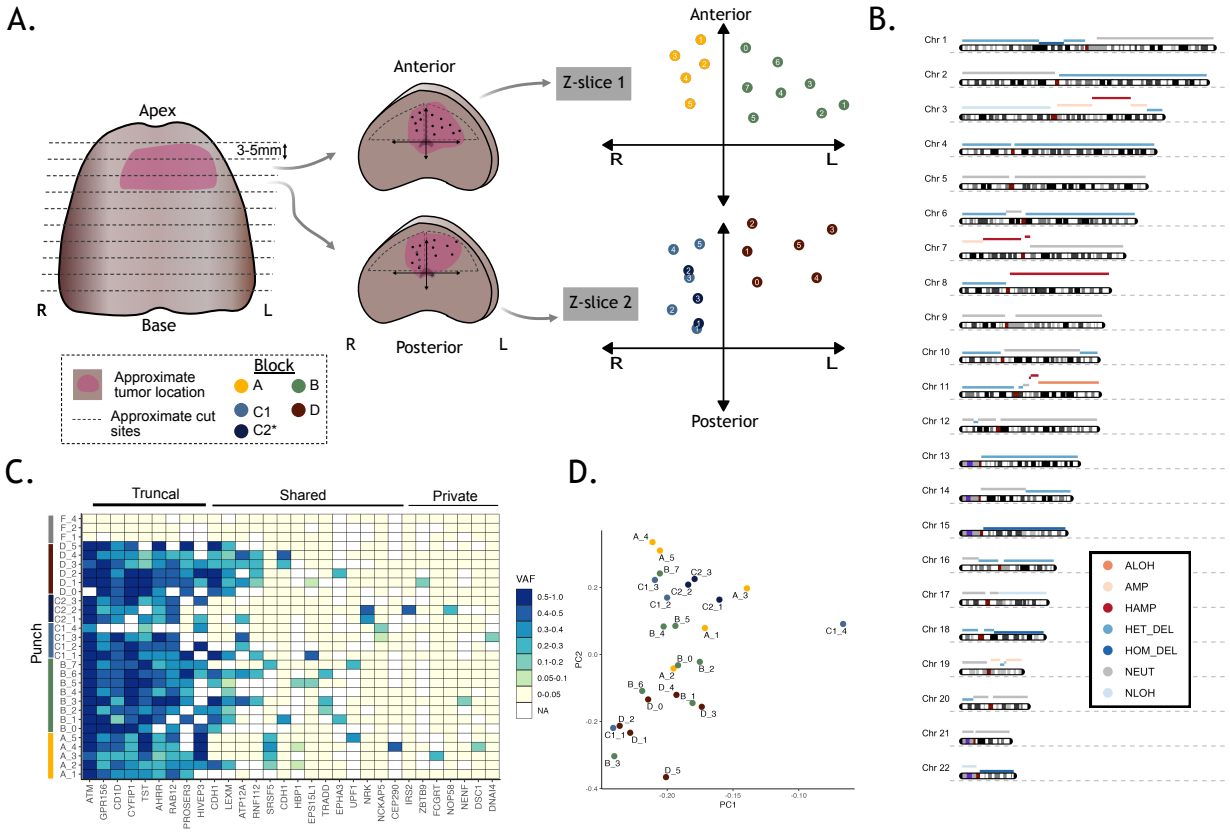

Figure 3.1: Spatial sampling scheme and genomic analysis of spatial heterogeneity. **A.** 2mm punch biopsies were collected from cyro-preserved transverse sections. Pink shading indicates approximate tumor extent. Dotted lines indicate approximate cuts for cryo-preserved blocks. Points mark locations of punch biopsies that generate a spatial map in 3D space reconstructed from H&E histology and punched blocks. **B.** Inferred global somatic copy-number aberrations (SCNAs) mapped across the autosomal chromosomes. Y-axis indicates the total copy number for each segment colored by inferred copy-number state (ALOH = amplification with LOH, AMP = amplification, HAMP = high amplification, HETDEL = heterozygous deletion, HOMDEL = homozygous deletion, NEUT = normal copy number, NLOH = normal copy number with LOH). **C.** High-confidence non-synonymous coding variants across whole-exome sequenced punch biopsies. Variants are ordered by the number of samples in which they are found and the mean variant allele frequency. **D.** PCA of VAFs called in bulk WES libraries of punch biopsies.

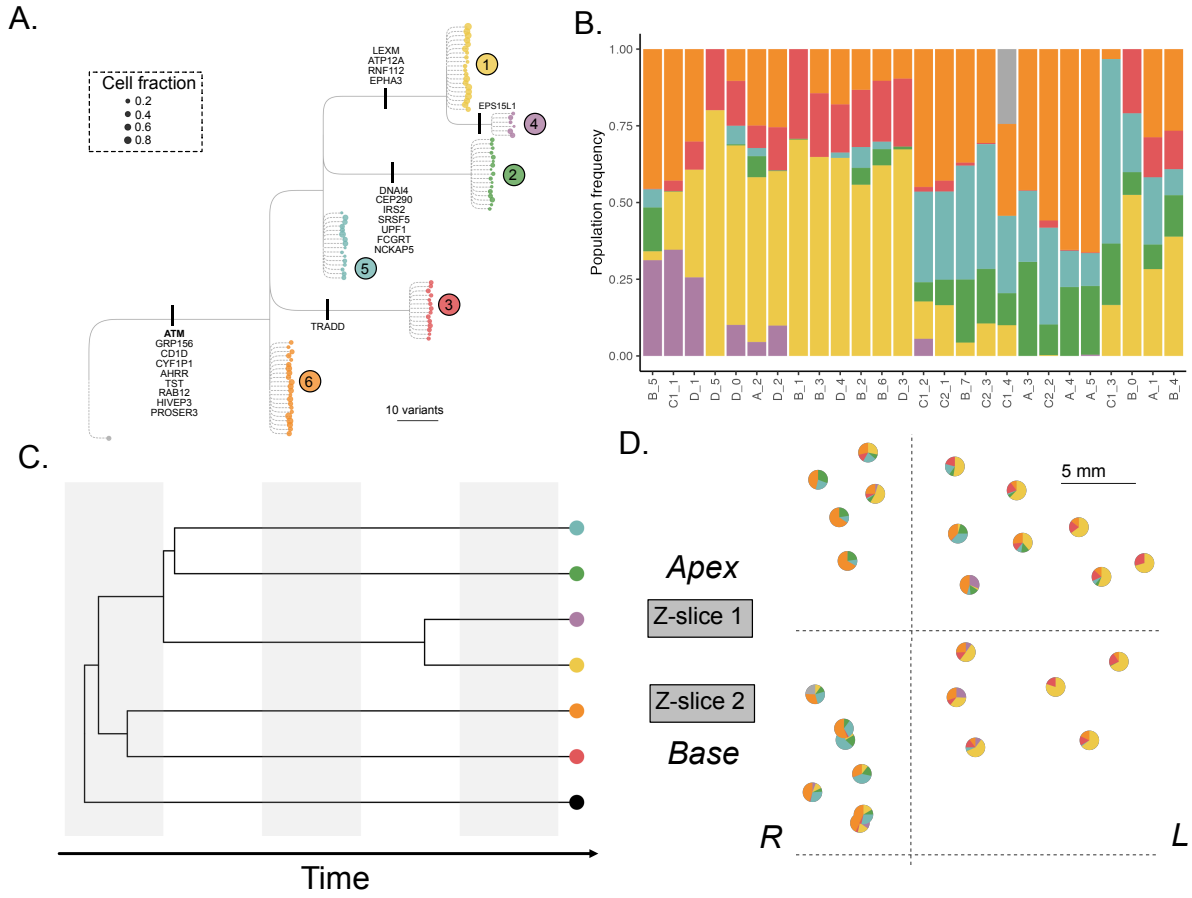

Figure 3.2: Subclonal genetic population structure. **A.** Inferred genetic clonal tree with all non-synonymous SNVs labeled on branches. Each tip indicates a different punch and size correlates with estimated cellular fraction. Color delineates a unique subclone and branch lengths correspond to the number of genetic mutations. **B.** Per punch population fraction of each subclone within each punch ordered using hierarchical clustering of punches. **C.** Distribution of genetic subclones in XY space. Each plot corresponds to a Z-plane (transverse sections). Each pie chart corresponds to an individual punch and represents the subclonal distribution within that population. **D.** Pairwise genetic versus spatial Euclidean distance.

two separate models to reconstruct 1) genetic divergence and 2) the spatial location of subclonal populations over time. First, to estimate the timings of subclonal diversification, we treated each subclonal population identified in Figure 3.2 as a single "species" with an associated psuedo-sequence created from identified SNVs (Figure 3.2A). We inferred a maximum clade credibility (MCC) tree with an exponential coalescent model (Figure 3.4A) and estimated that the most recent common ancestors (MRCAs) of subclones existed early in tumor growth. We note that although this first model does not account for spatial structure, the only subclone estimated to have diverged later in tumor evolution also has smaller spatial distribution (yellow subclone 4, Figure 3.2C).

Adaptation of continuous trait evolution phylodynamic model to investigate spatial dynamics

Next, we aimed to elucidate spatial dynamics by directly modeling cell migration as a continuous diffusion process. Specifically, we applied a Brownian diffusion or "random walk" model of continuous trait evolution (196), where the trait in this analysis is the XYZ coordinates of each punch biopsy. Similar models have been widely applied in other fields to infer biogeographic spread (197, 57), but have never been applied to cancer biology. These models are classically applied when each sampled tip on a tree has a trait measurement. However, subclones identified in this case are distributed across multiple locations and therefore a single coordinate cannot be assigned to each subclone. In order to incorporate spatial and genetic distribution, we create an augmented set of sampled tips, which we term "pseudo-cells". To create these tips, we assign 15 "cells" per punch location and distribute them to genetic subclones based on the estimated cellular fractions (Figure 3.3). From this analysis, we estimate both an MCC tree where the subclonal proportions and spatial distribution have been taken into account (Figure 3.4A) and can estimate the locations of the subclonal MRCAs in space (Figure 3.4B) and time (Figure 3.4C). From these results, we conclude that spatial and genetic divergence co-occurred in the tumor's clinical history, further sup-

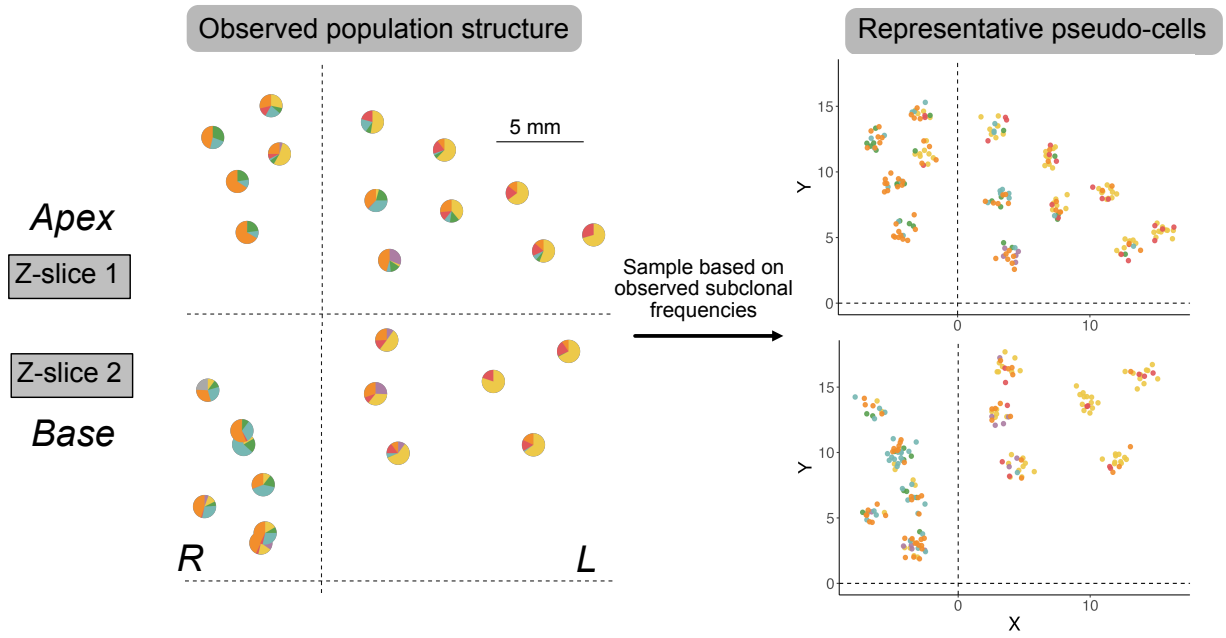

Figure 3.3: Data augmentation of sampled genetic and spatial diversity. The continuous trait evolution requires each input sequence to have a unique associated XYZ value. Instead, each identified genetic subclone is distributed across multiple punch biospies. To create the appropriate input dataset, we generated "pseudo-cells" that were assigned a subclonal identify and XYZ location. Psuedo-cells were distributed based on inferred subclonal frequencies in each punch population. We also added a small amount of random variation in the coordinates to prevent the exact same location from being used many times (something that is very unlikely in a continuous trait model).

porting spatially structured growth. Interestingly, when spatial information is included in the biogeographic model, the estimated timings common ancestors of subclones 2, 3, and 4 were much more similar than in the exponential coalescent model (Figure 3.2C, potentially indicating a discordance between the rates of spatial and genetic evolution.

3.3 Discussion

Here, we demonstrate that although most identified SCNAs and somatic variants are found across all spatially-sampled regions, indicating a common origin for all clones, we also identified further genetic ITH. With a high number of multi-region samples, we were able to undercover spatial population structures across tumor space that we tracked through the varying prevalence of genetic subclones. Through a novel application of biogeographic phylodynamic modeling, we further show that genetic population structure can be explained by an early temporal and spatial divergence.

The "big bang" theory of cancer growth hypothesizes that most ITH is a result of early diversification and spatial expansion (28). Initial inspection of the genetic data and subclonal distribution supports this model given the wide spatial distribution of subclones, yet, we also note that phylodynamic modeling of spatial migration as continuous trait evolution points to an ongoing, local expansion of subclones in the primary tumor. These findings are consistent with other observations of multiple lineages existing in the same spatial location within a primary prostate cancer (7), but we extend these insights by reconstructing the spatial diversification that explains these observations.

We note that although we sequenced at relatively high depth and spatial resolution, additional unsampled genetic diversity could have gone undetected. We also captured heterogeneous genetic populations within each punch biopsy, necessitating clonal deconvolution. To reconstruct subclonal populations, clonal deconvolution methods rely on the co-occurrence of variants across multiple punches. By nature, they are more likely to exclude lower frequency

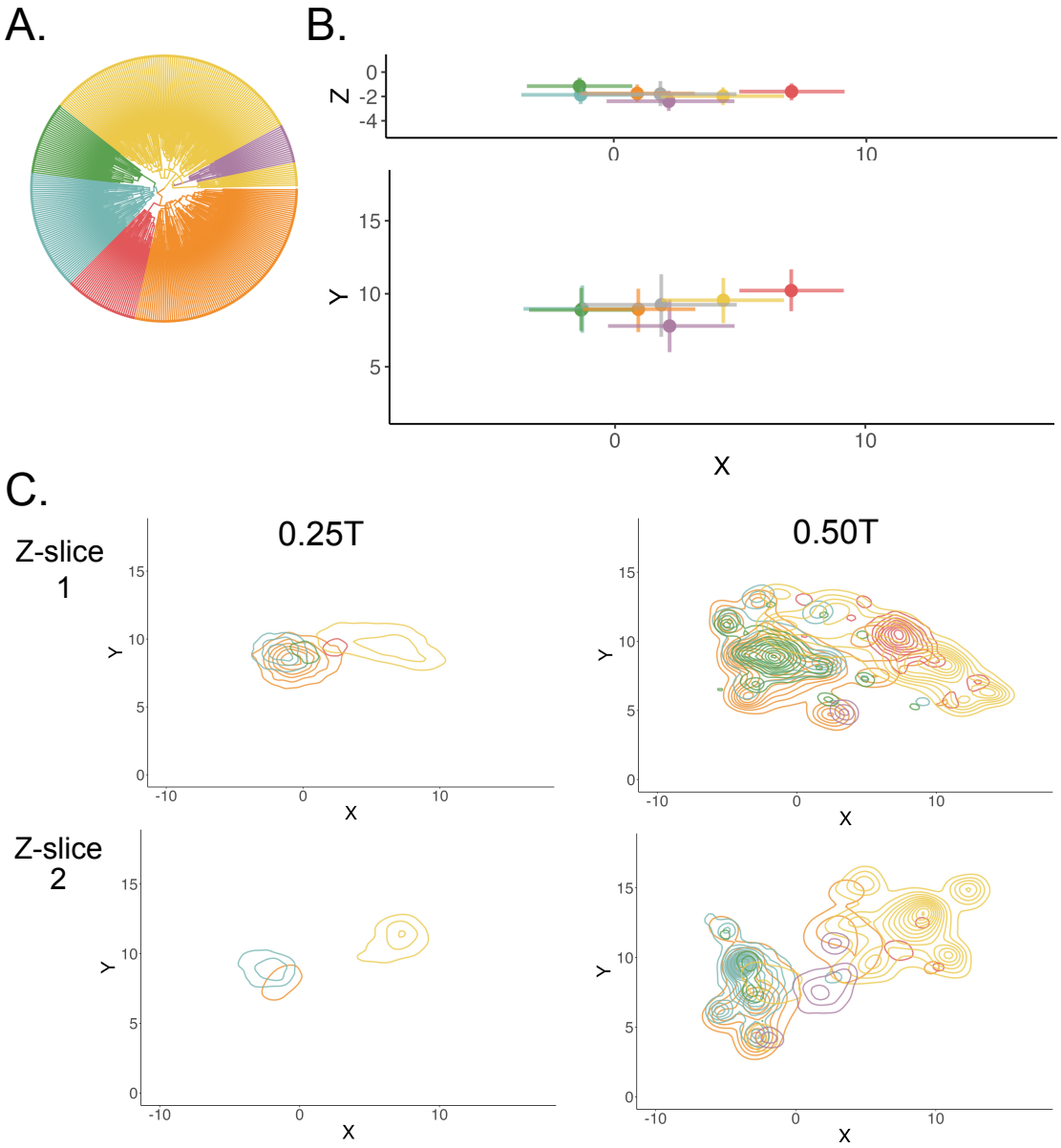

Figure 3.4: Biogeographic modeling reveals early spatial divergence of genetic subclones. A. Reconstructed MCC time tree using augmented tip samples, an exponential coalescent population model, and a Brownian diffusion (random walk) model of continuous trait evolution to infer past spatial locations. Ancestral subclonal identities are also reconstructed. **B.** Means (points) and 95% HPD intervals (bars) for MRCA of each subclone in XY (lower) and XZ (upper) space. **C.** Kernel density estimation plots representing the posterior distribution of reconstructed ancestral locations colored by subclone at 25% and 50% of inferred clinical time (columns) for each Z-slice (rows).

variants and have less power to call private subclones. We, therefore, cannot discount late diversification and ongoing evolution of subclone, as we have less power to observe more recent somatic events.

3.4 Methods

3.4.1 Patient information and specimen sampling

The patient underwent a radical prostatectomy performed via laparoscopy. The specimen was collected by the University of Washington Genitourinary Research lab and processed for research according to the study protocol. Briefly, the prostate was divided into 3-5mm transverse sections. These sections were further divided along the anterior-posterior and left-right axes and either frozen in OCT for research use or formalin-fixed for pathological assessment. The research report indicates focal Gleason 3, extensive Gleason 4 (cribriform/fused glands), and Gleason 5 patterns. Local lymph node dissemination was also noted. H&E histology was used to identify areas of carcinoma in OCT-frozen blocks. 2mm punched biopsies were collected from these regions (26 total punches, see Figure). Additionally, 4 punches were taken from a block with normal, glandular morphology. An approximate tumor map was interpreted from the prostatectomy research assessment form and H&E slides annotated by a pathologist. The tumor diameter was approximated to be 25mm based on the span of carcinoma in sampled frozen blocks (measured in WebDigitizer2). The tumor volume was calculated from this radius and the observation of the tumor spanned 4 transverse 3-5 mm sections ($\tilde{1}6$ mm), giving an estimated volume of 5.2 cc, on par with published volumes of prostate tumors (198). Prostatic cancer cells were assumed to have a diameter of 30 μM (199). From this, a spherical tumor cell volume would be $1.4E-5mm^3$ and the estimated population size is $3.7E8$. Following the same logic, the volume of each $2mm$ punch biopsy is expected to be approximately $12mm^3$, which would correspond to an estimated 900K cells.

3.4.2 Sample processing

Tissue from each punch biopsy was disrupted, homogenized, and lysed in a bead microtube homogenizer. Each lysate was split into two replicates per punch (Figure 3.5). DNA and RNA were extracted from each cell lysate using the Qiagen AllPrep DNA/RNA kit. Punches were processed in batches randomized across blocks. The tumor diameter was approximated to be 25mm based on the span of carcinoma in sampled frozen blocks.

3.4.3 Exome sequencing

Whole-exome sequencing libraries were generated with the KAPA HyperCap Workflow V.3.0, using the HyperPlus and KAPA HyperExome target enrichment probes from 500ng of DNA per replicate. Nanodrop and Tapestation D5000 were used to check library concentration and fragment size distributions pre- and post-probe capture.

3.4.4 Somatic variant calling

Joint variant calling was carried out using GATK Mutect2 (v4.1.8.1. (200)) against the matched "normal" punches. Variants were further filtered by a minimum VAF of 0.05 in at least two libraries and a mean depth of 10 across all libraries. Variants with an alternate allele count of at least 5 were considered present in an individual punch. Variants were only considered absent in samples with no variant reads and with a depth of at least 10 reads. Variants clustered with an indel variant (within 10 basepairs) were counted as a single variant. For high-confidence variant calls, the variant must be present in at least two libraries (either in replicates of the same punch or across different punches). To avoid missing private variants in punches with only one replicate, exceptions were made for variants over 0.15 VAF in a single library.

