FEATURE AND MODEL BASED BIOMEDICAL SYSTEM CHARACTERIZATION OF CANCER

by

Zaki Hasnain

A dissertation presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
(Mechanical Engineering)

December, 2018

Dedicated to my grandparents.
Forever in their shadow.
Forever in their light.

Acknowledgments

Every new doctoral student is terrified by the prospect of having a bad advisor. After many years of giving me intellectual freedom and timely direction in exchange for the occasional piece of chewing gum, I am proud to say my advisor Dr. Paul K. Newton is one of the best.

I also thank Dr. Peter Kuhn and Dr. Jorge Nieva for trusting me to perform in a domain where I had no previous knowledge and showing me how important academic research is for the lives of cancer patients.

After finding the perfect advisor, a doctoral student proceeds to worrying about lab-mates. I have been fortunate in this regard as well. I thank Dr. Angie Lee for setting the benchmark for composure during the latter years of doctoral studies; and thank her for keeping the thermostat fixed at 55° as it made me more productive than I ever thought was possible. I thank Dr. Jeremy Mason for always putting people first, especially in the quantitative four-cornered world where many of us take refuge from that very thing. I thank Dr. Jeffrey West for his casually serious approach to life and research, and for being by my side in the trenches.

I thank my office-mate Jonathan Hoy for being a loyal friend and being exponentially less distracting during the final year of my studies. I thank my office-mate Xiangyu Gao for his fortitude and wisdom.

I thank my teammates in the USC Taekwondo club for never asking me about my research but always leaving me ready to face my fiercest opponent for one more round.

Finally, I thank my family for reminding me to be grateful for studying in sunny southern California by drawing me away to visit them during cold winter holidays in Virginia. I also thank my family in Pakistan for reminding me that life is often nonlinear, as the temperature-productivity relationship suggested in the previous statement, among many other things, does not hold true for scorching south-Asian summers, where this story began.

Contents

List of Figures

List of Tables

Abstract

Progression of cancer occurs across multiple length and time scales and encompasses biodynamical processes with intrinsically unique implications for the host patient. The compartmentalized study of these processes has not only revealed scale-specific biomedical questions but also led to an equally diverse set of quantitative approaches derived from various fields in the applied spirit of engineering. Due to advances in imaging, experimental, and patient engagement technologies, a growing subset of cancer progression studies are empirical in nature and yield data which is often spatio-temporal. The focus of this thesis is to explore a general design approach for developing quantitative analysis tools which leverage data from these studies to provide quantified answers to biomedical questions across time and length scales, and to apply the methodology to three case studies representing three different scales.

Biodynamical processes or systems observed via experiments, clinical studies, and longitudinal medical assessments span the physical spectrum from nanometer scale molecular phenomena, human body locomotion at the meter scale, to population level disease trends. These processes may be characterized by features from statistics, mechanics, signal processing, or any domain-specific life science. The direct use of these features, optionally in combination with dimensionality reduction and distance metrics, forms the first arm of biodynamical analytical study design called the feature based approach. This is the more direct approach and leads to an intuitive understanding of a process' role in cancer progression, for instance, by comparing, correlating, and ranking samples or features themselves. The second arm of biodynamical analytical study design, the model based approach, uses stochastic, physics, or statistical modeling techniques to create models of the biodynamical processes or system features. This is the more complex design arm and requires a greater amount of data, but is distinct from the feature based approach due the more advanced questions the models are able to answer. Both approaches' ultimate goal is to provide a quantified output in the form of one or a mixture of the following: i) clustering or differentiating samples, ii) classifying samples as a type, iii) forecasting or predicting a numerical sample characteristic, and iv) characterizing the behavior of a given sample.

The first case study of morphogenesis of 40 organoids focuses on extracting geometric and signal processing based features from time series of images at the micrometer scale. The feature set is used to cluster organoids into groups of similar phenotypes in order to separate invasive and normal behaving organoids. Secondly, two model based approaches are used to understand the dynamics of growth: i) a statistical model of shape is created to corroborate the feature based clustering, and ii) a hidden Markov model of organoid shape is used to discover unmeasured genomic states and their corresponding shape dynamics. A standardized code is provided in order to apply the feature based clustering to image data from organoid

experiments.

The second case study is a human performance assessment of 37 cancer patients where a non-invasive motion capture camera is used to record patients performing clinically supervised exercises. The goal of the study is to identify patients with higher risk of deteriorating health while undergoing a chemotherapy regimen. A kinematic feature based approach is used to cluster patients into groups of low and high risk of unexpected hospitalization and limited physical activity. Secondly, a machine learning model based on kinematic and demographic features is used to classify individual patient fitness levels based on a single exercise instance. This work serves as a proof-of-concept of a non-invasive human performance based risk calculator of clinical utility.

The final case study uses a feature based approach to understand the importance of clinical measures in a longitudinal bladder cancer dataset of 3500 patients, and machine learning models to stratify patient survival and disease recurrence post-surgery. The combination of feature and model based approaches developed here can be applied to most clinical datasets which contain survival data, without additional domain specific knowledge of the cancer type. The accuracy and temporal resolution of predictions made with the developed models can dramatically change post-cystectomy patient care.