We identified 345 high-confidence variants that passed quality control filters. These included 232 SNVs and 87 indels. Bioconductor's VariantAnnotation (201) R package was

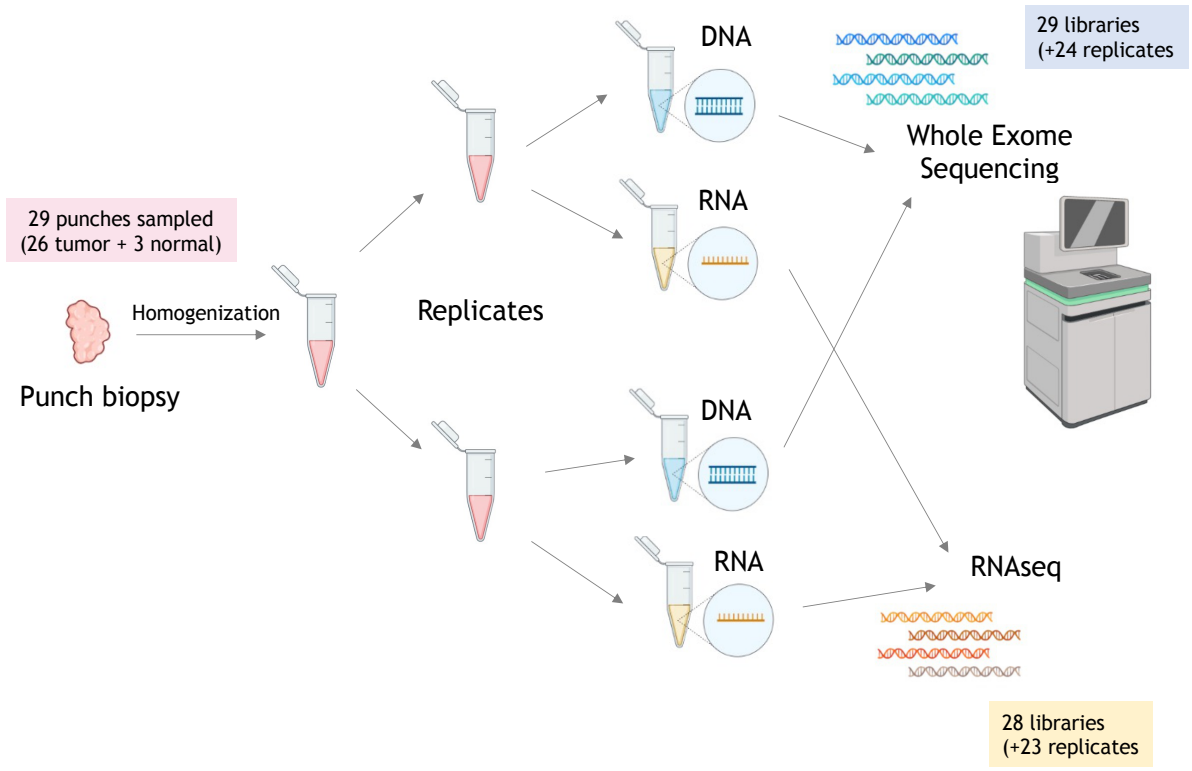

Figure 3.5: Tissue processing and sequencing. 29 total punch biopsies (26 tumor and 3 normal) were homogenized and lysed. The lysate was split into two replicates and DNA and RNA were extracted from each. WES and bulk RNAseq libraries were generated for each sample.

used to predict coding variants and identified 30 coding, non-synonymous variants (Figure 3.1 and 3.6).

3.4.5 Inference of copy number aberrations

We used Hatchet (202) to infer CNAs jointly across all tumor punch biopsies from the WES data. Through the Hatchet pipeline, we compared diploid and tetraploid models and selected the optimal number of clones. The optimal solution was diploid with 2 clones (1 tumor and 1 normal), which indicates that subclonal variation was not detectable.

3.4.6 Clonal deconvolution

All SNVs (non-coding, synonymous, and synonymous) were used as inputs for clonal deconvolution. SNV clusters were identified by PyClone-VI (194) and these clusters were input into Pairtree (195) to infer the tumor clonal tree.

3.4.7 Phylogenetic reconstruction

BEAST2 (203) was used to infer a time under a coalescent model with exponential population growth-based pseudo-sequences inferred per subclone. The tumor diameter was approximated to be 25mm based on the span of carcinoma in sampled frozen blocks (measured in WebDigitizer2). The tumor spanned 4 transverse 3-5 mm sections ($\sim$ 16 mm), giving an estimated volume of 5.2 cc, on par with published volumes of prostate tumor (198). Prostatic cancer cells were assumed to have a diameter of 30 μM (199). From this, a spherical tumor volume would be $1.4E-5mm^3$ and the estimated population size is $3.7E8$. Generation time is estimated to be 23 days, following the logic of (204) based on experimental and clinical evidence. For substitution rate, there is no direct experimental evidence to inform this parameter, so instead, we extrapolate from Alves et al., who assume 4.6E-10 substitutions per site per generation in colorectal cancer. Since prostate cancer has a lower

mutation burden than colorectal cancer (0.7 /Mb versus 3.1 / Mb (205)), we adjust this rate to $(4.6E-10)(0.7/3.1) = 1E-9$.

3.4.8 Phylodynamic modeling of subclonal migration

A Brownian random walk model of continuous trait evolution implemented in BEAST (206) was used to infer subclonal MRCAs' locations. An augmented dataset of sampled tips was inferred to represent the spatial and genetic distribution of the tumor population. We assigned 15 pseudo-cells per punch location and distributed them to genetic subclones based on the estimated cellular fraction within each punch. Random noise was added to each sample because under a random walk model tips being observed at exactly the same location in a continuous space is very unlikely.

3.4.9 Data and code availability

Code and data required to reproduce analyses are available at https://github.com/blab/primary_prostate_spatial. All other packages used for analysis and visualization are also open source.

3.4.10 Chapter acknowledgements

We would like to thank Lori Kollath and Andrew Ustach for their help in collecting the specimen samples, as well as the funding that supports the GU biorepository - the Pacific Northwest Prostate Cancer SPORE (P50CA97186), the PO1 NIH grant (PO1CA163227), and the Institute for Prostate Cancer Research.

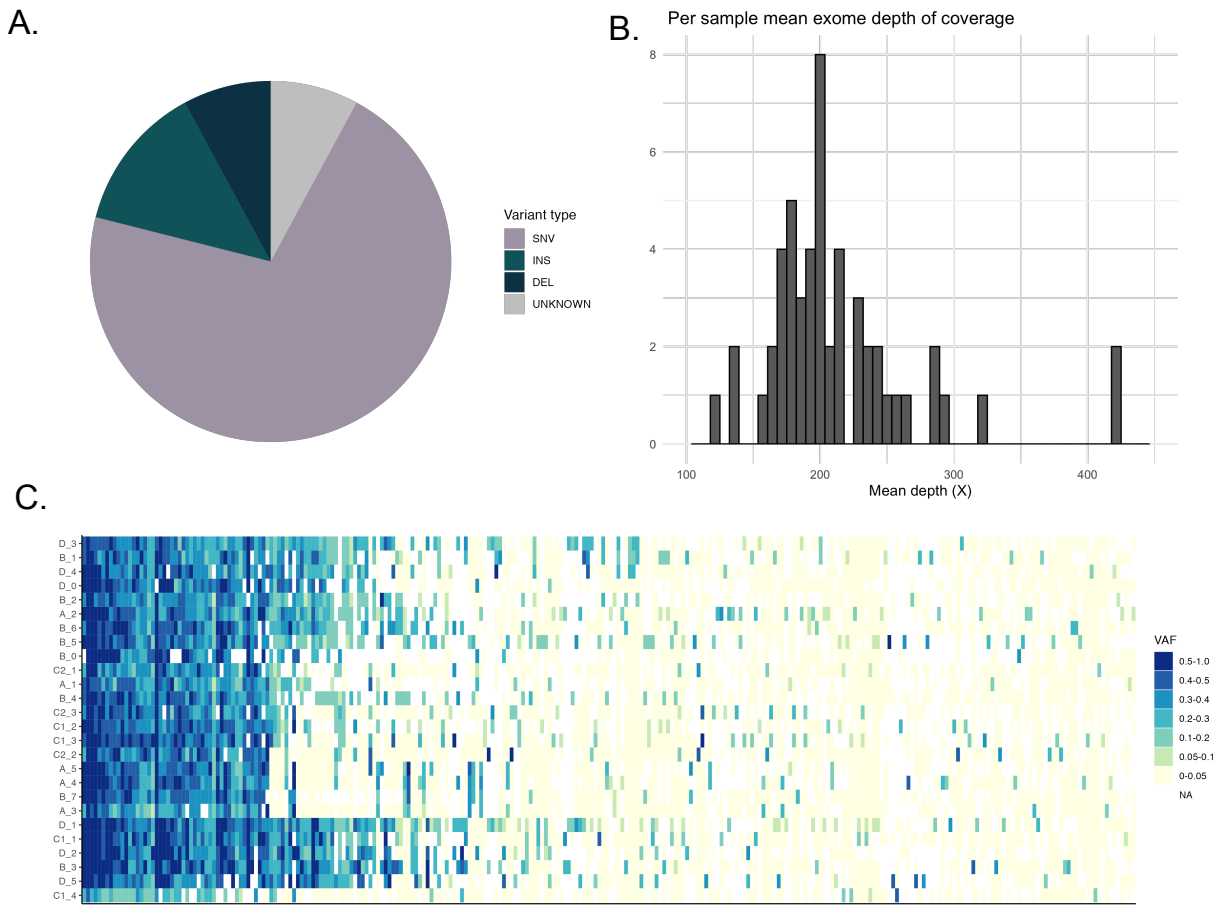

Figure 3.6: Extended analysis of WES data. **A.** Relative fractions of somatic SNVs and INDELS identified in variant calling of WES. **B.** Distribution of WES library depths. **C.** Heatmap of all somatic SNVs identified in the tumor. The color indicates VAF in the associated punch. **D.** Distribution of tumor purity estimated using each WES library.

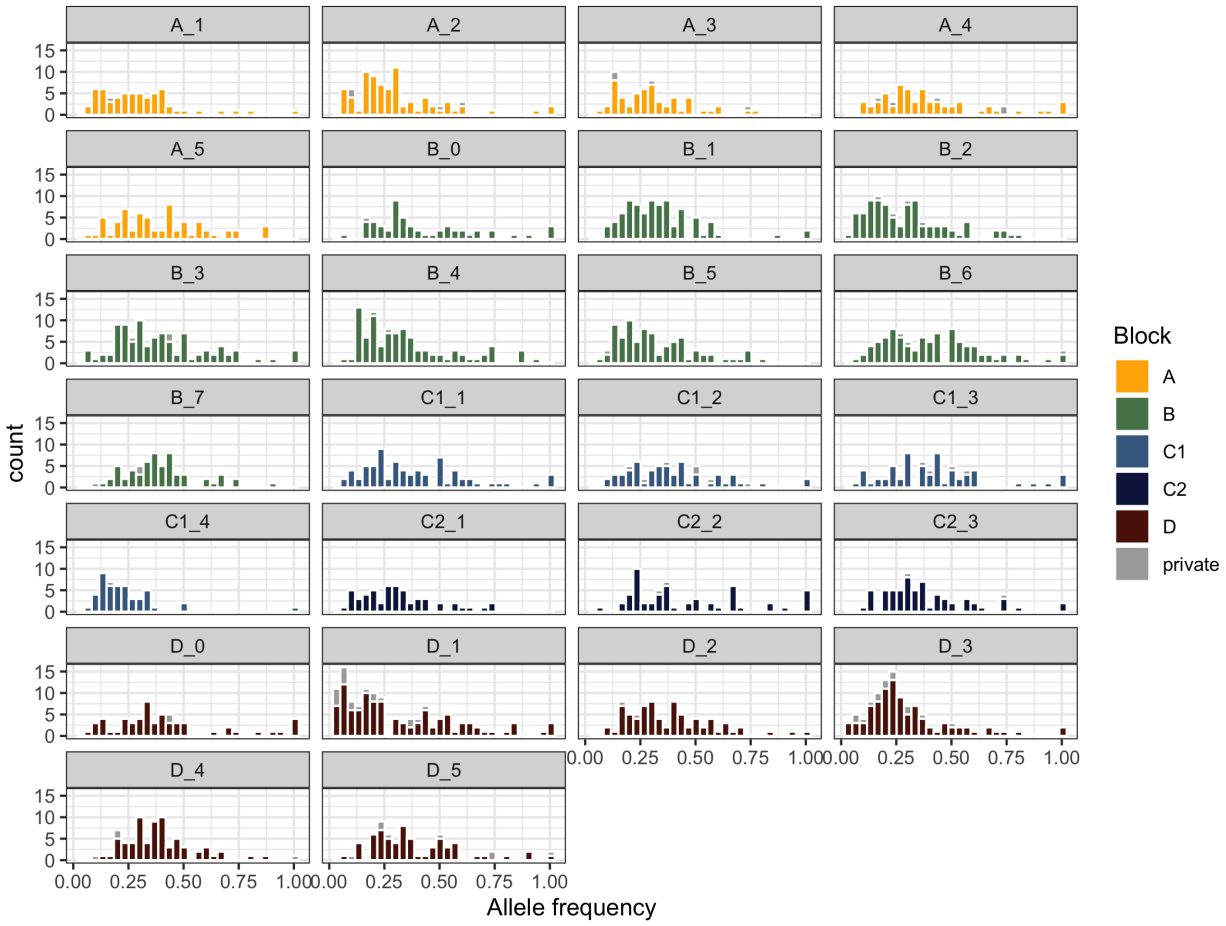

Figure 3.7: Variant allele frequency (VAF) histograms reveal punches have high subclonal diversity. Variant allele frequencies for all non-private (colored by punch) and private (grey) variants.

Chapter

LINKING TRANSCRIPTOMIC HETEROGENEITY TO THE SPATIAL DYNAMICS OF PRIMARY PROSTATE CANCER

4.1 Introduction

Prostate cancer (PCa) is remarkable for its clinical and biological heterogeneity. PCa cells evolve in a complex ecosystem that exhibits spatiotemporal variation (207, 208). Several studies have noted spatial variation in key cancer-related pathways, including those influencing metabolism, inflammation, androgen receptor signaling, neuroendocrine differentiation, and proliferation (209–211). Understanding this transcriptomic and phenotypic intra-tumor variability has high clinical importance, as transcriptomic signatures are increasingly used to predict disease progression (212, 213). There has been an enormous effort toward linking cancer phenotype to driver genes or other genetic determinants in prostate cancer. Most of these studies have included limited samples of a large number of tumors in an attempt to find consistent effects across many patients (214). Although this effort has successfully identified recurrent driver mutations, these drivers are seldom necessary or sufficient to observe a particular phenotype in clinical specimens, leaving a fundamental gap in understanding the extent to which genetic alterations can predict phenotype, or, more critically, how much of phenotypic heterogeneity can be explained by heritable genetic variation.

Here, we build on our investigation of genetic and spatial heterogeneity to understand spatial transcriptomic heterogeneity. From the same extensively-sampled, spatially-resolved punch biopsies described in Chapter 3, we analyzed bulk RNAseq profiles to characterize genes and pathways with a high intra-tumor variation. We apply an unsupervised statistical approach to identify pathways associated with underlying genetic or spatial features and

compare the relative variance explained. We find that although spatial and genetic dynamics are co-correlated, there is evidence of unique contributions to key cancer-related signaling pathways from spatial and genetic factors.

4.2 Results

4.2.1 Characterizing intra-tumor transcriptomic diversity

As described in Chapter 3 (Figure 3.5), we performed RNAseq on 26 punch biopsies from a single, mixed grade, primary tumor and on 3 punches from a pathologically normal prostate gland in the same punch. First, from comparing the expression profiles of the tumor versus the normal punches, we found that typical prostate cancer-related genes were differentially expressed (Figure 4.9B). For example, we found that tumor cells up-regulated XAGE1A, XAGE1B, CRISP3, and RELN, and down-regulated WIF1, all common patterns in prostate cancer (215–220).

We visualized global transcriptional heterogeneity in PCA space and observed that while normal punches were tightly clustered, there was extensive intra-tumor variability, with variation between punches being on a similar scale to that between normal and tumor populations (Figure 4.9). For each punch biopsy, we have mapped the spatial location and genetic subclonal populations (Figures 3.1 and 3.2) and next aimed to understand and link the observed intra-tumor transcriptomic diversity.

To begin to elucidate the relationship between ITH in genetic populations and RNA expression, we explored how the genetic composition of each population, described by the relative frequencies of genetic subclones described in Chapter 3 (Figure 3.2), was related to global transcriptional profiles. We plotted subclonal composition per punch in RNAseq PCA space (Figure 4.1A). We observed punch biopsies appear to be clustered based on the dominant genetic subclone. However, genetically similar populations are only slightly more likely to have similar transcriptional profiles (Figure 4.1B) as measured by pairwise Euclidean

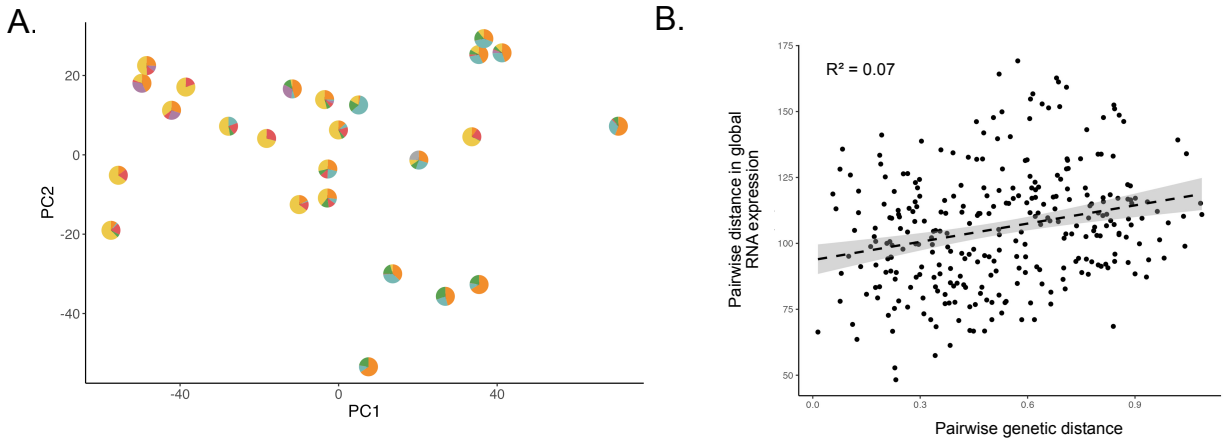

Figure 4.1: Characterization of transcriptomic heterogeneity. **A.** Intra-tumor transcriptomic heterogeneity in PCA space. Only tumor punches show with pie charts showing genetic clonal compositions. **B.** Pairwise Euclidean distance in genetic versus RNAseq datasets.

distance.

4.2.2 Uncovering transcriptomic pathways with genetic or spatial contributions

Next, we were interested in identifying which genes and pathways were differentially expressed in tumor space. To do so, we needed to find a way to represent spatial structure and drew from methods in ecology (221). Specifically, when thinking about variation over space, detecting variation along XYZ coordinates directly would only capture genes that were linearly associated along these axes, for example, if a gene was only expressed in the right but not the left side of the tumor. However, the potential spatial patterns are likely much more complex. Instead, we created Moran's Eigenvectors Maps (MEMs) as spatial features (222, 223). These convert spatial coordinates into orthogonal vectors calculated to maximize autocorrelation as measured by Moran's I (Figure 4.2).

We then used these MEMs as features in differential expression analysis and found spatial patterning of cancer-related genes (Figure 4.3). For example, among genes differential

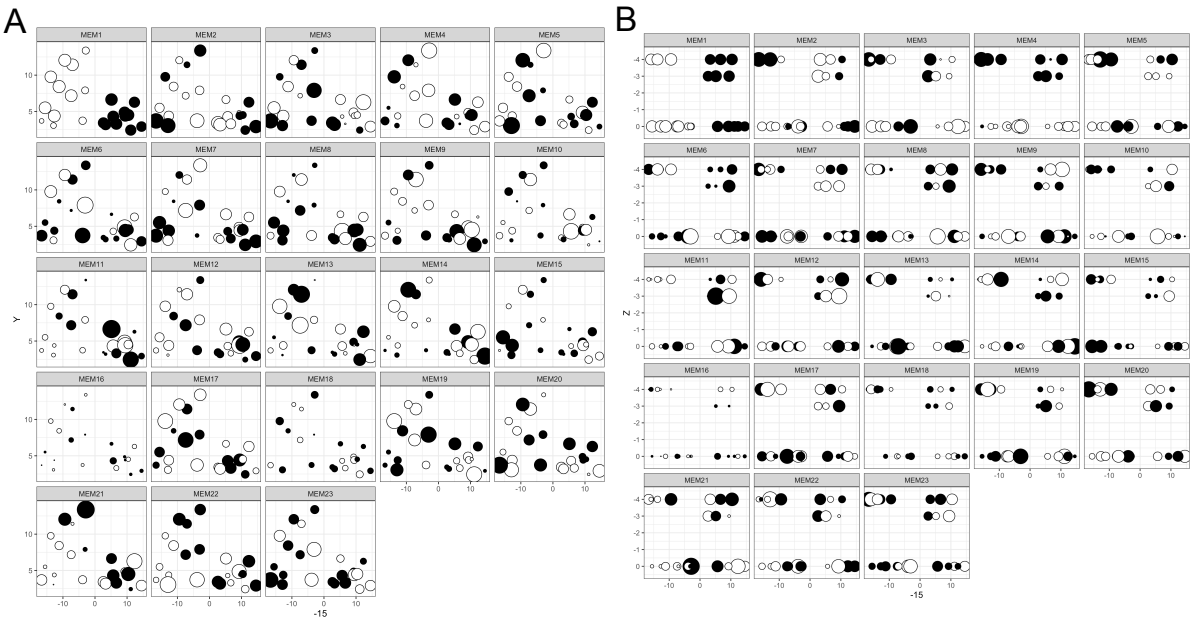

Figure 4.2: Moran's Eigenvectors Maps (MEM) as spatial predictors. MEMs were generated using a spatial weight matrix that represented the XYZ coordinates of the punch biopsies. Weights of each punch biopsy location in each MEM (headers), where point size corresponds to the absolute magnitude and black and white fill correspond to positive and negative values, respectively. XY space is shown in **A** and YZ space is shown in **B**.

expressed along the first four MEMs were several transcription factors involved in cancer progression, including FOSB, EGR1, and EGR3 (224–227).

Interestingly, neuroendocrine-related genes were also found in this set (CHGA and CHGB). Markers of neuronal-differentiation prostate cancer can be associated with neuroendocrine carcinoma, a more aggressive subtype of prostate cancer that is correlated with therapy resistance (228). Prostate tumors can often be subtyped by global gene expression related to androgen receptor (AR) signaling or NE differentiation, but instances of intra-tumor heterogeneity have been observed (229). We highlight that while there is homogeneous expression along the AR-signaling axis (Figure 4.4A, there is substantial variability in neuroendocrine-related genes (Figure 4.4B). We also observed intra-tumor variability in genes associated with cell-cycle proliferation (230).

Observed transcriptional intra-tumor heterogeneity is likely a combination of population composition and environmental and cell-intrinsic factors. To probe these interactions, we applied Canonical Correlation Analysis (CCA) to uncover underlying correlated intra-tumor genetic, spatial, and transcriptomic variation (Figure 4.5). CCA has been applied in multi-omics studies as an unsupervised approach to describe correlated features across datasets (231–233). Here, we hypothesize that co-linear sets of genetic variants (i.e. subclonal populations) may explain the expression of subsets of genes within the same pathway. We expect that most individual SNVs correlated with RNA expression are not causal, instead, they are markers for the genetic population that is tied to this expression change.

To first identify RNA expression pathways associated with genetic variation, we applied CCA to the genetic dataset which includes VAFs of somatic SNVs for each punch, and to the RNAseq count matrix. We analyzed the first three canonical pairs and visualized how genetic and transcriptomic features were weighted in these latent variable spaces. We applied GSEA to the canonical coefficients RNAseq genes ranked by their associated coefficients using MSigDB hallmark gene sets (234) to find candidate pathways with evidence of unique genetic

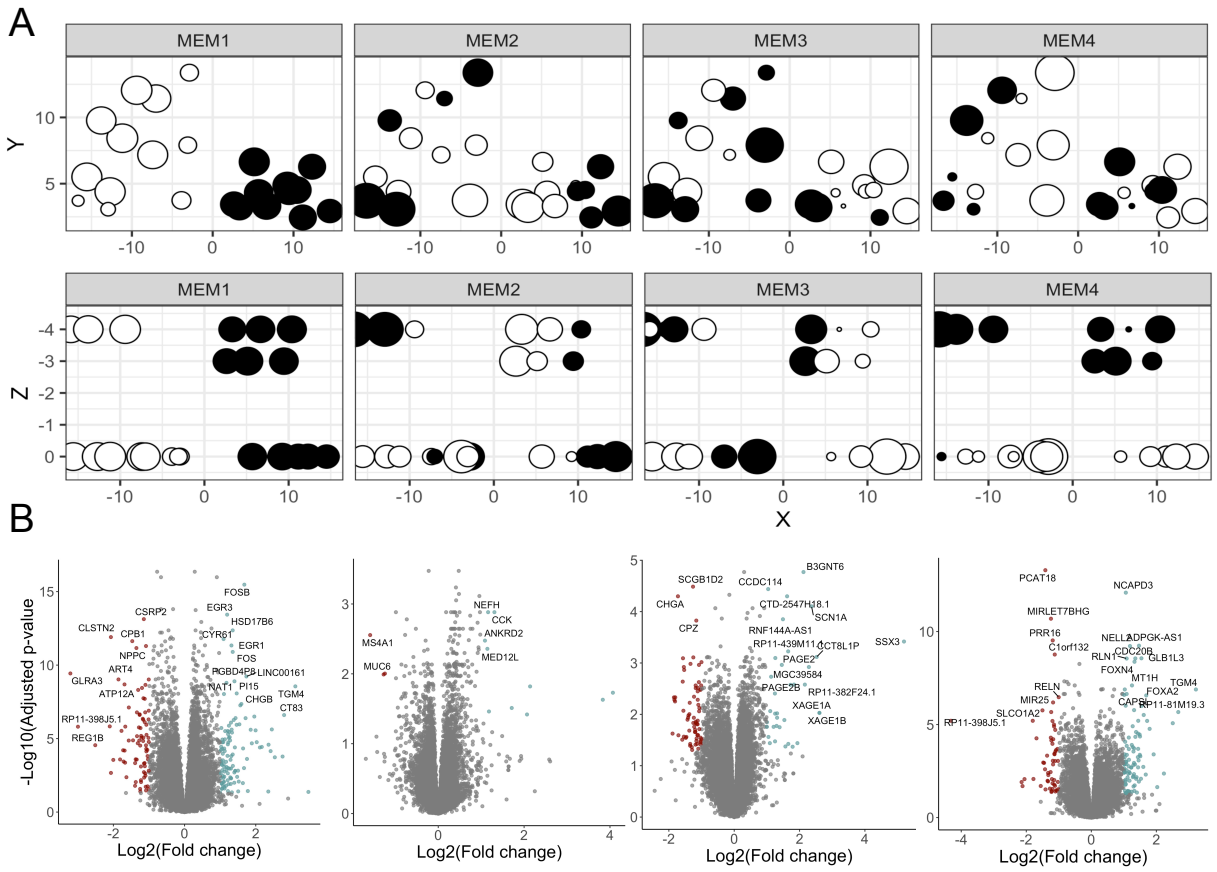

Figure 4.3: Cancer-related genes are differentially expressed across tumor space. we performed differential expression analysis using the spatial predictors identified in Figure 4.2. Differentially expressed genes for the first 4 MEMs (**A**) are represented as corresponding volcano plots (**B**).

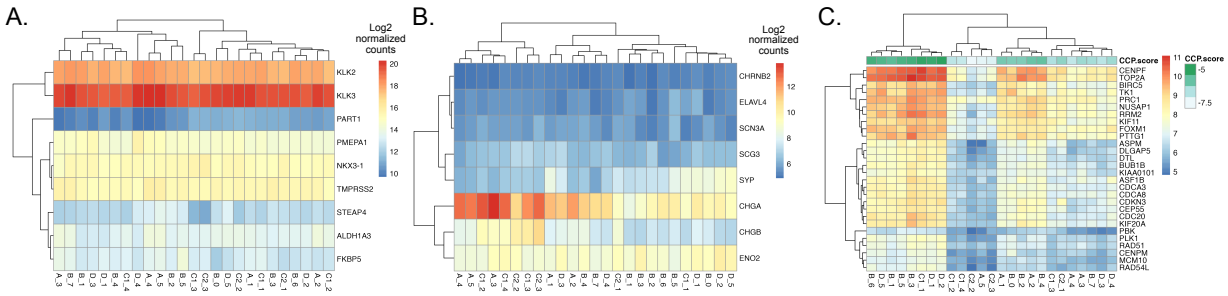

Figure 4.4: Intra-punch variability in neuroendocrine, but not AR signaling. Hierarchically clustered heatmaps of genes associated with **A.** Androgen Receptor (AR) signaling, **B.** neuroendocrine (NE) signaling, and **C.** cell-cyle progression (CCP). Columns are annotated with the cell-cycle progression score normalized by housekeeping genes (230).

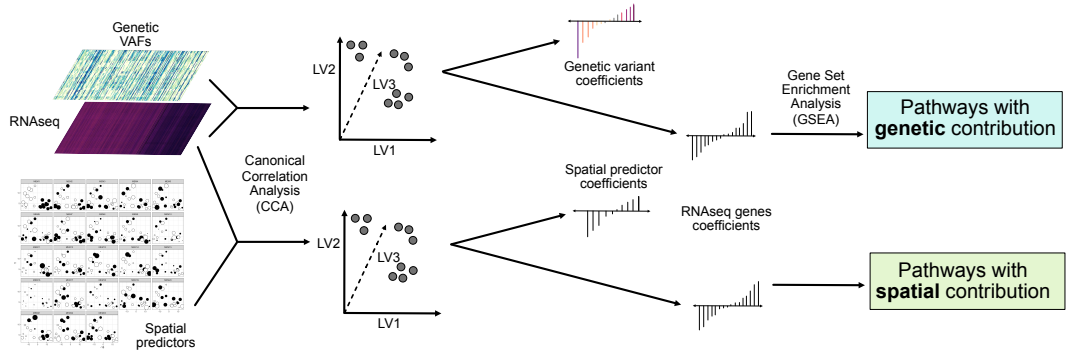

Figure 4.5: Unsupervised identification of correlated genetic, spatial, and transcriptomic features. Canonical Correlation Analysis (CCA) reduces large dimensional datasets (in this case genetic variant allele frequencies, spatial predictors (MEMs), and the normalized RNA counts matrix) into latent variable pairs that maximize their correlation. This is accomplished through linear combinations of each feature, which are given by a set of coefficients for each dataset. We input the ranked order of the RNAseq coefficients into Gene Set Enrichment Analysis (GSEA) to find candidate hallmark pathways that are correlated with genetic features.

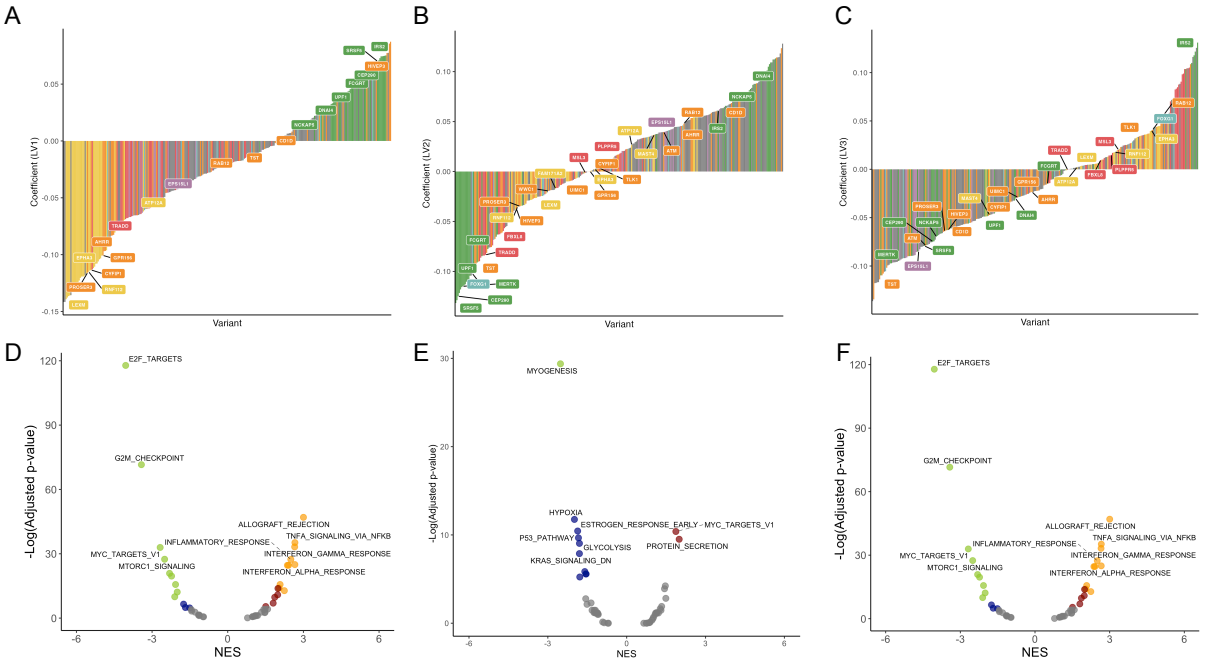

Figure 4.6: Unsupervised identification of correlated genetic and transcriptomic features. **A-C.** Coefficients associated with each genetic variant colored by subclonal assignment from Figure 3.2 for the first three latent variables. **D-F.** Volcano plots of enrichment of signature pathways based on the RNAseq coefficients for each corresponding LV

contribution (Figure 4.6). We observed that genes hallmark pathways "E2F targets" and "G2M checkpoint" were enriched along the same dimension that genetic variants assigned to the late-diverged orange subclone (subclone 1) were weighted highly (Figure 4.6A and D). On the other extreme of this dimension, genes in the immune-related pathways, such as "allograph rejection" and "interferon gamma response" along with variants associated with subclone (subclone 2) were weighted heavily.

We repeated the CCA analysis for spatial features, in which the spatial features were XYZ coordinates reduced to Moran's Eigenvector Maps (MEMs) (Figure 4.2). By using these spatial predictors as inputs to CCA, we generated candidate pathways linked to underlying spatial structure (Figure 4.7). We observed high overlap between candidate hallmark

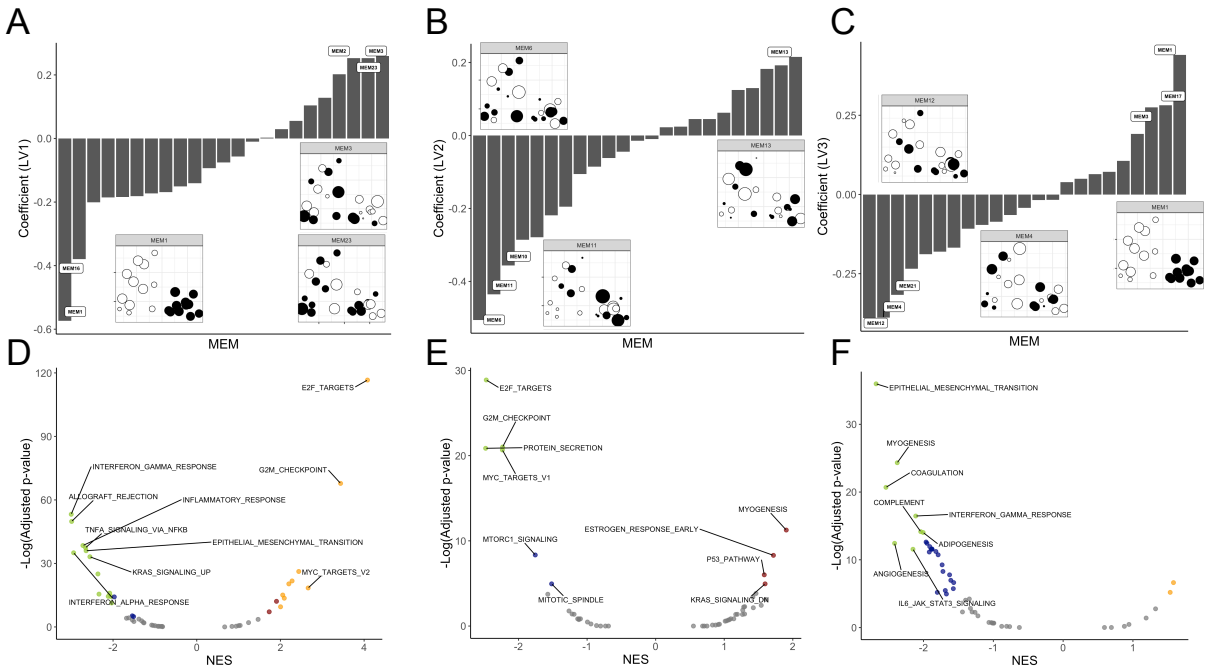

Figure 4.7: Identification of correlated spatial and transcriptomic features. Spatial predictors were Moran's Eigenvector Maps generated by maximizing autocorrelation from a spatial weight matrix describing relative punch locations (Figure 4.2). **A-C.** Coefficients of each spatial feature for the first three latent variable pairs. Insets show top-weighted MEMs. **D-E.** Candidate pathways with spatial correlation were found by applying GSEA based on the RNAseq coefficients found by CCA (as in Figure 4.6). Volcano plots show enrichment of MSigDB hallmark pathways in the first three canonical pairs.

pathways linked to genetic and spatial structure, as we would expect given that these features are correlated themselves (Figure 3.2D).

4.2.3 Comparing spatial and genetic contribution to variation in hallmark signaling pathway expression

Due to the extensive overlap of spatial and genetic structure, we aimed to determine which pathways had evidence of unique genetic or spatial contribution to their observed intra-tumor heterogeneity. We anticipate that if the variation of a pathway is explained by cell-intrinsic

factors (i.e. genetic or epigenetic), then pathway expression should have a high correlation with genetic features and minimal variance that is uniquely explained by spatial features. Alternatively, we expect that if a trait had a substantial environmental contribution, it should be correlated with spatial features independent of genetic background. We test these expectations through Variation Partitioning, a statistical method commonly used in ecology to investigate contributions of environment or genetic features to spatial variability (235, 236) (Figure 4.8A).

We rank candidate pathways by total genetic contribution (unique genetic and combined, Figure 4.8B) to highlight pathways that are more likely to be determined by cell-intrinsic mechanisms. We find mTOR signaling, Myc targets (V1), and androgen response as the top three pathways in these rankings. We also ranked candidate pathways by unique spatial contribution (Figure 4.8C) to highlight pathways that are more likely to be independent of genetic population background. We find TNFα signaling via NFKB, E2F targets, and UV response (up) as the top three pathways according to this logic. We note that many hallmark pathways that have statistically significant evidence of a unique genetic component still have unique spatial contributions, for example, "mitotic spindle", "oxidative phosphorylation", and "epithelial mesenchymal transition". Although

4.3 Discussion

We conceptualize ITH of RNA expression as a sum of cell-intrinsic and cell-extrinsic factors. Although global RNAseq heterogeneity was only weakly correlated with genetic background, we apply an unsupervised, statistical approach to uncover correlated genetic, spatial, and transcriptomic features. We further investigated hallmark RNA expression pathways to rank genetic and spatial contributions to biologically-relevant heterogeneity. We find evidence that cell-intrinsic (genetic) features explain a unique component of the variability in the expression of genes related to cell division, metabolism, inflammation, and hormonal signaling. We

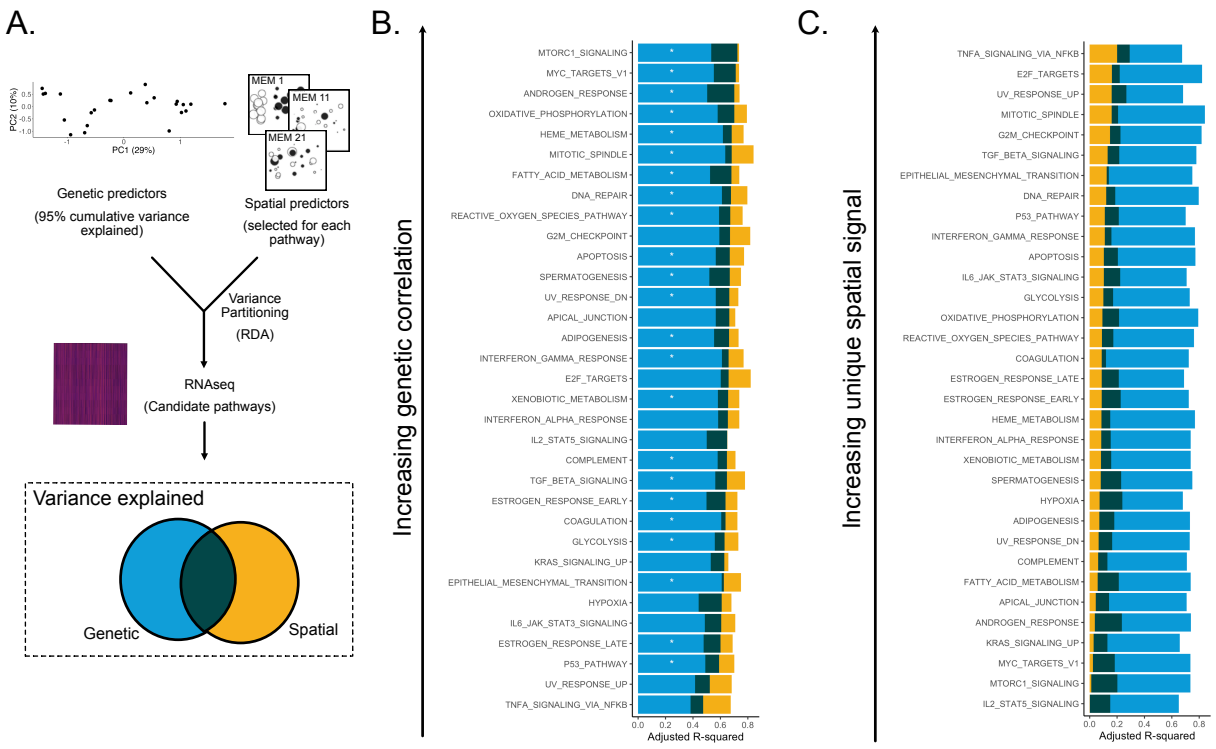

Figure 4.8: Comparison of genetic and spatial contribution through variance partitioning. **A.** Variance partitioning was used to identify unique and shared variation explained by genetic and spatial predictors for all candidate pathways identified from CCA (Figures 4.6 and 4.7). For this analysis components derived from PCA on VAFs were used for genetic predictors and MEMs selected to explain the maximum variance for each pathway were input as spatial predictors. **B.** Variance partitioning results for all candidate pathways (a subset of MSigDB hallmark gene sets) ranked by total (unique + combined) variance explained by genetic features. asterisks mark significance for unique genetic variance explained. **C.** Variance partitioning results for all candidate pathways ranked by unique variance explained by spatial predictors.

further link the higher expression of cell cycle-related genes this to more recently diverged genetic clones, connecting genetic and RNA expression ITH to the reconstructed spatiotemporal tumor dynamics and providing further evidence for ongoing subclonal evolution.

Although we found evidence of genetic contributions to several hallmark pathways, most pathways had unique variances explained by both spatial and genetic features. It is not unexpected that cancer-related signaling and expression profiles could vary based on cell-extrinsic and cell-intrinsic factors. For example, cancer cells often alter their metabolism by switching from oxidative phosphorylation to glycolysis (237). These are both pathways where the unique genetic contribution was statistically significant and this fits with past evidence that these traits are associated with driver mutations (238). At the same time, cells will also switch to glycolysis in hypoxic conditions. In these examples, we were able to detect these separate components. However, in cases where spatial and genetic features are extensively correlated, their individual effects cannot be estimated. In some ways, we encounter many of the same complexities as in an observational ecological study for which variance partitioning was originally designed (235, 236). Relying on statistical tools developed to address these challenges, we were able to disentangle the effects of spatial and genetic features, however, there are likely masked co-correlated effects that could not be detected. Tumor purity (Figure 3.6D) is another potential confounding factor, however, we attempt to correct for these effects by incorporating tumor purity as a batch effect.

We further note that we are limited to understanding the genetic, spatial, and transcriptomic interactions in a single patient. We do not comment on specific genetic features or drivers, but instead, focus on the overall heritability of these traits. Alternatively, we highlight that extensive multi-region sampling of a single tumor is uniquely useful in investigating the differential genetic and spatial contributions to transcriptomic variability that likely would have been harder to determine while accounting for inter-patient heterogeneity. Further investigation could expand on this study design for more tumors to better test the

generalizability of the dynamics observed here.

4.4 Methods

4.4.1 Bulk RNA sequencing

RNAseq libraries were generated using the Illumina TruSeq Stranded mRNA kit from 1ug RNA per replicate, where replicates are independent RNAseq library preparations from RNA isolated from the same punch biopsy (Figure 3.5). Technical replicates were generally well correlated (Figure 4.9E). Tapestation was used to calculate RINe scores for each RNA sample.

4.4.2 RNAseq analysis

The R package DESeq2 (239) was used to normalize count matrices and for downstream differential expression (DE) analysis. Tumor purity as esitmated with TITAN was included in the design as potential confounding variable. For all analyses, both replicates of C1_4 were excluded due to low cellularity.

4.4.3 Statistical analysis of spatial, genetic, and transcriptomic ITH

For the genetic dataset, VAFs were normalized by tumor purity. Canonical correlation analysis was implemented using the R package *PMA* (240). We analyzed the first three canonical pairs for each genetic and spatial analysis. We used the R package *fgsea* (241) to run a gene set enrichment analysis (GSEA) on the RNAseq genes weighted by coefficients in each canonical vector for hallmark gene sets. Significantly enriched pathways in either the genetic or spatial CCAs were defined as candidate pathways to which we applied Variance Partitioning (235). We performed the Variance Partitioning using the *vegan* R package. We selected the first 18 principle components that explained 95% of the genetic variance as genetic features to avoid problems co-linearity. For each pathway, we also selected MEMs (242) as spatial predictors based on maximizing the adjusted R-squared. Prior to analysis,

we used the R package limma (243) to remove variation explained by tumor purity. We applied the Benjamini-Hochberg procedure (FDR < 0.05) to correct for multiple hypothesis testing.

4.4.4 Data and code availability

Code and data required to reproduce analyses are available at https://github.com/blab/primary_prostate_spatial, and all other R packages used are open source.

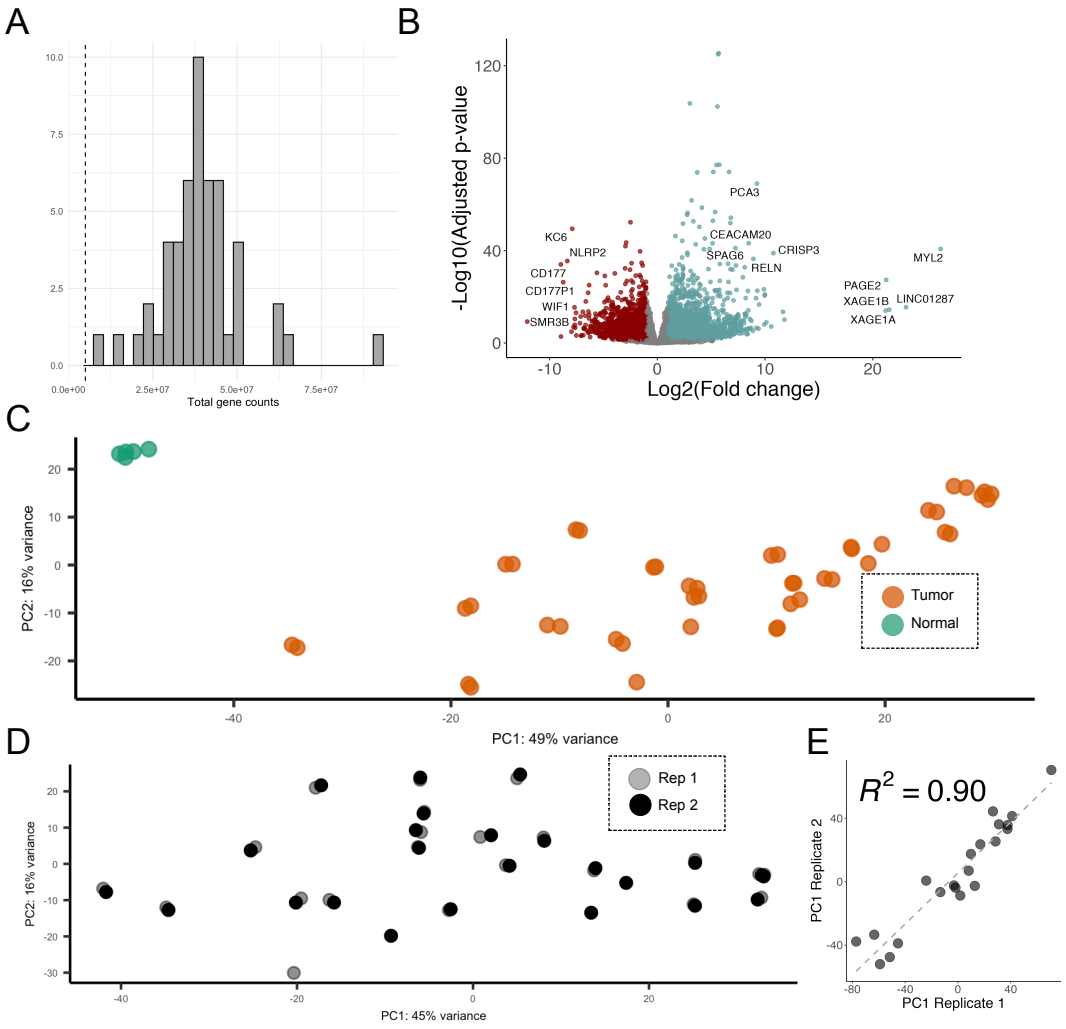

Figure 4.9: RNAseq library quality and consistency. **A.** Histogram of total gene counts for each bulk RNAseq library (most punches had two replicates). **B.** Volcano plot of differential expression analysis of tumor compared to pathologically normal punches. **C.** PCA plot of global RNAseq variability including separation between normal (green) and tumor (orange) punches, with replicates plotted. **D.** PCA of only tumor punches representing intra-tumor variability, showing low variability of technical replicates as compared to between punch biopsies (indicated by opacity). **E.** Scatter plot showing the correlation between the RNAseq technical replicates along the first principal component.

Chapter

CONCLUSION

In previous chapters, I have demonstrated a diverse set of phylodynamic and statistical approaches to elucidate evolutionary dynamics underlying the development of intra-tumor heterogeneity. It is clear from these results that both spatial and genetic features play important roles. In Chapter 2, we detected signals of boundary-driven modes of growth in two hepatocellular carcinomas. At the same time, the original study (34) found that observed spatial distribution was also likely influenced by selection of a center subclone. The same interplay between spatial and genetic features was seen in our case study of prostate cancer. In Chapter 3, we found that the distribution of subclones could be explained by diffusion and local expansion of tumor cells. Although these findings could be attributed completely by neutral, spatially-constrained growth, we then found evidence of genetically-linked proliferation-related signaling in Chapter 4. Here, the interaction of spatial and genetic features in spatiotemporal dynamics were further highlighted when almost all hallmark pathways investigated in Chapter 4 with a significant genetic partition also had some amount of variance uniquely explained by spatial factors. Taken together, these results confirm that the factors that determine clonal success and intra-tumor heterogeneity are multi-faceted and complex.

5.1 Understanding spatial evolutionary dynamics with the goal of improving therapy

From an evolutionary perspective, if drug resistance exists in the tumor population, treatment provides a large fitness advantage to the resistant cells and will lead to outgrowth of the

resistant population (244). Multi-drug therapy, also commonly used in antibiotic strategies (245), statistically decreases the chance that a single cell will resistance to all drugs used (246). However, once a cell has acquired resistance, selection for this clone is likely just as strong, and unfortunately, acquisition of resistance in cancer is quite common. Unlike in bacteria, which are evolutionary distant from their hosts and can be targeted more specifically, cancer cells are only slightly genetically and phenotypically different from their host cells. Therefore, therapeutic dose is often limited by off-target effects and toxicity, so phenotypic variability is often enough to allowed continued viability (247).

Evolutionary-informed therapy tries to take these selective dynamics into account (248). For example, clinical trials in prostate cancer have demonstrated that resistance can be delayed by periodically pausing treatment and allowing sensitive cells to compete with resistant cells (249, 250). However, these approaches require understanding the complexities of cell-to-cell competition and selection in cancer, as this will determine where cells fall on the tumor fitness landscape. In this respect, understanding and incorporating spatial dynamics into predictive cancer evolution models is critical. Consider the two extremes of spatial dynamics that could exist. In a tumors driven by selection and with high turnover, new driver mutations will rapidly sweep the population. In this case, changing forces of selection (i.e. drug treatment) will have a predictable and strong impact. On the other hand, if growth is extremely spatially constrained, for example, in the center of a tumor with limited turnover, cells would primarily compete locally, and relative fitness advantage will be more important than the global fitness landscape (251). The hepatocellular and prostate tumors investigated in Chapters 2 and 3, respectively, both appeared to be closer to the latter extreme. Therefore, if patients with similar tumors were to be treated with evolutionary-informed therapy, spatial constraints should certainly be taken into account.

5.2 How predictable is cancer evolution?

Although the complexity of cancer evolution is one major roadblock to controlling disease, there is another, more fundamental challenge. Evolutionary-informed therapy (252, 253, 250) relies on the ideology that if we knew enough about the underlying evolutionary processes, we could trap the cancer cells into a desired trajectory. However, even if we knew every possible characteristic about a tumor population, we still may reach a limit of predictability due to stochastic forces or the sheer number of interacting forces. This is reminiscent of evolutionary biologist Stephen Jay Gould's famously posed question: if we rewound the "tape of life" would we still be in the same world (254)? However, unlike Earth as a whole, in which we can only observe one outcome, we can watch the "tape" of cancer evolution replay many times over many different, slightly varying, contexts.

The results presented in this dissertation largely fall within the more optimistic perspective of this debate. The study design in Chapters 3 and 4 is unique in that is extensively sampled variation within a single environment instead of finding common effects across many tumors (41, 88). Therefore, while this design was not set up to find specific features, such as driver mutations, that predicted transcriptomic variability, it was more powered to detect *if* heterogeneity was predictable. That we were able to explain variance in several signaling pathways through genetic and spatial features is encouraging. For while the specific effect of a mutation may be harder to predict given that every context is slightly different, cancer phenotypes (25) or fitness (255) could be quantified and used to manipulate the evolutionary outcome (253).

5.3 The potential of phylodynamics in studying cancer evolution

The efforts presented in this dissertation are among the first to adapt phylodynamic models to cancer evolution (147) and highlight the potential insights to be gained by more broadly applying phylodynamic approaches. I focused on modeling spatial traits (i.e. edge and

center, and XYZ coordinates) in prostate and liver cancer, but both discrete and continuous trait evolution models could be applied much more broadly. Morphology or grade, metastatic potential, transcriptomic profiles, or mutational signatures are just a few examples of traits that could be modeled. In addition, a wide variety of population models have applied to study pathogen evolution that would also be useful in cancer, including estimating change in effective population size over time or enforcing population structure (perhaps as in stem-cell models of cancer evolution) (256).

A major challenge to enabling wide-spread use of phylodynamics is developing cancer-specific models or adaptations to allow application of cancer data. For example, in Chapter 2 we developed a novel state-dependent sequence evolution model that effectively linked increased birth with mutation, and this is essential for detecting boundary-driven growth in tumors. In Chapter 3, a continuous trait evolution model could not be immediately applied to clone trees, so augmented the sampled data into a set of representative tips that could be input into this model appropriately. These problems could be alleviated as single-cell data becomes more available and gains genetic resolution, but also overcome by specifically incorporating cancer-specific processes in in phylodynamic models. With the wide-spread adoption of genomic sequencing of viral outbreaks, phylodynamics has become a central tool in pathogen surveillance (118). Once these models have been adapted to cancer and the data becomes available, it could be similarly applied to track cancer evolution in clinical tumors. Therefore, future work could both apply methods adapted in this study to more tumors and cancer types, but it should also continue to push forward and expand phylodynamic methods in cancer.

BIBLIOGRAPHY

[1] Garth L. Nicolson. Generation of phenotypic diversity and progression in metastatic tumor cells. *Cancer and Metastasis Reviews*, 3(1):25–42, 1984. doi: 10.1007/BF00047691. URL https://doi.org/10.1007/BF00047691.

[2] Charles Swanton. Intratumor Heterogeneity: Evolution through Space and Time. *Cancer Research*, 72(19):4875–4882, 09 2012. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-12-2217. URL https://doi.org/10.1158/0008-5472.CAN-12-2217.

[3] Nicholas McGranahan and Charles Swanton. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*, 27(1):15–26, 2023/01/13 2015. doi: 10.1016/j.ccell.2014.12.001. URL https://doi.org/10.1016/j.ccell.2014.12.001.

[4] Johannes G. Reiter, Wei-Ting Hung, I-Hsiu Lee, Shriya Nagpal, Peter Giunta, Sebastian Degner, Gang Liu, Emma C. E. Wassenaar, William R. Jeck, Martin S. Taylor, Alexander A. Farahani, Hetal D. Marble, Simon Knott, Onno Kranenburg, Jochen K. Lennerz, and Kamila Naxerova. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nature Genetics*, 52(7):692–700, 2020. doi: 10.1038/s41588-020-0633-2. URL https://doi.org/10.1038/s41588-020-0633-2.

[5] Ruping Sun, Zheng Hu, Andrea Sottoriva, Trevor A Graham, Arbel Harpak, Zhicheng Ma, Jared M Fischer, Darryl Shibata, and Christina Curtis. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature Genetics*, 49(7):1015–1024, 2017. doi: 10.1038/ng.3891. URL https://doi.org/10.1038/ng.3891.

[6] Ikram Ullah, Govindasamy-Muralidharan Karthik, Amjad Alkodsi, Una Kjällquist, Gustav Stålhammar, John Lövrot, Nelson-Fuentes Martinez, Jens Lagergren, Sampsa Hautaniemi, Johan Hartman, and Jonas Bergh. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *The Journal of clinical investigation*, 128(4):1355–1370, 04 2018. doi: 10.1172/JCI96149. URL https://pubmed.ncbi.nlm.nih.gov/29480816.

[7] D. J. Woodcock, E. Riabchenko, S. Taavitsainen, M. Kankainen, G. Gundem, D. S. Brewer, P. Ellonen, M. Lepistö, Y. A. Golubeva, A. C. Warner, T. Tolonen, J. Jasu, W. B. Isaacs, M. R. Emmert-Buck, M. Nykter, T. Visakorpi, G. S. Bova, and D. C. Wedge. Prostate cancer evolution from multilineage primary to single lineage metastases with implications for liquid biopsy. *Nature Communications*, 11 (1):5070, 2020. doi: 10.1038/s41467-020-18843-5. URL https://doi.org/10.1038/s41467-020-18843-5.

[8] Zi-Ming Zhao, Bixiao Zhao, Yalai Bai, Atila Iamarino, Stephen G Gaffney, Joseph Schlessinger, Richard P Lifton, David L Rimm, and Jeffrey P Townsend. Early and multiple origins of metastatic lineages within primary tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 113(8):2140–2145, 02 2016. doi: 10.1073/pnas.1525677113. URL https://pubmed.ncbi.nlm.nih.gov/26858460.

[9] Joao M. Alves, Sonia Prado-López, JoséManuel Cameselle-Teijeiro, and David Posada. Rapid evolution and biogeographic spread in a colorectal cancer. *Nature Communications*, 10(1):5139, 2019. doi: 10.1038/s41467-019-12926-8. URL https://doi.org/10.1038/s41467-019-12926-8.

[10] Matthew W. Fittall and Peter Van Loo. Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Medicine*, 11(1):20, 2019. doi: 10.1186/s13073-019-0632-z. URL https://doi.org/10.1186/s13073-019-0632-z.

[11] Michael C. Haffner, Wilbert Zwart, Martine P. Roudier, Lawrence D. True, William G. Nelson, Jonathan I. Epstein, Angelo M. De Marzo, Peter S. Nelson, and Srinivasan Yegnasubramanian. Genomic and phenotypic heterogeneity in prostate cancer. *Nature Reviews Urology*, 18(2):79–92, 2021. doi: 10.1038/s41585-020-00400-w. URL https://doi.org/10.1038/s41585-020-00400-w.

[12] Joan R. Shapiro, Wai-Kwan A. Yung, and William R. Shapiro. Isolation, Karyotype, and Clonal Growth of Heterogeneous Subpopulations of Human Malignant Gliomas1. *Cancer Research*, 41(6):2349–2359, 06 1981. ISSN 0008-5472.

[13] A P Albino, K O Lloyd, A N Houghton, H F Oettgen, and L J Old. Heterogeneity in surface antigen and glycoprotein expression of cell lines derived from different melanoma metastases of the same patient. Implications for the study of tumor antigens. *Journal of Experimental Medicine*, 154(6):1764–1778, 12 1981. ISSN 0022-1007. doi: 10.1084/jem.154.6.1764. URL https://doi.org/10.1084/jem.154.6.1764.

[14] J. E. Griffin, D. R. Allman, J. L. Durrant, and J. D. Wilson. Variation in steroid 5 alpha-reductase activity in cloned human skin fibroblasts. shift in phenotypic expression from high to low activity upon subcloning. *Journal of Biological Chemistry*, 256(8):3662–3666, 1981. doi: https://doi.org/10.1016/S0021-9258(19)69504-9. URL https://www.sciencedirect.com/science/article/pii/S0021925819695049.

[15]

[16] Garth L. Nicolson. Cell surface molecules and tumor metastasis: Regulation of metastatic phenotypic diversity. *Experimental Cell Research*, 150(1):3–22, 1984. ISSN 0014-4827. doi: https://doi.org/10.1016/0014-4827(84)90696-7. URL https://www.sciencedirect.com/science/article/pii/0014482784906967.

[17] Leon P Bignold, Brian LD Coghlan, and Hubertus PA Jersmann. Hansemann's ideas of the nature of cancer: description and analysis, 2007.

[18] Panagiotis Karras, Ignacio Bordeu, Joanna Pozniak, Ada Nowosad, Cecilia Pazzi, Nina Van Raemdonck, Ewout Landeloos, Yannick Van Herck, Dennis Pedri, Greet Bervoets, Samira Makhzami, Jia Hui Khoo, Benjamin Pavie, Jochen Lamote, Oskar Marin-Bejar, Michael Dewaele, Han Liang, Xingju Zhang, Yichao Hua, Jasper Wouters, Robin Browaeys, Gabriele Bergers, Yvan Saeys, Francesca Bosisio, Joost van den Oord, Diether Lambrechts, Anil K. Rustgi, Oliver Bechter, Cedric Blanpain, Benjamin D. Simons, Florian Rambow, and Jean-Christophe Marine. A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature*, 610(7930):190–198, 2022. doi: 10.1038/s41586-022-05242-7. URL https://doi.org/10.1038/s41586-022-05242-7.

[19] Christine L Chaffer and Robert A Weinberg. A perspective on cancer cell metastasis. *science*, 331(6024):1559–1564, 2011.

[20] Peter C. Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260): 23–28, 2023/01/13 1976. doi: 10.1126/science.959840. URL https://doi.org/10.1126/science.959840.

[21] Lauren M. F. Merlo, John W. Pepper, Brian J. Reid, and Carlo C. Maley. Cancer as an evolutionary and ecological process. *Nature Reviews Cancer*, 6(12):924–935, 2006. doi: 10.1038/nrc2013. URL https://doi.org/10.1038/nrc2013.

[22] James R. M. Black and Nicholas McGranahan. Genetic and non-genetic clonal diversity in cancer evolution. *Nature Reviews Cancer*, 21(6):379–392, 2021. doi: 10.1038/s41568-021-00336-2. URL https://doi.org/10.1038/s41568-021-00336-2.

[23] Alfred G. Knudson. Mutation and cancer: Statistical study of retinoblastoma. *Pro-*

ceedings of the National Academy of Sciences, 68(4):820–823, 2023/01/17 1971. doi: 10.1073/pnas.68.4.820. URL https://doi.org/10.1073/pnas.68.4.820.

[24] R M Sutherland and R E Durand. Growth and cellular characteristics of multicell spheroids. *Recent Results Cancer Res*, 95:24–49, 1984.

[25] Douglas Hanahan and Robert A. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674, 2011. doi: https://doi.org/10.1016/j.cell.2011.02.013. URL https://www.sciencedirect.com/science/article/pii/S0092867411001279.

[26] Samra Turajlic, Andrea Sottoriva, Trevor Graham, and Charles Swanton. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics*, 20(7):404–416, 2019.

[27] Ignacio Vázquez-García, Florian Uhlitz, Nicholas Ceglia, Jamie L. P. Lim, Michelle Wu, Neeman Mohibullah, Juliana Niyazov, Arvin Eric B. Ruiz, Kevin M. Boehm, Viktoria Bojilova, Christopher J. Fong, Tyler Funnell, Diljot Grewal, Eliyahu Havasov, Samantha Leung, Arfath Pasha, Druv M. Patel, Maryam Pourmaleki, Nicole Rusk, Hongyu Shi, Rami Vanguri, Marc J. Williams, Allen W. Zhang, Vance Broach, Dennis S. Chi, Arnaud Da Cruz Paula, Ginger J. Gardner, Sarah H. Kim, Matthew Lennon, Kara Long Roche, Yukio Sonoda, Oliver Zivanovic, Ritika Kundra, Agnes Viale, Fatemeh N. Derakhshan, Luke Geneslaw, Shirin Issa Bhaloo, Ana Maroldi, Rahelly Nunez, Fresia Pareja, Anthe Stylianou, Mahsa Vahdatinia, Yonina Bykov, Rachel N. Grisham, Ying L. Liu, Yulia Lakhman, Ines Nikolovski, Daniel Kelly, Jianjiong Gao, Andrea Schietinger, Travis J. Hollmann, Samuel F. Bakhoum, Robert A. Soslow, Lora H. Ellenson, Nadeem R. Abu-Rustum, Carol Aghajanian, Claire F. Friedman, Andrew McPherson, Britta Weigelt, Dmitriy Zamarin, and Sohrab P. Shah. Ovarian cancer mutational processes drive site-specific immune evasion. *Nature*, 612(7941):778–786, 2022. doi: 10.1038/s41586-022-05496-1. URL https://doi.org/10.1038/s41586-022-05496-1.

[28] Andrea Sottoriva, Haeyoun Kang, Zhicheng Ma, Trevor A Graham, Matthew P Salomon, Junsong Zhao, Paul Marjoram, Kimberly Siegmund, Michael F Press, Darryl Shibata, et al. A big bang model of human colorectal tumor growth. *Nature genetics*, 47(3):209, 2015.

[29] Marc J Williams, Benjamin Werner, Chris P Barnes, Trevor A Graham, and Andrea Sottoriva. Identification of neutral tumor evolution across cancer types. *Nature Genetics*, 48(3):238–244, 2016. doi: 10.1038/ng.3489. URL https://doi.org/10.1038/ng.3489.

[30] Phuong H. D. Nguyen, Martin Wasser, Chong Teik Tan, Chun Jye Lim, Hannah L. H. Lai, Justine Jia Wen Seow, Ramanuj DasGupta, Cheryl Z. J. Phua, Siming Ma, Jicheng Yang, Sheena D/O Suthen, Wai Leong Tam, Tony K. H. Lim, Joe Yeong, Wei Qiang Leow, Yin Huei Pang, Gwyneth Soon, Tracy Jiezhen Loh, Wei Keat Wan, Chung Yip Chan, Peng Chung Cheow, Han Chong Toh, Alfred Kow, Yock Young Dan, Juinn Huar Kam, Shridhar Iyer, Krishnakumar Madhavan, Alexander Chung, Glenn K. Bonney, Brian K. P. Goh, Naiyang Fu, Victor C. Yu, Weiwei Zhai, Salvatore Albani, Pierce K. H. Chow, and Valerie Chew. Trajectory of immune evasion and cancer progression in hepatocellular carcinoma. *Nature Communications*, 13(1):1441, 2022. doi: 10.1038/s41467-022-29122-w. URL https://doi.org/10.1038/s41467-022-29122-w.

[31] Jorge Luis Galeano Niño, Hanrui Wu, Kaitlyn D. LaCourse, Andrew G. Kempchinsky, Alexander Baryiames, Brittany Barber, Neal Futran, Jeffrey Houlton, Cassie Sather, Ewa Sicinska, Alison Taylor, Samuel S. Minot, Christopher D. Johnston, and Susan Bullman. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature*, 611(7937):810–817, 2022. doi: 10.1038/s41586-022-05435-0. URL https://doi.org/10.1038/s41586-022-05435-0.

[32] Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko,

Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, Vasyl Vaskivskyi, Tong Li, Omer Ali Bayraktar, Sarah Pinder, Andrea L. Richardson, Sandro Santagata, Peter J. Campbell, Hege Russnes, Moritz Gerstung, Mats Nilsson, and Lucy R. Yates. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature*, 611(7936):594–602, 2022. doi: 10.1038/s41586-022-05425-2. URL https://doi.org/10.1038/s41586-022-05425-2.

[33] Alexander Davis, Ruli Gao, and Nicholas Navin. Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1867(2):151–161, 2017.

[34] Guanghao Li, Zuyu Yang, Dafei Wu, Sixue Liu, Xuening Li, Tao Li, Yawei Li, Liji Liang, Weilong Zou, Chung-I Wu, Hurng-Yi Wang, and Xuemei Lu. Evolution under Spatially Heterogeneous Selection in Solid Tumors. *Molecular Biology and Evolution*, 39(1), 11 2021. ISSN 1537-1719. doi: 10.1093/molbev/msab335. URL https://doi.org/10.1093/molbev/msab335. msab335.

[35] Shaoping Ling, Zheng Hu, Zuyu Yang, Fang Yang, Yawei Li, Pei Lin, Ke Chen, Lili Dong, Lihua Cao, Yong Tao, et al. Extremely high genetic diversity in a single tumor points to prevalence of non-darwinian cell evolution. *Proceedings of the National Academy of Sciences*, 112(47):E6496–E6505, 2015.

[36] Robert Noble, Dominik Burri, Cécile Le Sueur, Jeanne Lemant, Yannick Viossat, Jakob Nikolas Kather, and Niko Beerenwinkel. Spatial structure governs the mode of tumour evolution. *Nature Ecology & Evolution*, 6(2):207–217, 2022.

[37] Xiao Fu, Yue Zhao, Jose I. Lopez, Andrew Rowan, Lewis Au, Annika Fendler, Steve Hazell, Hang Xu, Stuart Horswell, Scott T. C. Shepherd, Charlotte E. Spencer, Lavinia Spain, Fiona Byrne, Gordon Stamp, Tim O'Brien, David Nicol, Marcel-

lus Augustine, Ashish Chandra, Sarah Rudman, Antonia Toncheva, Andrew J. S. Furness, Lisa Pickering, Santosh Kumar, Dow-Mu Koh, Christina Messiou, Derfel ap Dafydd, Matthew R. Orton, Simon J. Doran, James Larkin, Charles Swanton, Erik Sahai, Kevin Litchfield, Samra Turajlic, Ben Challacombe, Simon Chowdhury, William Drake, Archana Fernando, Nicos Fotiadis, Emine Hatipoglu, Karen Harrison-Phipps, Peter Hill, Catherine Horsfield, Teresa Marafioti, Jonathon Olsburgh, Alexander Polson, Sergio Quezada, Mary Varia, Hema Verma, Paul A. Bates, and on behalf of the TRACERx Renal Consortium. Spatial patterns of tumour growth impact clonal diversification in a computational model and the tracerx renal study. *Nature Ecology & Evolution*, 6(1):88–102, 2022. doi: 10.1038/s41559-021-01586-x. URL https://doi.org/10.1038/s41559-021-01586-x.

[38] Marco Gerlinger, Andrew J Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl j Med*, 366:883–892, 2012.

[39] Katarina Bacevic, Robert Noble, Ahmed Soffar, Orchid Wael Ammar, Benjamin Boszonyik, Susana Prieto, Charles Vincent, Michael E Hochberg, Liliana Krasinska, and Daniel Fisher. Spatial competition constrains resistance to targeted cancer therapy. *Nature communications*, 8(1):1–15, 2017.

[40] Bartlomiej Waclaw, Ivana Bozic, Meredith E Pittman, Ralph H Hruban, Bert Vogelstein, and Martin A Nowak. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261–264, 09 2015. doi: 10.1038/nature14971. URL https://pubmed.ncbi.nlm.nih.gov/26308893.

[41] Anchal Sharma, Elise Merritt, Xiaoju Hu, Angelique Cruz, Chuan Jiang, Halle Sarkodie, Zhan Zhou, Jyoti Malhotra, Gregory M Riedlinger, and Subhajyoti De. Non-

genetic intra-tumor heterogeneity is a major predictor of phenotypic heterogeneity and ongoing evolutionary dynamics in lung tumors. *Cell reports*, 29(8):2164–2174, 2019.

[42] Stefanie Tiede, Ravi Kiran Reddy Kalathur, Fabiana Lüönd, Luca von Allmen, Barbara Maria Szczerba, Mathias Hess, Tatjana Vlajnic, Benjamin Müller, James Canales Murillo, Nicola Aceto, and Gerhard Christofori. Multi-color clonal tracking reveals intra-stage proliferative heterogeneity during mammary tumor progression. *Oncogene*, 40(1):12–27, 2021. doi: 10.1038/s41388-020-01508-4. URL https://doi.org/10.1038/s41388-020-01508-4.

[43] Piyush B Gupta, Ievgenia Pastushenko, Adam Skibinski, Cedric Blanpain, and Charlotte Kuperwasser. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. *Cell stem cell*, 24(1):65–78, 2019.

[44] Jacob Househam, Timon Heide, George D. Cresswell, Inmaculada Spiteri, Chris Kimberley, Luis Zapata, Claire Lynn, Chela James, Maximilian Mossner, Javier Fernandez-Mateos, Alessandro Vinceti, Ann-Marie Baker, Calum Gabbutt, Alison Berner, Melissa Schmidt, Bingjie Chen, Eszter Lakatos, Vinaya Gunasri, Daniel Nichol, Helena Costa, Miriam Mitchinson, Daniele Ramazzotti, Benjamin Werner, Francesco Iorio, Marnix Jansen, Giulio Caravagna, Chris P. Barnes, Darryl Shibata, John Bridgewater, Manuel Rodriguez-Justo, Luca Magnani, Andrea Sottoriva, and Trevor A. Graham. Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature*, 611(7937):744–753, 2022. doi: 10.1038/s41586-022-05311-x. URL https://doi.org/10.1038/s41586-022-05311-x.

[45] Ronald Aylmer Fisher. *The genetical theory of natural selection.* , 1958.

[46] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97, 1931.

[47] J. F. C. Kingman. On the genealogy of large populations. *Journal of Applied Probability*, 19:27–43, 1982. ISSN 00219002. URL http://www.jstor.org/stable/3213548.

[48] David G Kendall. On the generalized" birth-and-death" process. *The annals of mathematical statistics*, 19(1):1–15, 1948.

[49] Wayne P. Maddison, Peter E. Midford, and Sarah P. Otto. Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, 56(5):701–710, 11/23/2021 2007. doi: 10.1080/10635150701607033. URL https://doi.org/10.1080/10635150701607033.

[50] Tanja Stadler. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of theoretical biology*, 261(1):58–66, 2009.

[51] Ailene MacPherson, Stilianos Louca, Angela McLaughlin, Jeffrey B Joy, and Matthew W Pennell. Unifying Phylogenetic Birth–Death Models in Epidemiology and Macroevolution. *Systematic Biology*, 71(1):172–189, 06 2021. ISSN 1063-5157. doi: 10.1093/sysbio/syab049. URL https://doi.org/10.1093/sysbio/syab049.

[52] Sudhir Kumar. Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, 6(8):654–662, 2005. doi: 10.1038/nrg1659. URL https://doi.org/10.1038/nrg1659.

[53] THOMAS H. JUKES and CHARLES R. CANTOR. Chapter 24 - evolution of protein molecules. In H.N. MUNRO, editor, *Mammalian Protein Metabolism*, pages 21–132. Academic Press, 1969. ISBN 978-1-4832-3211-9. doi: https://doi.org/10.1016/B978-1-4832-3211-9.50009-7. URL https://www.sciencedirect.com/science/article/pii/B9781483232119500097.

[54] David A Rasmussen and Tanja Stadler. Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models. *eLife*, 8:e45562, aug 2019.

ISSN 2050-084X. doi: 10.7554/eLife.45562. URL https://doi.org/10.7554/eLife.45562.

[55] Nicola F Müller, David A Rasmussen, and Tanja Stadler. The structured coalescent and its approximations. *Molecular biology and evolution*, 34(11):2970–2981, 2017.

[56] Gytis Dudas, Luiz Max Carvalho, Andrew Rambaut, and Trevor Bedford. Mers-cov spillover at the camel-human interface. *eLife*, 7:e31257, jan 2018. ISSN 2050-084X. doi: 10.7554/eLife.31257. URL https://doi.org/10.7554/eLife.31257.

[57] Allison Black, Louise H. Moncla, Katherine Laiton-Donato, Barney Potter, Lissethe Pardo, Angelica Rico, Catalina Tovar, Diana P. Rojas, Ira M. Longini, M. Elizabeth Halloran, Dioselina Peláez-Carvajal, Juan D. Ramírez, Marcela Mercado-Reyes, and Trevor Bedford. Genomic epidemiology supports multiple introductions and cryptic transmission of zika virus in colombia. *BMC Infectious Diseases*, 19(1):963, 2019. doi: 10.1186/s12879-019-4566-2. URL https://doi.org/10.1186/s12879-019-4566-2.

[58] Louise H Moncla, Allison Black, Chas DeBolt, Misty Lang, Nicholas R Graff, Ailyn C Pérez-Osorio, Nicola F Müller, Dirk Haselow, Scott Lindquist, and Trevor Bedford. Repeated introductions and intensive community transmission fueled a mumps virus outbreak in washington state. *eLife*, 10:e66448, apr 2021. ISSN 2050-084X. doi: 10.7554/eLife.66448. URL https://doi.org/10.7554/eLife.66448.

[59] Nicola F. Müller, Cassia Wagner, Chris D. Frazar, Pavitra Roychoudhury, Jover Lee, Louise H. Moncla, Benjamin Pelle, Matthew Richardson, Erica Ryke, Hong Xie, Lasata Shrestha, Amin Addetia, Victoria M. Rachleff, Nicole A. P. Lieberman, Meei-Li Huang, Romesh Gautom, Geoff Melly, Brian Hiatt, Philip Dykema, Amanda Adler, Elisabeth Brandstetter, Peter D. Han, Kairsten Fay, Misja Ilcisin, Kirsten Lacombe, Thomas R. Sibley, Melissa Truong, Caitlin R. Wolf, Michael Boeckh, Janet A. Englund, Michael

Famulare, Barry R. Lutz, Mark J. Rieder, Matthew Thompson, Jeffrey S. Duchin, Lea M. Starita, Helen Y. Chu, Jay Shendure, Keith R. Jerome, Scott Lindquist, Alexander L. Greninger, Deborah A. Nickerson, and Trevor Bedford. Viral genomes reveal patterns of the sars-cov-2 outbreak in washington state. *Science Translational Medicine*, 13(595):eabf0202, 2023/01/13 2021. doi: 10.1126/scitranslmed.abf0202. URL https://doi.org/10.1126/scitranslmed.abf0202.

[60] Boris Igic and Jeremiah W. Busch. Is self-fertilization an evolutionary dead end? *New Phytologist*, 198(2):386–397, 2013. doi: https://doi.org/10.1111/nph.12182. URL https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.12182.

[61] Tracy A Heath, John P Huelsenbeck, and Tanja Stadler. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111(29):E2957–E2966, 2014.

[62] Zaira Seferbekova, Artem Lomakin, Lucy R. Yates, and Moritz Gerstung. Spatial biology of cancer evolution. *Nature Reviews Genetics*, 2022. doi: 10.1038/s41576-022-00553-x. URL https://doi.org/10.1038/s41576-022-00553-x.

[63] Colin S Cooper, Rosalind Eeles, David C Wedge, Peter Van Loo, Gunes Gundem, Ludmil B Alexandrov, Barbara Kremeyer, Adam Butler, Andrew G Lynch, Niedzica Camacho, Charlie E Massie, Jonathan Kay, Hayley J Luxton, Sandra Edwards, Zsofia Kote-Jarai, Nening Dennis, Sue Merson, Daniel Leongamornlert, Jorge Zamora, Cathy Corbishley, Sarah Thomas, Serena Nik-Zainal, Manasa Ramakrishna, Sarah O'Meara, Lucy Matthews, Jeremy Clark, Rachel Hurst, Richard Mithen, Robert G Bristow, Paul C Boutros, Michael Fraser, Susanna Cooke, Keiran Raine, David Jones, Andrew Menzies, Lucy Stebbings, Jon Hinton, Jon Teague, Stuart McLaren, Laura Mudie, Claire Hardy, Elizabeth Anderson, Olivia Joseph, Victoria Goody, Ben Robinson, Mark Maddison, Stephen Gamble, Christopher Greenman, Dan Berney, Steven

Hazell, Naomi Livni, Cyril Fisher, Christopher Ogden, Pardeep Kumar, Alan Thompson, Christopher Woodhouse, David Nicol, Erik Mayer, Tim Dudderidge, Nimish C Shah, Vincent Gnanapragasam, Thierry Voet, Peter Campbell, Andrew Futreal, Douglas Easton, Anne Y Warren, Christopher S Foster, Michael R Stratton, Hayley C Whitaker, Ultan McDermott, Daniel S Brewer, David E Neal, and the ICGC Prostate Group. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nature Genetics*, 47(4):367–372, 2015. doi: 10.1038/ng.3221. URL https://doi.org/10.1038/ng.3221.

[64] Michael C. Haffner, Timothy Mosbruger, David M. Esopi, Helen Fedor, Christopher M. Heaphy, David A. Walker, Nkosi Adejola, Meltem Gürel, Jessica Hicks, Alan K. Meeker, Marc K. Halushka, Jonathan W. Simons, William B. Isaacs, Angelo M. De Marzo, William G. Nelson, and Srinivasan Yegnasubramanian. Tracking the clonal origin of lethal prostate cancer. *The Journal of Clinical Investigation*, 123(11):4918–4922, 11 2013. doi: 10.1172/JCI70354. URL https://www.jci.org/articles/view/70354.

[65] Gunes Gundem, Peter Van Loo, Barbara Kremeyer, Ludmil B. Alexandrov, Jose M. C. Tubio, Elli Papaemmanuil, Daniel S. Brewer, Heini M. L. Kallio, Gunilla Högnäs, Matti Annala, Kati Kivinummi, Victoria Goody, Calli Latimer, Sarah O'Meara, Kevin J. Dawson, William Isaacs, Michael R. Emmert-Buck, Matti Nykter, Christopher Foster, Zsofia Kote-Jarai, Douglas Easton, Hayley C. Whitaker, David E. Neal, Colin S. Cooper, Rosalind A. Eeles, Tapio Visakorpi, Peter J. Campbell, Ultan McDermott, David C. Wedge, G. Steven Bova, and ICGC Prostate UK Group. The evolutionary history of lethal metastatic prostate cancer. *Nature*, 520(7547):353–357, 2015. doi: 10.1038/nature14347. URL https://doi.org/10.1038/nature14347.

[66] Kamila Naxerova, Johannes G. Reiter, Elena Brachtel, Jochen K. Lennerz, Marc

van de Wetering, Andrew Rowan, Tianxi Cai, Hans Clevers, Charles Swanton, Martin A. Nowak, Stephen J. Elledge, and Rakesh K. Jain. Origins of lymphatic and distant metastases in human colorectal cancer. *Science*, 357(6346):55–60, 2017. doi: 10.1126/science.aai8515. URL https://www.science.org/doi/abs/10.1126/science.aai8515.

[67] H. P. Greenspan. Models for the growth of a solid tumor by diffusion. *Studies in Applied Mathematics*, 51(4):317–340, 1972. doi: https://doi.org/10.1002/sapm1972514317. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm1972514317.

[68] J P Freyer and R M Sutherland. Proliferative and clonogenic heterogeneity of cells from EMT6/Ro multicellular spheroids induced by the glucose and oxygen supply. *Cancer Res*, 46(7):3513–3520, July 1986.

[69] J P Freyer and R M Sutherland. Regulation of growth saturation and development of necrosis in EMT6/Ro multicellular spheroids by the glucose and oxygen supply. *Cancer Res*, 46(7):3504–3512, July 1986.

[70] J. P. Ward and J. R. King. Mathematical modelling of avascular-tumour growth. *Mathematical Medicine and Biology: A Journal of the IMA*, 14(1):39–69, 03 1997. ISSN 1477-8599. doi: 10.1093/imammb/14.1.39. URL https://doi.org/10.1093/imammb/14.1.39.

[71] Christian A Petrulio, Seunghee Kim-Schulze, and Howard L Kaufman. The tumour microenvironment and implications for cancer immunotherapy. *Expert Opinion on Biological Therapy*, 6(7):671–684, 2006. doi: 10.1517/14712598.6.7.671. URL https://doi.org/10.1517/14712598.6.7.671. PMID: 16805707.

[72] Andriy Marusyk, Doris P Tabassum, Philipp M Altrock, Vanessa Almendro, Franziska

Michor, and Kornelia Polyak. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520):54–58, July 2014.

[73] Kristiaan J Lenos, Daniël M Miedema, Sophie C Lodestijn, Lisanne E Nijman, Tom van den Bosch, Xavier Romero Ros, Filipe C Lourenço, Maria C Lecca, Maartje van der Heijden, Sanne M van Neerven, et al. Stem cell functionality is microenvironmentally defined during tumour expansion and therapy response in colon cancer. *Nature cell biology*, 20(10):1193–1202, 2018.

[74] Christopher A Edmonds, Anita S Lillie, and L Luca Cavalli-Sforza. Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences*, 101(4):975–979, 2004.

[75] Seraina Klopfstein, Mathias Currat, and Laurent Excoffier. The fate of mutations surfing on the wave of a range expansion. *Molecular biology and evolution*, 23(3): 482–490, 2006.

[76] Oskar Hallatschek, Pascal Hersen, Sharad Ramanathan, and David R Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, 2007.

[77] Kirill S Korolev, Melanie JI Müller, Nilay Karahan, Andrew W Murray, Oskar Hallatschek, and David R Nelson. Selective sweeps in growing microbial colonies. *Physical biology*, 9(2):026008, 2012.

[78] Matti Gralka, Fabian Stiewe, Fred Farrell, Wolfram Möbius, Bartlomiej Waclaw, and Oskar Hallatschek. Allele surfing promotes microbial adaptation from standing variation. *Ecology letters*, 19(8):889–898, 2016.

[79] Bartlomiej Waclaw, Ivana Bozic, Meredith E Pittman, Ralph H Hruban, Bert Vogel-

stein, and Martin A Nowak. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261, 2015.

[80] Ruping Sun, Zheng Hu, Andrea Sottoriva, Trevor A Graham, Arbel Harpak, Zhicheng Ma, Jared M Fischer, Darryl Shibata, and Christina Curtis. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature genetics*, 49(7): 1015, 2017.

[81] Zafarali Ahmed and Simon Gravel. Intratumor heterogeneity and circulating tumor cell clusters. *Molecular biology and evolution*, 35(9):2135–2144, 2018.

[82] Ketevan Chkhaidze, Timon Heide, Benjamin Werner, Marc J. Williams, Weini Huang, Giulio Caravagna, Trevor A. Graham, and Andrea Sottoriva. Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. *PLOS Computational Biology*, 15(7):e1007243–, 07 2019. URL https://doi.org/10.1371/journal.pcbi.1007243.

[83] Jona Kayser, Carl F Schreck, Matti Gralka, Diana Fusco, and Oskar Hallatschek. Collective motion conceals fitness differences in crowded cellular populations. *Nature ecology & evolution*, 3(1):125–134, 2019.

[84] Maximilian AR Strobl, Jill Gallaher, Jeffrey West, Mark Robertson-Tessi, Philip K Maini, and Alexander RA Anderson. Spatial structure impacts adaptive therapy by shaping intra-tumoral competition. *Communications medicine*, 2(1):1–18, 2022.

[85] Diana Fusco, Matti Gralka, Jona Kayser, Alex Anderson, and Oskar Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial luria–delbrück experiments. *Nature communications*, 7(1):1–9, 2016.

[86] Marc J Williams, Benjamin Werner, Chris P Barnes, Trevor A Graham, and Andrea

Sottoriva. Identification of neutral tumor evolution across cancer types. *Nature genetics*, 48(3):238–244, 2016.

[87] Li You, Joel S. Brown, Frank Thuijsman, Jessica J. Cunningham, Robert A. Gatenby, Jingsong Zhang, and Kateřina Staňková. Spatial vs. non-spatial eco-evolutionary dynamics in a tumor growth model. *Journal of Theoretical Biology*, 435:78–97, 2017. doi: https://doi.org/10.1016/j.jtbi.2017.08.022. URL https://www.sciencedirect.com/science/article/pii/S0022519317303971.

[88] Samra Turajlic, Hang Xu, Kevin Litchfield, Andrew Rowan, Stuart Horswell, Tim Chambers, Tim O'Brien, Jose I. Lopez, Thomas B.K. Watkins, David Nicol, Mark Stares, Ben Challacombe, Steve Hazell, Ashish Chandra, Thomas J. Mitchell, Lewis Au, Claudia Eichler-Jonsson, Faiz Jabbar, Aspasia Soultati, Simon Chowdhury, Sarah Rudman, Joanna Lynch, Archana Fernando, Gordon Stamp, Emma Nye, Aengus Stewart, Wei Xing, Jonathan C. Smith, Mickael Escudero, Adam Huffman, Nik Matthews, Greg Elgar, Ben Phillimore, Marta Costa, Sharmin Begum, Sophia Ward, Max Salm, Stefan Boeing, Rosalie Fisher, Lavinia Spain, Carolina Navas, Eva Grönroos, Sebastijan Hobor, Sarkhara Sharma, Ismaeel Aurangzeb, Sharanpreet Lall, Alexander Polson, Mary Varia, Catherine Horsfield, Nicos Fotiadis, Lisa Pickering, Roland F. Schwarz, Bruno Silva, Javier Herrero, Nick M. Luscombe, Mariam Jamal-Hanjani, Rachel Rosenthal, Nicolai J. Birkbak, Gareth A. Wilson, Orsolya Pipek, Dezso Ribli, Marcin Krzystanek, Istvan Csabai, Zoltan Szallasi, Martin Gore, Nicholas McGranahan, Peter Van Loo, Peter Campbell, James Larkin, and Charles Swanton. Deterministic evolutionary trajectories influence primary tumor growth: Tracerx renal. *Cell*, 173(3):595–610.e11, 2018. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2018.03.043. URL https://www.sciencedirect.com/science/article/pii/S0092867418303751.

[89] Moritz Gerstung, Clemency Jolly, Ignaty Leshchiner, Stefan C. Dentro, Santiago Gon-

zalez, Daniel Rosebrock, Thomas J. Mitchell, Yulia Rubanova, Pavana Anur, Kaixian Yu, Maxime Tarabichi, Amit Deshwar, Jeff Wintersinger, Kortine Kleinheinz, Ignacio Vázquez-García, Kerstin Haase, Lara Jerman, Subhajit Sengupta, Geoff Macintyre, Salem Malikic, Nilgun Donmez, Dimitri G. Livitz, Marek Cmero, Jonas Demeulemeester, Steven Schumacher, Yu Fan, Xiaotong Yao, Juhee Lee, Matthias Schlesner, Paul C. Boutros, David D. Bowtell, Hongtu Zhu, Gad Getz, Marcin Imielinski, Rameen Beroukhim, S. Cenk Sahinalp, Yuan Ji, Martin Peifer, Florian Markowetz, Ville Mustonen, Ke Yuan, Wenyi Wang, Quaid D. Morris, Stefan C. Dentro, Amit G. Deshwar, David J. Adams, Paul C. Boutros, David D. Bowtell, Peter J. Campbell, Shaolong Cao, Elizabeth L. Christie, Yupeng Cun, Kevin J. Dawson, Ruben M. Drews, Roland Eils, Matthew Fittall, Dale W. Garsed, Gavin Ha, Henry Lee-Six, Dimitri G. Livitz, Inigo Martincorena, Thomas J. Mitchell, Layla Oesper, Myron Peto, Benjamin J. Raphael, S. Cenk Sahinalp, Adriana Salcedo, Ruian Shi, Seung Jun Shin, Oliver Spiro, Lincoln D. Stein, Shankar Vembu, David A. Wheeler, Tsun-Po Yang, Quaid D. Morris, Paul T. Spellman, David C. Wedge, Peter Van Loo, Paul T. Spellman, David C. Wedge, PCAWG Evolution & Heterogeneity Working Group, and PCAWG Consortium. The evolutionary history of 2,658 cancers. *Nature*, 578(7793):122–128, 2020. doi: 10.1038/s41586-019-1907-7. URL https://doi.org/10.1038/s41586-019-1907-7.

[90] Shinichi Yachida, Siân Jones, Ivana Bozic, Tibor Antal, Rebecca Leary, Baojin Fu, Mihoko Kamiyama, Ralph H. Hruban, James R. Eshleman, Martin A. Nowak, Victor E. Velculescu, Kenneth W. Kinzler, Bert Vogelstein, and Christine A. Iacobuzio-Donahue. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, 467(7319):1114–1117, 2010.

[91] Raymond H Thomlinson and Louis H Gray. The histological structure of some human lung cancers and the possible implications for radiotherapy. *British journal of cancer*, 9(4):539, 1955.

[92] Charles S. Parkins, John L. Darling, Steven S. Gill, Thomas Revesz, and David G. Thomas. Cell proliferation in serial biopsies through human malignant brain tumours: Measurement using ki67 antibody labelling. *British Journal of Neurosurgery*, 5(3): 289–298, 01 1991. doi: 10.3109/02688699109005189. URL https://doi.org/10.3109/02688699109005189.

[93] A.J.M. Connor, S.E. Pinder, C.W. Elston, J.A. Bell, P. Wencyk, J.F.R. Robertson, R.W. Blarney, R.I. Nicholson, and I.O. Ellis. Intratumoural heterogeneity of proliferation in invasive breast carcinoma evaluated with mibi antibody. *The Breast*, 6(4):171–176, 1997. ISSN 0960-9776. doi: https://doi.org/10.1016/S0960-9776(97)90568-3. URL https://www.sciencedirect.com/science/article/pii/S0960977697905683.

[94] Wolfgang Mueller-Klieser. Multicellular spheroids. *Journal of Cancer Research and Clinical Oncology*, 113(2):101–122, 1987.

[95] Stefan Florian, Yoshiko Iwamoto, Margaret Coughlin, Ralph Weissleder, and Timothy J Mitchison. A human organoid system that self-organizes to recapitulate growth and differentiation of a benign mammary tumor. *Proceedings of the National Academy of Sciences*, 116(23):11444–11453, 2019.

[96] Jennifer Laurent, Céline Frongia, Martine Cazales, Odile Mondesert, Bernard Ducommun, and Valérie Lobjois. Multicellular tumor spheroid models to explore cell cycle checkpoints in 3D. *BMC Cancer*, 13:73, February 2013.

[97] Rouven Hoefflin, Bernd Lahrmann, Gregor Warsow, Daniel Hübschmann, Cathleen Spath, Britta Walter, Xin Chen, Luisa Hofer, Stephan Macher-Goeppinger, Yanis Tolstov, Nina Korzeniewski, Anette Duensing, Carsten Grüllich, Dirk Jäger, Sven Perner, Gita Schönberg, Joanne Nyarangi-Dix, Sanjay Isaac, Gencay Hatiboglu, Dogu Teber, Boris Hadaschik, Sascha Pahernik, Wilfried Roth, Roland Eils, Matthias

Schlesner, Holger Sültmann, Markus Hohenfellner, Niels Grabe, and Stefan Duensing. Spatial niche formation but not malignant progression is a driving force for intratumoural heterogeneity. *Nature Communications*, 7(1):ncomms11845, 2016. doi: 10.1038/ncomms11845. URL https://doi.org/10.1038/ncomms11845.

[98] Soniya Bastola, Marat S. Pavlyukov, Daisuke Yamashita, Sadashib Ghosh, Heejin Cho, Noritaka Kagaya, Zhuo Zhang, Mutsuko Minata, Yeri Lee, Hirokazu Sadahiro, Shinobu Yamaguchi, Svetlana Komarova, Eddy Yang, James Markert, Louis B. Nabors, Krishna Bhat, James Lee, Qin Chen, David K. Crossman, Kazuo Shin-Ya, Do-Hyun Nam, and Ichiro Nakano. Glioma-initiating cells at tumor edge gain signals from tumor core cells to promote their malignancy. *Nature Communications*, 11(1):4660, 2020. doi: 10.1038/s41467-020-18189-y. URL https://doi.org/10.1038/s41467-020-18189-y.

[99] Emelie Berglund, Jonas Maaskola, Niklas Schultz, Stefanie Friedrich, Maja Marklund, Joseph Bergenstråhle, Firas Tarish, Anna Tanoglidi, Sanja Vickovic, Ludvig Larsson, Fredrik Salmén, Christoph Ogris, Karolina Wallenborg, Jens Lagergren, Patrik Ståhl, Erik Sonnhammer, Thomas Helleday, and Joakim Lundeberg. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications*, 9(1):2419, 2018.

[100] Liang Wu, Jiayan Yan, Yinqi Bai, Feiyu Chen, JiangShan Xu, Xuanxuan Zou, Ao Huang, Liangzhen Hou, Yu Zhong, Zehua Jing, Xiaorui Zhou, Haixiang Sun, Mengnan Cheng, Yuan Ji, Rongkui Luo, Qinqin Li, Liang Wu, Pengxiang Wang, Dezhen Guo, Waidong Huang, Junjie Lei, Sha Liao, Yuxiang Li, Zhifeng Jiang, Na Yao, Yang Yu, Yao Li, Fengming Liu, Mingyuan Zhang, Huanming Yang, Shuang Yang, Xun Xu, Longqi Liu, Xiangdong Wang, Jian Wang, Jia Fan, Shiping Liu, Xinrong Yang, Ao Chen, and Jian Zhou. Spatially-resolved transcriptomics analyses of in-

vasive fronts in solid tumors. *bioRxiv*, 2021. doi: 10.1101/2021.10.21.465135. URL https://www.biorxiv.org/content/early/2021/10/22/2021.10.21.465135.

[101] Sunny Z Wu, Ghamdan Al-Eryani, Daniel Lee Roden, Simon Junankar, Kate Harvey, Alma Andersson, Aatish Thennavan, Chenfei Wang, James R Torpy, Nenad Bartonicek, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics*, 53(9):1334–1347, 2021.

[102] Jacob Househam, Timon Heide, George D. Cresswell, Inmaculada Spiteri, Chris Kimberley, Luis Zapata, Claire Lynn, Chela James, Maximilian Mossner, Javier Fernandez-Mateos, Alessandro Vinceti, Ann-Marie Baker, Calum Gabbutt, Alison Berner, Melissa Schmidt, Bingjie Chen, Eszter Lakatos, Vinaya Gunasri, Daniel Nichol, Helena Costa, Miriam Mitchinson, Daniele Ramazzotti, Benjamin Werner, Francesco Iorio, Marnix Jansen, Giulio Caravagna, Chris P. Barnes, Darryl Shibata, John Bridgewater, Manuel Rodriguez-Justo, Luca Magnani, Andrea Sottoriva, and Trevor A. Graham. Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature*, 2022. doi: 10.1038/s41586-022-05311-x. URL https://doi.org/10.1038/s41586-022-05311-x.

[103] Sebastian Lamprecht, Eva Marina Schmidt, Cristina Blaj, Heiko Hermeking, Andreas Jung, Thomas Kirchner, and David Horst. Multicolor lineage tracing reveals clonal architecture and dynamics in colon cancer. *Nature communications*, 8(1):1406, 2017.

[104] Maartje van der Heijden, Daniël M Miedema, Bartlomiej Waclaw, Veronique L Veenstra, Maria C Lecca, Lisanne E Nijman, Erik van Dijk, Sanne M van Neerven, Sophie C Lodestijn, Kristiaan J Lenos, et al. Spatiotemporal regulation of clonogenicity in colorectal cancer xenografts. *Proceedings of the National Academy of Sciences*, 116(13): 6140–6145, 2019.

[105] Melissa Q Reeves, Eve Kandyba, Sophie Harris, Reyno Del Rosario, and Allan Balmain.

Multicolour lineage tracing reveals clonal dynamics of squamous carcinoma evolution from initiation to metastasis. *Nature cell biology*, 20(6):699, 2018.

[106] Yue Zhao, Xiao Fu, Jose I Lopez, Andrew Rowan, Lewis Au, Annika Fendler, Steve Hazell, Hang Xu, Stuart Horswell, Scott TC Shepherd, et al. Selection of metastasis competent subclones in the tumour interior. *Nature ecology & evolution*, 5(7):1033–1045, 2021.

[107] Mark C. Lloyd, Jessica J. Cunningham, Marilyn M. Bui, Robert J. Gillies, Joel S. Brown, and Robert A. Gatenby. Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely Random Mutations but Also Variable Environmental Selection Forces. *Cancer Research*, 76(11):3136–3144, 05 2016. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-15-2962. URL https://doi.org/10.1158/0008-5472.CAN-15-2962.

[108] Anna Chen, Jaclyn Sceneay, Nathan Gödde, Tanja Kinwel, Sunyoung Ham, Erik W. Thompson, Patrick O. Humbert, and Andreas Möller. Intermittent hypoxia induces a metastatic phenotype in breast cancer. *Oncogene*, 37(31):4214–4225, 2018.

[109] Panagiotis Karras, Ignacio Bordeu, Joanna Pozniak, Ada Nowosad, Cecilia Pazzi, Nina Van Raemdonck, Ewout Landeloos, Yannick Van Herck, Dennis Pedri, Greet Bervoets, Samira Makhzami, Jia Hui Khoo, Benjamin Pavie, Jochen Lamote, Oskar Marin-Bejar, Michael Dewaele, Han Liang, Xingju Zhang, Yichao Hua, Jasper Wouters, Robin Browaeys, Gabriele Bergers, Yvan Saeys, Francesca Bosisio, Joost van den Oord, Diether Lambrechts, Anil K. Rustgi, Oliver Bechter, Cedric Blanpain, Benjamin D. Simons, Florian Rambow, and Jean-Christophe Marine. A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature*, 610(7930):190–198, 2022.

[110] Andrea Sottoriva, Inmaculada Spiteri, Sara G M Piccirillo, Anestis Touloumis, V Peter

Collins, John C Marioni, Christina Curtis, Colin Watts, and Simon Tavaré. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*, 110(10):4009–4014, February 2013.

[111] Kyung In Kim and Richard Simon. Using single cell sequencing data to model the evolutionary history of a tumor. *BMC bioinformatics*, 15(1):1–13, 2014.

[112] Samra Turajlic, Hang Xu, Kevin Litchfield, Andrew Rowan, Tim Chambers, Jose I Lopez, David Nicol, Tim O'Brien, James Larkin, Stuart Horswell, et al. Tracking cancer evolution reveals constrained routes to metastases: Tracerx renal. *Cell*, 173(3): 581–594, 2018.

[113] Virginia A Turati, José Afonso Guerra-Assunção, Nicola E Potter, Rajeev Gupta, Simone Ecker, Agne Daneviciute, Maxime Tarabichi, Amy P Webster, Chuling Ding, Gillian May, et al. Chemotherapy induces canalization of cell state in childhood b-cell precursor acute lymphoblastic leukemia. *Nature Cancer*, 2(8):835–852, 2021.

[114] Marco L Leung, Alexander Davis, Ruli Gao, Anna Casasent, Yong Wang, Emi Sei, Eduardo Vilar, Dipen Maru, Scott Kopetz, and Nicholas E Navin. Single-cell dna sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome research*, 27(8):1287–1299, 2017.

[115] Zheng Hu, Jie Ding, Zhicheng Ma, Ruping Sun, Jose A Seoane, J Scott Shaffer, Carlos J Suarez, Anna S Berghoff, Chiara Cremolini, Alfredo Falcone, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nature genetics*, 51(7):1113–1122, 2019.

[116] Joao M. Alves, Sonia Prado-López, JoséManuel Cameselle-Teijeiro, and David Posada. Rapid evolution and biogeographic spread in a colorectal cancer. *Nature Communi-*

cations, 10(1):5139, 2019. doi: 10.1038/s41467-019-12926-8. URL https://doi.org/10.1038/s41467-019-12926-8.

[117] T. Stadler, O. G. Pybus, and M. P. H. Stumpf. Phylodynamics for cell biologists. *Science*, 371(6526):eaah6266, 2021. doi: 10.1126/science.aah6266. URL https://www.science.org/doi/abs/10.1126/science.aah6266.

[118] Stephen W. Attwood, Sarah C. Hill, David M. Aanensen, Thomas R. Connor, and Oliver G. Pybus. Phylogenetic and phylodynamic approaches to understanding and combating the early sars-cov-2 pandemic. *Nature Reviews Genetics*, 2022.

[119] Tanja Stadler and Sebastian Bonhoeffer. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1614):20120198, 2013.

[120] Denise Kühnert, Tanja Stadler, Timothy G. Vaughan, and Alexei J. Drummond. Phylodynamics with migration: A computational framework to quantify population structure from genomic data. *Molecular Biology and Evolution*, 33(8):2102–2116, 11/23/2021 2016. doi: 10.1093/molbev/msw064. URL https://doi.org/10.1093/molbev/msw064.

[121] Remco Bouckaert, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, Michael Matschiner, Fábio K. Mendes, Nicola F. Müller, Huw A. Ogilvie, Louis du Plessis, Alex Popinga, Andrew Rambaut, David Rasmussen, Igor Siveroni, Marc A. Suchard, Chieh-Hsi Wu, Dong Xie, Chi Zhang, Tanja Stadler, and Alexei J. Drummond. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4):1–28, 04 2019. doi: 10.1371/journal.pcbi.1006650. URL https://doi.org/10.1371/journal.pcbi.1006650.

[122] G Bard Ermentrout and Leah Edelstein-Keshet. Cellular automata approaches to biological modeling. *Journal of theoretical Biology*, 160(1):97–133, 1993.

[123] Alexander RA Anderson and Mark AJ Chaplain. Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bulletin of mathematical biology*, 60(5): 857–899, 1998.

[124] T. Vaughan. Bdmm-prime. https://github.com/tgvaughan/BDMM-Prime, 2022.

[125] Sebastian Höhna, Tanja Stadler, Fredrik Ronquist, and Tom Britton. Inferring Speciation and Extinction Rates under Different Sampling Schemes. *Molecular Biology and Evolution*, 28(9):2577–2589, 04 2011. ISSN 0737-4038. doi: 10.1093/molbev/msr095. URL https://doi.org/10.1093/molbev/msr095.

[126] Ahmadreza Ghaffarizadeh, Randy Heiland, Samuel H Friedman, Shannon M Mumenthaler, and Paul Macklin. Physicell: An open source physics-based cell simulator for 3-d multicellular systems. *PLoS computational biology*, 14(2):e1005991, 2018.

[127] Katarzyna A Rejniak and Alexander RA Anderson. Hybrid models of tumor growth. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(1):115–125, 2011.

[128] Junhwan Jeon, Vito Quaranta, and Peter T Cummings. An off-lattice hybrid discrete-continuum model of tumor growth and invasion. *Biophysical journal*, 98(1):37–47, 2010.

[129] Jonathan Ozik, Nicholson Collier, Justin M Wozniak, Charles Macal, Chase Cockrell, Samuel H Friedman, Ahmadreza Ghaffarizadeh, Randy Heiland, Gary An, and Paul Macklin. High-throughput cancer hypothesis testing with an integrated physicell-emews workflow. *BMC bioinformatics*, 19(18):81–97, 2018.

[130] Dian Yang, Matthew G Jones, Santiago Naranjo, William M Rideout III, Kyung Hoi Joseph Min, Raymond Ho, Wei Wu, Joseph M Replogle, Jennifer L Page, Jeffrey J Quinn, et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell*, 2022.

[131] Xiao Fu, Yue Zhao, Jose I Lopez, Andrew Rowan, Lewis Au, Annika Fendler, Steve Hazell, Hang Xu, Stuart Horswell, Scott TC Shepherd, et al. Spatial patterns of tumour growth impact clonal diversification in a computational model and the tracerx renal study. *Nature ecology & evolution*, 6(1):88–102, 2022.

[132] Bartomeu Colom, Maria P Alcolea, Gabriel Piedrafita, Michael WJ Hall, Agnieszka Wabik, Stefan C Dentro, Joanna C Fowler, Albert Herms, Charlotte King, Swee Hoe Ong, et al. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nature genetics*, 52(6):604–614, 2020.

[133] Rumen Kostadinov, Carlo C. Maley, and Mary K. Kuhner. Bulk genotyping of biopsies can create spurious evidence for hetereogeneity in mutation content. *PLOS Computational Biology*, 12(4):1–15, 04 2016. doi: 10.1371/journal.pcbi.1004413. URL https://doi.org/10.1371/journal.pcbi.1004413.

[134] Marc D. Ryser, Diego Mallo, Allison Hall, Timothy Hardman, Lorraine M. King, Sergei Tatishchev, Inmaculada C. Sorribes, Carlo C. Maley, Jeffrey R. Marks, E. Shelley Hwang, and Darryl Shibata. Minimal barriers to invasion during human colorectal tumor growth. *Nature Communications*, 11(1):1280, 2020.

[135] Jeffrey West, Ryan O. Schenck, Chandler Gatenbee, Mark Robertson-Tessi, and Alexander R. A. Anderson. Normal tissue architecture determines the evolutionary course of cancer. *Nature Communications*, 12(1):2060, 2021.

[136] Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

[137] Stilianos Louca and Matthew W. Pennell. Extant timetrees are consistent with a myriad of diversification histories. *Nature*, 580(7804):502–505, 2020.

[138] Stilianos Louca, Angela McLaughlin, Ailene MacPherson, Jeffrey B Joy, and Matthew W Pennell. Fundamental Identifiability Limits in Molecular Epidemiology. *Molecular Biology and Evolution*, 38(9):4010–4024, 05 2021. ISSN 1537-1719. doi: 10.1093/molbev/msab149. URL https://doi.org/10.1093/molbev/msab149.

[139] H Lote, I Spiteri, L Ermini, A Vatsiou, A Roy, A McDonald, N Maka, M Balsitis, N Bose, M Simbolo, et al. Carbon dating cancer: defining the chronology of metastatic progression in colorectal cancer. *Annals of Oncology*, 28(6):1243–1249, 2017.

[140] Zheng Hu, Zan Li, Zhicheng Ma, and Christina Curtis. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nature genetics*, 52(7):701–708, 2020.

[141] Michael Worobey, Guan-Zhu Han, and Andrew Rambaut. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature*, 508(7495):254–257, 2014.

[142] Melissa R. Junttila and Frederic J. de Sauvage. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*, 501(7467):346–354, 2013.

[143] Nathan Farrokhian, Jeff Maltas, Mina Dinh, Arda Durmaz, Patrick Ellsworth, Masahiro Hitomi, Erin McClure, Andriy Marusyk, Artem Kaznatcheev, and Jacob G Scott. Measuring competitive exclusion in non-small cell lung cancer. *bioRxiv*, 2022. doi: 10.1101/2020.09.18.303966. URL https://www.biorxiv.org/content/early/2022/04/29/2020.09.18.303966.1.

[144] Federico Abascal, Luke MR Harvey, Emily Mitchell, Andrew RJ Lawson, Stefanie V Lensing, Peter Ellis, Andrew JC Russell, Raul E Alcantara, Adrian Baez-Ortega, Yichen Wang, et al. Somatic mutation landscapes at single-molecule resolution. *Nature*, 593(7859):405–410, 2021.

[145] Kylie Chen, David Welch, and Alexei J Drummond. Ignoring errors causes inaccurate timing of single-cell phylogenies. *bioRxiv*, 2021.

[146] Jiří C. Moravec, Rob Lanfear, David L. Spector, Sarah D. Diermeier, and Alex Gavryushkin. Testing for phylogenetic signal in single-cell rna-seq data. *bioRxiv*, 2022. doi: 10.1101/2021.01.07.425804. URL https://www.biorxiv.org/content/early/2022/02/07/2021.01.07.425804.

[147] João M Alves, Tamara Prieto, and David Posada. Multiregional tumor trees are not phylogenies. *Trends in cancer*, 3(8):546–550, 2017.

[148] Giulio Caravagna, Timon Heide, Marc J Williams, Luis Zapata, Daniel Nichol, Ketevan Chkhaidze, William Cross, George D Cresswell, Benjamin Werner, Ahmet Acar, et al. Subclonal reconstruction of tumors by using machine learning and population genetics. *Nature Genetics*, 52(9):898–907, 2020.

[149] João M Alves and David Posada. Sensitivity to sequencing depth in single-cell cancer genomics. *Genome medicine*, 10(1):1–11, 2018.

[150] Bora Lim, Yiyun Lin, and Nicholas Navin. Advancing cancer research and medicine with single-cell genomics. *Cancer cell*, 37(4):456–470, 2020.

[151] Ivana Bozic, Tibor Antal, Hisashi Ohtsuki, Hannah Carter, Dewey Kim, Sining Chen, Rachel Karchin, Kenneth W Kinzler, Bert Vogelstein, and Martin A Nowak. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences*, 107(43):18545–18550, 2010.

[152] Christopher D McFarland, Kirill S Korolev, Gregory V Kryukov, Shamil R Sunyaev, and Leonid A Mirny. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A*, 110(8):2910–2915, February 2013.

[153] Cristian Tomasetti, Bert Vogelstein, and Giovanni Parmigiani. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A*, 110(6):1999–2004, January 2013.

[154] Niko Beerenwinkel, Tibor Antal, David Dingli, Arne Traulsen, Kenneth W Kinzler, Victor E Velculescu, Bert Vogelstein, and Martin A Nowak. Genetic progression and the waiting time to cancer. *PLoS computational biology*, 3(11):e225, 2007.

[155] Jochen Weile, Song Sun, Atina G Cote, Jennifer Knapp, Marta Verby, Joseph C Mellor, Yingzhou Wu, Carles Pons, Cassandra Wong, Natascha van Lieshout, et al. A framework for exhaustively mapping functional missense variants. *Molecular systems biology*, 13(12):957, 2017.

[156] Tanja Gernhard. The conditioned reconstructed process. *Journal of theoretical biology*, 253(4):769–778, 2008.

[157] Rasmus Nielsen. Mapping mutations on phylogenies. *Systematic biology*, 51(5):729–739, 2002.

[158] John P Huelsenbeck, Rasmus Nielsen, and Jonathan P Bollback. Stochastic mapping of morphological characters. *Systematic biology*, 52(2):131–158, 2003.

[159] Ugnė Stolz, Tanja Stadler, Nicola F Müller, and Timothy G Vaughan. Joint inference of migration and reassortment patterns for viruses with segmented genomes. *Molecular biology and evolution*, 39(1):msab342, 2022.

[160] Timothy G Vaughan and Alexei J Drummond. A stochastic simulator of birth–death master equations with application to phylodynamics. *Molecular biology and evolution*, 30(6):1480–1493, 2013.

[161] Morgan N. Price, Paramvir S. Dehal, and Adam P. Arkin. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, 26(7):1641–1650, 04 2009. ISSN 0737-4038. doi: 10.1093/molbev/msp077. URL https://doi.org/10.1093/molbev/msp077.

[162] John Huddleston, James Hadfield, Thomas R. Sibley, Jover Lee, Kairsten Fay, Misja Ilcisin, Elias Harkins, Trevor Bedford, Richard A. Neher, and Emma B. Hodcroft. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *Journal of Open Source Software*, 6(57):2906, 2021. doi: 10.21105/joss.02906. URL https://doi.org/10.21105/joss.02906.

[163] Guangchuang Yu, David K. Smith, Huachen Zhu, Yi Guan, and Tommy Tsan-Yuk Lam. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36, 2017. doi: https://doi.org/10.1111/2041-210X.12628. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12628.

[164] Li-Gen Wang, Tommy Tsan-Yuk Lam, Shuangbin Xu, Zehan Dai, Lang Zhou, Tingze Feng, Pingfan Guo, Casey W Dunn, Bradley R Jones, Tyler Bradley, Huachen Zhu, Yi Guan, Yong Jiang, and Guangchuang Yu. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, 37(2):599–603, 10 2019. ISSN 0737-4038. doi: 10.1093/molbev/msz240. URL https://doi.org/10.1093/molbev/msz240.

[165] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2020.

CA: A Cancer Journal for Clinicians, 70(1):7–30, 2020/11/10 2020. doi: https://doi.org/10.3322/caac.21590. URL https://doi.org/10.3322/caac.21590.

[166] Charles R Pound, Alan W Partin, Mario A Eisenberger, Daniel W Chan, Jay D Pearson, and Patrick C Walsh. Natural history of progression after psa elevation following radical prostatectomy. *Jama*, 281(17):1591–1597, 1999.

[167] EMIEL TH RUIJTER, CHRISTINA A VAN DE KAA, JACK A SCHALKEN, FRANS M DEBRUYNE, and DIRK J RUITER. Histological grade heterogeneity in multifocal prostate cancer. biological and clinical implications. *The Journal of pathology*, 180(3):295–299, 1996.

[168] Marco Gerlinger, James W. Catto, Torben F. Orntoft, Francisco X. Real, Ellen C. Zwarthoff, and Charles Swanton. Intratumour heterogeneity in urologic cancers: From molecular evidence to clinical implications. *European Urology*, 67(4):729–737, 2015. ISSN 0302-2838. doi: https://doi.org/10.1016/j.eururo.2014.04.014. URL https://www.sciencedirect.com/science/article/pii/S0302283814003959.

[169] Kristina Totland Carm, Andreas M. Hoff, Anne Cathrine Bakken, Ulrika Axcrona, Karol Axcrona, Ragnhild A. Lothe, Rolf I. Skotheim, and Marthe Løvf. Interfocal heterogeneity challenges the clinical usefulness of molecular classification of primary prostate cancer. *Scientific Reports*, 9(1):13579, 2019. doi: 10.1038/s41598-019-49964-7. URL https://doi.org/10.1038/s41598-019-49964-7.

[170] Sylvan C. Baca, Davide Prandi, Michael S. Lawrence, Juan Miguel Mosquera, Alessandro Romanel, Yotam Drier, Kyung Park, Naoki Kitabayashi, Theresa Y. MacDonald, Mahmoud Ghandi, Eliezer Van Allen, Gregory V. Kryukov, Andrea Sboner, Jean-Philippe Theurillat, T. David Soong, Elizabeth Nickerson, Daniel Auclair, Ashutosh Tewari, Himisha Beltran, Robert C. Onofrio, Gunther Boysen, Candace Guiducci,

Christopher E. Barbieri, Kristian Cibulskis, Andrey Sivachenko, Scott L. Carter, Gordon Saksena, Douglas Voet, Alex H. Ramos, Wendy Winckler, Michelle Cipicchio, Kristin Ardlie, Philip W. Kantoff, Michael F. Berger, Stacey B. Gabriel, Todd R. Golub, Matthew Meyerson, Eric S. Lander, Olivier Elemento, Gad Getz, Francesca Demichelis, Mark A. Rubin, and Levi A. Garraway. Punctuated evolution of prostate cancer genomes. *Cell*, 153(3):666–677, 2013. doi: https://doi.org/10.1016/j.cell.2013.03.021. URL https://www.sciencedirect.com/science/article/pii/S0092867413003437.

[171] Paul C Boutros, Michael Fraser, Nicholas J Harding, Richard de Borja, Dominique Trudel, Emilie Lalonde, Alice Meng, Pablo H Hennings-Yeomans, Andrew McPherson, Veronica Y Sabelnykova, Amin Zia, Natalie S Fox, Julie Livingstone, Yu-Jia Shiah, Jianxin Wang, Timothy A Beck, Cherry L Have, Taryne Chong, Michelle Sam, Jeremy Johns, Lee Timms, Nicholas Buchner, Ada Wong, John D Watson, Trent T Simmons, Christine P'ng, Gaetano Zafarana, Francis Nguyen, Xuemei Luo, Kenneth C Chu, Stephenie D Prokopec, Jenna Sykes, Alan Dal Pra, Alejandro Berlin, Andrew Brown, Michelle A Chan-Seng-Yue, Fouad Yousif, Robert E Denroche, Lauren C Chong, Gregory M Chen, Esther Jung, Clement Fung, Maud H W Starmans, Hanbo Chen, Shaylan K Govind, James Hawley, Alister D'Costa, Melania Pintilie, Daryl Waggott, Faraz Hach, Philippe Lambin, Lakshmi B Muthuswamy, Colin Cooper, Rosalind Eeles, David Neal, Bernard Tetu, Cenk Sahinalp, Lincoln D Stein, Neil Fleshner, Sohrab P Shah, Colin C Collins, Thomas J Hudson, John D McPherson, Theodorus van der Kwast, and Robert G Bristow. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nature Genetics*, 47(7):736–745, 2015. doi: 10.1038/ng.3315. URL https://doi.org/10.1038/ng.3315.

[172] Marthe Løvf, Sen Zhao, Ulrika Axcrona, Bjarne Johannessen, Anne Cathrine Bakken, Kristina Totland Carm, Andreas M. Hoff, Ola Myklebost, Leonardo A. Meza-Zepeda,

A. Kathrine Lie, Karol Axcrona, Ragnhild A. Lothe, and Rolf I. Skotheim. Multifocal primary prostate cancer exhibits high degree of genomic heterogeneity. *European Urology*, 75(3):498–505, 2019. ISSN 0302-2838. doi: https://doi.org/10.1016/j.eururo.2018.08.009. URL https://www.sciencedirect.com/science/article/pii/S0302283818305943.

[173] Michael C. Haffner, Wilbert Zwart, Martine P. Roudier, Lawrence D. True, William G. Nelson, Jonathan I. Epstein, Angelo M. De Marzo, Peter S. Nelson, and Srinivasan Yegnasubramanian. Genomic and phenotypic heterogeneity in prostate cancer. *Nature Reviews Urology*, 18(2):79–92, 2021. doi: 10.1038/s41585-020-00400-w. URL https://doi.org/10.1038/s41585-020-00400-w.

[174] Andrew Erickson, Mengxiao He, Emelie Berglund, Maja Marklund, Reza Mirzazadeh, Niklas Schultz, Linda Kvastad, Alma Andersson, Ludvig Bergenstråhle, Joseph Bergenstråhle, Ludvig Larsson, Leire Alonso Galicia, Alia Shamikh, Elisa Basmaci, Teresita Díaz De Ståhl, Timothy Rajakumar, Dimitrios Doultsinos, Kim Thrane, Andrew L. Ji, Paul A. Khavari, Firaz Tarish, Anna Tanoglidi, Jonas Maaskola, Richard Colling, Tuomas Mirtti, Freddie C. Hamdy, Dan J. Woodcock, Thomas Helleday, Ian G. Mills, Alastair D. Lamb, and Joakim Lundeberg. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature*, 608(7922):360–367, 2022. doi: 10.1038/s41586-022-05023-2. URL https://doi.org/10.1038/s41586-022-05023-2.

[175] Lei Wei, Jianmin Wang, Erika Lampert, Simon Schlanger, Adam D. DePriest, Qiang Hu, Eduardo Cortes Gomez, Mitsuko Murakam, Sean T. Glenn, Jeffrey Conroy, Carl Morrison, Gissou Azabdaftari, James L. Mohler, Song Liu, and Hannelore V. Heemers. Intratumoral and intertumoral genomic heterogeneity of multifocal localized prostate cancer impacts molecular classifications and genomic prognosticators. *European Urology*, 71(2):183–192, 2017. ISSN 0302-2838. doi: https://doi.org/10.1016/j.

eururo.2016.07.008. URL https://www.sciencedirect.com/science/article/pii/S0302283816304067.

[176] Akash Kumar, Ilsa Coleman, Colm Morrissey, Xiaotun Zhang, Lawrence D True, Roman Gulati, Ruth Etzioni, Hamid Bolouri, Bruce Montgomery, Thomas White, Jared M Lucas, Lisha G Brown, Ruth F Dumpit, Navonil DeSarkar, Celestia Higano, Evan Y Yu, Roger Coleman, Nikolaus Schultz, Min Fang, Paul H Lange, Jay Shendure, Robert L Vessella, and Peter S Nelson. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nature Medicine*, 22(4):369–378, 2016. doi: 10.1038/nm.4053. URL https://doi.org/10.1038/nm.4053.

[177] J Clark, S Merson, S Jhavar, P Flohr, S Edwards, CS Foster, R Eeles, Frank L Martin, DH Phillips, M Crundwell, et al. Diversity of tmprss2-erg fusion transcripts in the human prostate. *Oncogene*, 26(18):2667–2673, 2007.

[178] Rohit Mehra, Bo Han, Scott A Tomlins, Lei Wang, Anjana Menon, Matthew J Wasco, Ronglai Shen, James E Montie, Arul M Chinnaiyan, and Rajal B Shah. Heterogeneity of tmprss2 gene rearrangements in multifocal prostate adenocarcinoma: molecular evidence for an independent group of diseases. *Cancer research*, 67(17):7991–7995, 2007.

[179] Shadrielle Melijah G. Espiritu, Lydia Y. Liu, Yulia Rubanova, Vinayak Bhandari, Erle M. Holgersen, Lesia M. Szyca, Natalie S. Fox, Melvin L. K. Chua, Takafumi N. Yamaguchi, Lawrence E. Heisler, Julie Livingstone, Jeff Wintersinger, Fouad Yousif, Emilie Lalonde, Alexandre Rouette, Adriana Salcedo, Kathleen E. Houlahan, Constance H. Li, Vincent Huang, Michael Fraser, Theodorus van der Kwast, Quaid D. Morris, Robert G. Bristow, and Paul C. Boutros. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell*, 173(4):1003–1013.e15, 2018.

doi: https://doi.org/10.1016/j.cell.2018.03.029. URL http://www.sciencedirect.com/science/article/pii/S009286741830309X.

[180] Nicholas McGranahan and Charles Swanton. Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell*, 168(4):613–628, 2020/11/10 2017. doi: 10.1016/j.cell.2017.01.018. URL https://doi.org/10.1016/j.cell.2017.01.018.

[181] Jingsong Zhang, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications*, 8(1):1816, 2017. doi: 10.1038/s41467-017-01968-5. URL https://doi.org/10.1038/s41467-017-01968-5.

[182] N N Nupponen, L Kakkola, P Koivisto, and T Visakorpi. Genetic alterations in hormone-refractory recurrent prostate carcinomas. 153(1):141–148, 1998. doi: 10.1016/S0002-9440(10)65554-X. URL http://dx.doi.org/10.1016/S0002-9440(10)65554-X.

[183] Zhihu Ding, Chang-Jiun Wu, Gerald C Chu, Yonghong Xiao, Dennis Ho, Jingfang Zhang, Samuel R Perry, Emma S Labrot, Xiaoqiu Wu, Rosina Lis, et al. Smad4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature*, 470(7333):269–273, 2011.

[184] Goutam Chakraborty, Joshua Armenia, Ying Z. Mazzu, Subhiksha Nandakumar, Konrad H. Stopsack, Mohammad O. Atiq, Kazumasa Komura, Lina Jehane, Rahim Hirani, Kalyani Chadalavada, Yuki Yoshikawa, Nabeela A. Khan, Yu Chen, Wassim Abida, Lorelei A. Mucci, Gwo-Shu Mary Lee, Gouri J. Nanjangud, and Philip W. Kantoff. Significance of BRCA2 and RB1 Co-loss in Aggressive Prostate Cancer Progression. *Clinical Cancer Research*, 26(8):2047–2064, 04 2020. ISSN 1078-0432.

doi: 10.1158/1078-0432.CCR-19-1570. URL https://doi.org/10.1158/1078-0432.CCR-19-1570.

[185] Dev Karan, Ming-Fong Lin, Sonny L Johansson, and Surinder K Batra. Current status of the molecular genetics of human prostatic adenocarcinomas. 103(3):285–293, 2003. doi: 10.1002/ijc.10813. URL http://dx.doi.org/10.1002/ijc.10813.

[186] Katja Lundgren, Karolina Holm, Bo Nordenskjöld, Åke Borg, and Göran Landberg. Gene products of chromosome 11q and their association with ccnd1gene amplification and tamoxifen resistance in premenopausal breast cancer. *Breast Cancer Research*, 10 (5):R81, 2008. doi: 10.1186/bcr2150. URL https://doi.org/10.1186/bcr2150.

[187] Xintao Qiu, Nadia Boufaied, Tarek Hallal, Avery Feit, Anna de Polo, Adrienne M. Luoma, Walaa Alahmadi, Janie Larocque, Giorgia Zadra, Yingtian Xie, Shengqing Gu, Qin Tang, Yi Zhang, Sudeepa Syamala, Ji-Heui Seo, Connor Bell, Edward O'Connor, Yang Liu, Edward M. Schaeffer, R. Jeffrey Karnes, Sheila Weinmann, Elai Davicioni, Colm Morrissey, Paloma Cejas, Leigh Ellis, Massimo Loda, Kai W. Wucherpfennig, Mark M. Pomerantz, Daniel E. Spratt, Eva Corey, Matthew L. Freedman, X. Shirley Liu, Myles Brown, Henry W. Long, and David P. Labbé. Myc drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets. *Nature Communications*, 13(1):2559, 2022. doi: 10.1038/s41467-022-30257-z. URL https://doi.org/10.1038/s41467-022-30257-z.

[188] Yun Peng, Yuxuan Song, and Haitao Wang. Systematic elucidation of the aneuploidy landscape and identification of aneuploidy driver genes in prostate cancer. *Frontiers in Cell and Developmental Biology*, 9:723466, 2022.

[189] Oliver Richmond, Maryam Ghotbaddini, Cidney Allen, Alice Walker, Shokouh Zahir, and Joann B. Powell. The aryl hydrocarbon receptor is constitutively active in ad-

vanced prostate cancer cells. *PLOS ONE*, 9(4):1–11, 04 2014. doi: 10.1371/journal.pone.0095058. URL https://doi.org/10.1371/journal.pone.0095058.

[190] Nitin Patel, Tatsuo Itakura, Shinwu Jeong, Chun-Peng Liao, Pradip Roy-Burman, Ebrahim Zandi, Susan Groshen, Jacek Pinski, Gerhard A. Coetzee, Mitchell E. Gross, and M. Elizabeth Fini. Expression and functional role of orphan receptor gpr158 in prostate cancer growth and progression. *PLOS ONE*, 10(2):1–30, 02 2015. doi: 10.1371/journal.pone.0117758. URL https://doi.org/10.1371/journal.pone.0117758.

[191] Karim Zuhra, Catarina S. Tomé, Elena Forte, João B. Vicente, and Alessandro Giuffrè. The multifaceted roles of sulfane sulfur species in cancer-associated processes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1862(2):148338, 2021. ISSN 0005-2728. doi: https://doi.org/10.1016/j.bbabio.2020.148338. URL https://www.sciencedirect.com/science/article/pii/S0005272820301882.

[192] Juan M. Jiménez-Vacas, Vicente Herrero-Aguayo, Antonio J. Montero-Hidalgo, Enrique Gómez-Gómez, Antonio C. Fuentes-Fayos, Antonio J. León-González, Prudencio Sáez-Martínez, Emilia Alors-Pérez, Sergio Pedraza-Arévalo, Teresa González-Serrano, Oscar Reyes, Ana Martínez-López, Rafael Sánchez-Sánchez, Sebastián Ventura, Elena M. Yubero-Serrano, María J. Requena-Tapia, Justo P. Castaño, Manuel D. Gahete, and Raúl M. Luque. Dysregulation of the splicing machinery is directly associated to aggressiveness of prostate cancer. *eBioMedicine*, 51, 2022/08/17 2020. doi: 10.1016/j.ebiom.2019.11.008. URL https://doi.org/10.1016/j.ebiom.2019.11.008.

[193] Diping Wang, R. Bruce Montgomery, Lucy J. Schmidt, Elahe A. Mostaghel, Haojie Huang, Peter S. Nelson, and Donald J. Tindall. Reduced Tumor Necrosis Factor Receptor–Associated Death Domain Expression Is Associated with Prostate Cancer Progression. *Cancer Research*, 69(24):9448–9456, 12 2009. ISSN 0008-5472.

doi: 10.1158/0008-5472.CAN-09-1903. URL https://doi.org/10.1158/0008-5472.CAN-09-1903.

[194] Sierra Gillis and Andrew Roth. Pyclone-vi: scalable inference of clonal population structures using whole genome data. *BMC Bioinformatics*, 21(1):571, 2020. doi: 10.1186/s12859-020-03919-2. URL https://doi.org/10.1186/s12859-020-03919-2.

[195] Jeff A. Wintersinger, Stephanie M. Dobson, Ethan Kulman, Lincoln D. Stein, John E. Dick, and Quaid Morris. Reconstructing Complex Cancer Evolutionary Histories from Multiple Bulk DNA Samples Using Pairtree. *Blood Cancer Discovery*, 3(3):208–219, 05 2022. ISSN 2643-3230. doi: 10.1158/2643-3230.BCD-21-0092. URL https://doi.org/10.1158/2643-3230.BCD-21-0092.

[196] Joseph Felsenstein. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American journal of human genetics*, 25(5):471, 1973.

[197] Philippe Lemey, Andrew Rambaut, John J Welch, and Marc A Suchard. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol*, 27 (8):1877–1885, March 2010.

[198] Ersin Konyalioglu, Huseyin Tarhan, Ozgur Cakmak, Emel Ebru Pala, and Ferruh Zorlu. Prostate cancer volume estimations based on transrectal ultrasonography-guided biopsy in order to predict clinically significant prostate cancer. *International braz j urol : official journal of the Brazilian Society of Urology*, 41(3):442–448, May-Jun 2015. doi: 10.1590/S1677-5538.IBJU.2014.0251. URL https://pubmed.ncbi.nlm.nih.gov/26200537.

[199] Giseok Kang, Young-Jae Kim, Hong-Sang Moon, Jeong-Woo Lee, Tag-Keun Yoo, Kwangsung Park, and Jong-Hyun Lee. Discrimination between the human prostate

normal cell and cancer cell by using a novel electrical impedance spectroscopy controlling the cross-sectional area of a microfluidic channel. *Biomicrofluidics*, 7(4):44126–44126, 08 2013. doi: 10.1063/1.4818838. URL https://pubmed.ncbi.nlm.nih.gov/24404059.

[200] Ryan Poplin, Valentin Ruano-Rubio, Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, Laura D. Gauthier, Ami Levy-Moonshine, David Roazen, Khalid Shakir, Joel Thibault, Sheila Chandran, Chris Whelan, Monkol Lek, Stacey Gabriel, Mark J Daly, Ben Neale, Daniel G. MacArthur, and Eric Banks. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 2018. doi: 10.1101/201178. URL https://www.biorxiv.org/content/early/2018/07/24/201178.

[201] Valerie Obenchain, Michael Lawrence, Vincent Carey, Stephanie Gogarten, Paul Shannon, and Martin Morgan. VariantAnnotation : a Bioconductor package for exploration and annotation of genetic variants . *Bioinformatics*, 30(14):2076–2078, 03 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu168. URL https://doi.org/10.1093/bioinformatics/btu168.

[202] Simone Zaccaria and Benjamin J. Raphael. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nature Communications*, 11(1):4301, 2020. doi: 10.1038/s41467-020-17967-y. URL https://doi.org/10.1038/s41467-020-17967-y.

[203] Remco Bouckaert, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, Michael Matschiner, Fábio K. Mendes, Nicola F. Müller, Huw A. Ogilvie, Louis du Plessis, Alex Popinga, Andrew Rambaut, David Rasmussen, Igor Siveroni, Marc A. Suchard, Chieh-Hsi Wu, Dong Xie, Chi Zhang, Tanja Stadler,

and Alexei J. Drummond. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4):1–28, 04 2019. doi: 10.1371/journal.pcbi.1006650. URL https://doi.org/10.1371/journal.pcbi.1006650.

[204] Cristian Tomasetti, Lu Li, and Bert Vogelstein. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*, 355(6331):1330–1334, 2017. doi: 10.1126/science.aaf9011. URL https://www.science.org/doi/abs/10.1126/science.aaf9011.

[205] Michael S Lawrence, Petar Stojanov, Craig H Mermel, James T Robinson, Levi A Garraway, Todd R Golub, Matthew Meyerson, Stacey B Gabriel, Eric S Lander, and Gad Getz. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 505(7484):495–501, 01 2014. doi: 10.1038/nature12912. URL https://pubmed.ncbi.nlm.nih.gov/24390350.

[206] Andrew Erickson, Mengxiao He, Emelie Berglund, Maja Marklund, Reza Mirzazadeh, Niklas Schultz, Linda Kvastad, Alma Andersson, Ludvig Bergenstråhle, Joseph Bergenstråhle, Ludvig Larsson, Leire Alonso Galicia, Alia Shamikh, Elisa Basmaci, Teresita Díaz De Ståhl, Timothy Rajakumar, Dimitrios Doultsinos, Kim Thrane, Andrew L. Ji, Paul A. Khavari, Firaz Tarish, Anna Tanoglidi, Jonas Maaskola, Richard Colling, Tuomas Mirtti, Freddie C. Hamdy, Dan J. Woodcock, Thomas Helleday, Ian G. Mills, Alastair D. Lamb, and Joakim Lundeberg. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature*, 608(7922):360–367, 2022. doi: 10.1038/s41586-022-05023-2. URL https://doi.org/10.1038/s41586-022-05023-2.

[207] Cindy H. Chau and William Figg, II. Molecular and phenotypic heterogeneity of metastatic prostate cancer. *Cancer Biology & Therapy*, 4(2):174–175, 02 2005. doi: 10.4161/cbt.4.2.1571. URL https://doi.org/10.4161/cbt.4.2.1571.

[208] Jonathan Shoag and Christopher E Barbieri. Clinical variability and molecular heterogeneity in prostate cancer. *Asian journal of andrology*, 18(4):543–548, Jul-Aug 2016. doi: 10.4103/1008-682X.178852. URL https://pubmed.ncbi.nlm.nih.gov/27080479.

[209] Emelie Berglund, Jonas Maaskola, Niklas Schultz, Stefanie Friedrich, Maja Marklund, Joseph Bergenstråhle, Firas Tarish, Anna Tanoglidi, Sanja Vickovic, Ludvig Larsson, Fredrik Salmén, Christoph Ogris, Karolina Wallenborg, Jens Lagergren, Patrik Ståhl, Erik Sonnhammer, Thomas Helleday, and Joakim Lundeberg. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications*, 9(1):2419, 2018. doi: 10.1038/s41467-018-04724-5. URL https://doi.org/10.1038/s41467-018-04724-5.

[210] Yuliang Wang, Shuyi Ma, and Walter L. Ruzzo. Spatial modeling of prostate cancer metabolic gene expression reveals extensive heterogeneity and selective vulnerabilities. *Scientific Reports*, 10(1):3490, 2020. doi: 10.1038/s41598-020-60384-w. URL https://doi.org/10.1038/s41598-020-60384-w.

[211] Lauren Brady, Michelle Kriner, Ilsa Coleman, Colm Morrissey, Martine Roudier, Lawrence D. True, Roman Gulati, Stephen R. Plymate, Zoey Zhou, Brian Birditt, Rhonda Meredith, Gary Geiss, Margaret Hoang, Joseph Beechem, and Peter S. Nelson. Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling. 12(1):1426, 2021. doi: 10.1038/s41467-021-21615-4. URL https://doi.org/10.1038/s41467-021-21615-4.

[212] Laura J Van't Veer, Hongyue Dai, Marc J Van De Vijver, Yudong D He, Augustinus AM Hart, Mao Mao, Hans L Peterse, Karin Van Der Kooy, Matthew J Marton, Anke T Witteveen, et al. Gene expression profiling predicts clinical outcome of breast cancer. *nature*, 415(6871):530–536, 2002.

[213] Wensheng Zhang and Kun Zhang. A transcriptomic signature for prostate cancer relapse prediction identified from the differentially expressed genes between tp53 mutant and wild-type tumors. *Scientific Reports*, 12(1):10561, 2022. doi: 10.1038/s41598-022-14436-y. URL https://doi.org/10.1038/s41598-022-14436-y.

[214] Joshua Armenia, Stephanie A. M. Wankowicz, David Liu, Jianjiong Gao, Ritika Kundra, Ed Reznik, Walid K. Chatila, Debyani Chakravarty, G. Celine Han, Ilsa Coleman, Bruce Montgomery, Colin Pritchard, Colm Morrissey, Christopher E. Barbieri, Himisha Beltran, Andrea Sboner, Zafeiris Zafeiriou, Susana Miranda, Craig M. Bielski, Alexander V. Penson, Charlotte Tolonen, Franklin W. Huang, Dan Robinson, Yi Mi Wu, Robert Lonigro, Levi A. Garraway, Francesca Demichelis, Philip W. Kantoff, Mary-Ellen Taplin, Wassim Abida, Barry S. Taylor, Howard I. Scher, Peter S. Nelson, Johann S. de Bono, Mark A. Rubin, Charles L. Sawyers, Arul M. Chinnaiyan, Nikolaus Schultz, Eliezer M. Van Allen, and PCF/SU2C International Prostate Cancer Dream Team. The long tail of oncogenic drivers in prostate cancer. *Nature Genetics*, 50(5):645–651, 2018. doi: 10.1038/s41588-018-0078-z. URL https://doi.org/10.1038/s41588-018-0078-z.

[215] Jacques B De Kok, Gerald W Verhaegh, Rian W Roelofs, Daphne Hessels, Lambertus A Kiemeney, Tilly W Aalders, Dorine W Swinkels, and Jack A Schalken. Dd3 pca3, a very sensitive and specific marker to detect prostate tumors. *Cancer research*, 62(9): 2695–2698, 2002.

[216] Christoph Wissmann, Peter Johannes Wild, Simone Kaiser, Stefan Roepcke, Robert Stoehr, Matthias Woenckhaus, Glen Kristiansen, Jen-Chih Hsieh, Ferdinand Hofstaedter, Arndt Hartmann, Ruth Knuechel, André Rosenthal, and Christian Pilarsky. Wif1, a component of the wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer. *The Journal of Pathology*, 201(2):204–212, 2003. doi:

https://doi.org/10.1002/path.1449. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/path.1449.

[217] Giuseppe Perrone, Bruno Vincenzi, Mariagiovanna Zagami, Daniele Santini, Roger Panteri, Gerardo Flammia, Alfio Verzì, Daniela Lepanto, Sergio Morini, Antonio Russo, Vivian Bazan, Rosa M Tomasino, Vincenza Morello, Giuseppe Tonini, and Carla Rabitti. Reelin expression in human prostate cancer: a marker of tumor aggressiveness based on correlation with grade. *Modern Pathology*, 20(3):344–351, 2023/01/27 2007. doi: 10.1038/modpathol.3800743. URL https://doi.org/10.1038/modpathol.3800743.

[218] C. Xie and G. Wang. Xage-1b cancer/testis antigen is a potential target for immunotherapy in prostate cancer. *Urologia Internationalis*, 94(3):354–362, 2015. doi: 10.1159/000363333. URL https://www.karger.com/DOI/10.1159/000363333.

[219] Leonard S Marks and David G Bostwick. Prostate cancer specificity of PCA3 gene testing: Examples from clinical practice. *Rev Urol*, 10(3):175–181, 2008.

[220] Marianna Volpert, Luc Furic, Jinghua Hu, Anne E O'Connor, Richard J Rebello, Shivakumar Keerthikumar, Jemma Evans, D Jo Merriner, John Pedersen, Gail P Risbridger, Peter McIntyre, and Moira K O'Bryan. CRISP3 expression drives prostate cancer invasion and progression. *Endocr Relat Cancer*, 27(7):415–430, July 2020.

[221] Daniel Borcard, Pierre Legendre, and Pierre Drapeau. Partialling out the spatial component of ecological variation. *Ecology*, 73(3):1045–1055, 1992.

[222] Daniel A Griffith. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer/Le Géographe canadien*, 40(4):351–367, 1996.

[223] Daniel A. Griffith and Pedro R. Peres-Neto. Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology*, 87(10):2603–2613, 2006. ISSN 00129658, 19399170. URL http://www.jstor.org/stable/20069270.

[224] Eitan Shaulian and Michael Karin. Ap-1 as a regulator of cell life and death. *Nature cell biology*, 4(5):E131–E136, 2002.

[225] Xiaoyong Fu, Rinath Jeselsohn, Resel Pereira, Emporia F Hollingsworth, Chad J Creighton, Fugen Li, Martin Shea, Agostina Nardone, Carmine De Angelis, Laura M Heiser, et al. Foxa1 overexpression mediates endocrine resistance by altering the er transcriptome and il-8 expression in er-positive breast cancer. *Proceedings of the National Academy of Sciences*, 113(43):E6600–E6609, 2016.

[226] Seung-Hyun Shin, Iljin Kim, Jae Eun Lee, Mingyu Lee, and Jong-Wan Park. Loss of egr3 is an independent risk factor for metastatic progression in prostate cancer. *Oncogene*, 39(36):5839–5854, 2020. doi: 10.1038/s41388-020-01418-5. URL https://doi.org/10.1038/s41388-020-01418-5.

[227] Richard J. Rebello, Christoph Oing, Karen E. Knudsen, Stacy Loeb, David C. Johnson, Robert E. Reiter, Silke Gillessen, Theodorus Van der Kwast, and Robert G. Bristow. Prostate cancer. *Nature Reviews Disease Primers*, 7(1):9, 2021. doi: 10.1038/s41572-020-00243-0. URL https://doi.org/10.1038/s41572-020-00243-0.

[228] Xue Shui, Rong Xu, Caiqin Zhang, Han Meng, Jumei Zhao, and Changhong Shi. Advances in neuroendocrine prostate cancer research: From model construction to molecular network analyses. *Laboratory Investigation*, 102(4):332–340, 2022. doi: 10.1038/s41374-021-00716-0. URL https://doi.org/10.1038/s41374-021-00716-0.

[229] Lauren Brady, Michelle Kriner, Ilsa Coleman, Colm Morrissey, Martine Roudier, Lawrence D. True, Roman Gulati, Stephen R. Plymate, Zoey Zhou, Brian Birditt,

Rhonda Meredith, Gary Geiss, Margaret Hoang, Joseph Beechem, and Peter S. Nelson. Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling. *Nature Communications*, 12(1):1426, 2021. doi: 10.1038/s41467-021-21615-4. URL https://doi.org/10.1038/s41467-021-21615-4.

[230] Jack Cuzick, Gregory P Swanson, Gabrielle Fisher, Arthur R Brothman, Daniel M Berney, Julia E Reid, David Mesher, VO Speights, Elzbieta Stankiewicz, Christopher S Foster, Henrik Møller, Peter Scardino, Jorja D Warren, Jimmy Park, Adib Younus, Darl D Flake, Susanne Wagner, Alexander Gutin, Jerry S Lanchbury, and Steven Stone. Prognostic value of an rna expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *The Lancet Oncology*, 12(3):245–255, 2011. ISSN 1470-2045. doi: https://doi.org/10.1016/S1470-2045(10)70295-3. URL https://www.sciencedirect.com/science/article/pii/S1470204510702953.

[231] Marc Chadeau-Hyam, Gianluca Campanella, Thibaut Jombart, Leonardo Bottolo, Lutzen Portengen, Paolo Vineis, Benoit Liquet, and Roel C H Vermeulen. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ Mol Mutagen*, 54(7):542–557, August 2013.

[232] Maria J Rosa, Mitul A Mehta, Emilio M Pich, Celine Risterucci, Fernando Zelaya, Antje A T S Reinders, Steve C R Williams, Paola Dazzan, Orla M Doyle, and Andre F Marquand. Estimating multivariate similarity between neuroimaging datasets with sparse canonical correlation analysis: an application to perfusion imaging. *Front Neurosci*, 9:366, October 2015.

[233] Florian Rohart, Benoît Gautier, Amrit Singh, and Kim-Anh Lê Cao. mixomics: An r package for 'omics feature selection and multiple data integration. *PLOS Com-*

putational Biology, 13(11):1–19, 11 2017. doi: 10.1371/journal.pcbi.1005752. URL https://doi.org/10.1371/journal.pcbi.1005752.

[234] Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell Systems*, 1(6):417–425, 2022/12/19 2015. doi: 10.1016/j.cels.2015.12.004. URL https://doi.org/10.1016/j.cels.2015.12.004.

[235] Daniel Borcard, Pierre Legendre, and Pierre Drapeau. Partialling out the spatial component of ecological variation. *Ecology*, 73(3):1045–1055, 1992. doi: https://doi.org/10.2307/1940179. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1940179.

[236] Pierre Legendre. Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, 1(1):3–8, 07 2007. ISSN 1752-9921. doi: 10.1093/jpe/rtm001. URL https://doi.org/10.1093/jpe/rtm001.

[237] Otto Warburg. On the origin of cancer cells. *Science*, 123(3191):309–314, 1956.

[238] Yang Dong, Rongfu Tu, Hudan Liu, and Guoliang Qing. Regulation of cancer cell metabolism: oncogenic myc in the driver's seat. *Signal Transduction and Targeted Therapy*, 5(1):124, 2020. doi: 10.1038/s41392-020-00235-2. URL https://doi.org/10.1038/s41392-020-00235-2.

[239] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15 (12):550, 2014. doi: 10.1186/s13059-014-0550-8. URL https://doi.org/10.1186/s13059-014-0550-8.

[240] Daniela M Witten, Robert Tibshirani, and Trevor Hastie. A penalized matrix decom-

position, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3):515–534, April 2009.

[241] Alexey A. Sergushichev. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*, 2016. doi: 10.1101/060012. URL https://www.biorxiv.org/content/early/2016/06/20/060012.

[242] S. Dray, R. Pélissier, P. Couteron, M.-J. Fortin, P. Legendre, P. R. Peres-Neto, E. Bellier, R. Bivand, F. G. Blanchet, M. De Cáceres, A.-B. Dufour, E. Heegaard, T. Jombart, F. Munoz, J. Oksanen, J. Thioulouse, and H. H. Wagner. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, 82(3):257–275, 2012. doi: https://doi.org/10.1890/11-1183.1. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/11-1183.1.

[243] Matthew E Ritchie, Belinda Phipson, DI Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.

[244] Robert A Gatenby. A change of strategy in the war on cancer. *Nature*, 459(7246): 508–509, 2009.

[245] Alfonso Santos-Lopez, Christopher W Marshall, Allison L Haas, Caroline Turner, Javier Rasero, and Vaughn S Cooper. The roles of history, chance, and natural selection in the evolution of antibiotic resistance. *eLife*, 10:e70676, aug 2021. ISSN 2050-084X. doi: 10.7554/eLife.70676. URL https://doi.org/10.7554/eLife.70676.

[246] Ivana Bozic, Johannes G Reiter, Benjamin Allen, Tibor Antal, Krishnendu Chatterjee, Preya Shah, Yo Sup Moon, Amin Yaqubie, Nicole Kelly, Dung T Le, Evan J Lipson, Paul B Chapman, Jr Diaz, Luis A, Bert Vogelstein, and Martin A Nowak. Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife*, 2:e00747, jun

2013. ISSN 2050-084X. doi: 10.7554/eLife.00747. URL https://doi.org/10.7554/eLife.00747.

[247] Yaara Oren, Michael Tsabar, Michael S. Cuoco, Liat Amir-Zilberstein, Heidie F. Cabanos, Jan-Christian Hütter, Bomiao Hu, Pratiksha I. Thakore, Marcin Tabaka, Charles P. Fulco, William Colgan, Brandon M. Cuevas, Sara A. Hurvitz, Dennis J. Slamon, Amy Deik, Kerry A. Pierce, Clary Clish, Aaron N. Hata, Elma Zaganjor, Galit Lahav, Katerina Politi, Joan S. Brugge, and Aviv Regev. Cycling cancer persister cells arise from lineages with distinct programs. *Nature*, 596(7873):576–582, 2021. doi: 10.1038/s41586-021-03796-6. URL https://doi.org/10.1038/s41586-021-03796-6.

[248] Roberto Vendramin, Kevin Litchfield, and Charles Swanton. Cancer evolution: Darwin and beyond. *The EMBO Journal*, 40(18):e108389, 2021.

[249] Jingsong Zhang, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications*, 8(1):1816, 2017.

[250] Jingsong Zhang, Jessica Cunningham, Joel Brown, and Robert Gatenby. Evolution-based mathematical models significantly prolong response to abiraterone in metastatic castrate-resistant prostate cancer and identify strategies to further improve outcomes. *eLife*, 11:e76284, jun 2022. ISSN 2050-084X. doi: 10.7554/eLife.76284. URL https://doi.org/10.7554/eLife.76284.

[251] Maximilian A. R. Strobl, Jeffrey West, Yannick Viossat, Mehdi Damaghi, Mark Robertson-Tessi, Joel S. Brown, Robert A. Gatenby, Philip K. Maini, and Alexander R. A. Anderson. Turnover modulates the need for a cost of resistance in adaptive therapy. *Cancer Research*, 81(4):1135, 02 2021. doi: 10.1158/0008-5472.CAN-20-0806. URL http://cancerres.aacrjournals.org/content/81/4/1135.abstract.

[252] Jingsong Zhang, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications*, 8(1):1816, 2017.

[253] Ahmet Acar, Daniel Nichol, Javier Fernandez-Mateos, George D. Cresswell, Iros Barozzi, Sung Pil Hong, Nicholas Trahearn, Inmaculada Spiteri, Mark Stubbs, Rosemary Burke, Adam Stewart, Giulio Caravagna, Benjamin Werner, Georgios Vlachogiannis, Carlo C. Maley, Luca Magnani, Nicola Valeri, Udai Banerji, and Andrea Sottoriva. Exploiting evolutionary steering to induce collateral drug sensitivity in cancer. *Nature Communications*, 11(1):1923, 2020. doi: 10.1038/s41467-020-15596-z. URL https://doi.org/10.1038/s41467-020-15596-z.

[254] Stephen Jay Gould. *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company, 1990.

[255] Sohrab Salehi, Farhia Kabeer, Nicholas Ceglia, Mirela Andronescu, Marc J. Williams, Kieran R. Campbell, Tehmina Masud, Beixi Wang, Justina Biele, Jazmine Brimhall, David Gee, Hakwoo Lee, Jerome Ting, Allen W. Zhang, Hoa Tran, Ciara O'Flanagan, Fatemeh Dorri, Nicole Rusk, Teresa Ruiz de Algara, So Ra Lee, Brian Yu Chieh Cheng, Peter Eirew, Takako Kono, Jenifer Pham, Diljot Grewal, Daniel Lai, Richard Moore, Andrew J. Mungall, Marco A. Marra, Gregory J. Hannon, Giorgia Battistoni, Dario Bressan, Ian Gordon Cannell, Hannah Casbolt, Atefeh Fatemi, Cristina Jauset, Tatjana Kovačević, Claire M. Mulvey, Fiona Nugent, Marta Paez Ribes, Isabella Pearsall, Fatime Qosaj, Kirsty Sawicka, Sophia A. Wild, Elena Williams, Emma Laks, Yangguang Li, Ciara H. O'Flanagan, Austin Smith, Teresa Ruiz, Andrew Roth, Shankar Balasubramanian, Maximillian Lee, Bernd Bodenmiller, Marcel Burger, Laura Kuett, Sandra Tietscher, Jonas Windhager, Edward S. Boyden, Shahar Alon, Yi Cui, Amauche Emenari, Dan Goodwin, Emmanouil D. Karagian-

nis, Anubhav Sinha, Asmamaw T. Wassie, Carlos Caldas, Alejandra Bruna, Maurizio Callari, Wendy Greenwood, Giulia Lerda, Yaniv Eyal-Lubling, Oscar M. Rueda, Abigail Shea, Owen Harris, Robby Becker, Flaminia Grimaldi, Suvi Harris, Sara Lisa Vogl, Joanna Weselak, Johanna A. Joyce, Spencer S. Watson, Ignacio Vázquez-Garćıa, Simon Tavaré, Khanh N. Dinh, Eyal Fisher, Russell Kunes, Nicholas A. Walton, Mohammad Al Sa'd, Nick Chornay, Ali Dariush, Eduardo A. González-Solares, Carlos González-Fernández, Aybüke Küpcü Yoldas, Neil Millar, Tristan Whitmarsh, Xiaowei Zhuang, Jean Fan, Hsuan Lee, Leonardo A. Sepúlveda, Chenglong Xia, Pu Zheng, Andrew McPherson, Alexandre Bouchard-Côté, Samuel Aparicio, Sohrab P. Shah, and IMAXT Consortium. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature*, 595(7868):585–590, 2021. doi: 10.1038/s41586-021-03648-3. URL https://doi.org/10.1038/s41586-021-03648-3.

[256] Leo A Featherstone, Joshua M Zhang, Timothy G Vaughan, and Sebastian Duchene. Epidemiological inference from pathogen genomes: A review of phylodynamic models and applications. *Virus Evolution*, 8(1):veac045, 2022.