1 Introduction

1.1 Biomedical system characterization

The analytical toolsets developed in this thesis are motivated by scientific and clinical questions which arise through biomedical experiments, clinical studies, and medical assessments. Fig.1 shows the relevant spectrum of biological data sources of various length scales and the subsequent process of designing tools to produce quantified outputs. The projects presented in this thesis include data from i) organoid laboratory experiments at the μm scale, ii) clinical human motion capture studies at the m scale, and iii) preoperative imaging and operative data from surgery at a population scale. Indeed the range of biomedical data is vast (Fig.1), spanning nanometer scale molecular data to population scale demographics in space, and fractions of a second to years in time. Consequently the choice of study design is dependent, in part on whether a dataset is multi-scale or not. For instance, model based approaches including stochastic, physics, and statistical models are difficult to adapt to multi-scale data, where as feature based approaches easily combine data from various scales. Furthermore biomedical data collection methods are often fraught with high noise, which lead to lower resolution and less informative models. Nevertheless, feature based studies fail to describe or simulate biomedical behavior. Generally, feature based studies are more suitable for clinical applications where multi-modal data is ubiquitous and ultimate decision making must involve the corresponding complex higher level interactions which are difficult to model. In contrast, model based studies provide an opportunity to discover and describe biomechanisms and behavior which remain relevant to future studies. The two approaches are not entirely independent, and the ultimate goal of biomedical characterization studies is to provide a quantified summary output from a single or series of empirical observations. The next sections provide brief introductions to feature extraction (Section 1.2), dimensionality reduction (Section 1.2), model based approaches (Section 1.3), and clustering and classification (Section 1.3) topics relevant to studying biomedical datasets as shown in Fig.1.

1.2 Feature based study of biomedical data

1.2.1 Feature types

A complete or sufficient set of descriptive features of a biomedical system or process may be derived from a combination of different disciplines including biology, mechanics, geometry, information theory, time series analysis, statistics, and signal processing in addition to domain-specific biomedical parameters. Table 1 lists some of the commonly used features extracted from biomedical time series or spatio-temporal data to describe a system. Although the source disciplines of a set of features may be diverse, generally speaking the

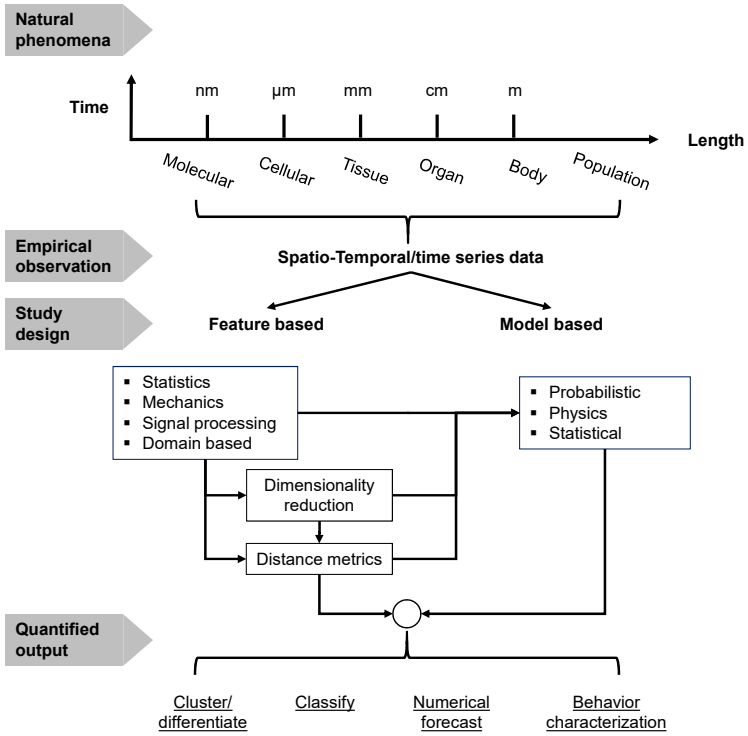


Figure 1: Schematic of analytical tool design options for studying biomedical processes of cancer progression. Steps include empirical observation of processes across various scales, feature or model based study design, and quantified output options.

features themselves can be characterized as being in either the time or frequency domain, and being scalar or time series. Furthermore, scalar features may be taken over a moving window to produce time series, e.g. minimum of every five time points. The number of feature types required to describe a system may be known *a priori*, or be discovered as part of the analysis. Nevertheless, due to the complexity of biomedical systems, often a multitude of perspectives are helpful to form a complete understanding or to discover differences among a set of samples. Once a set of features is selected, it may be used directly to draw conclusions, or may be processed with dimensionality reduction, distance metrics, or be used to train a model (Figure 1).

System feature types	
Source discipline	**Feature**
Statistics and Probability	min/max/median/mean/mode standard deviation pearson correlation moment: skewness/kurtosis/etc Kullback–Leibler divergence
Mechanics	(linear/angular) velocity (linear/angular) acceleration (linear/angular) momentum kinematic/potential energy kinetics
Geometry	perimeter/area/volume circularity curvature
Information theory	mutual information entropy
Time series analysis	ARIMA orders autoregressive parameters moving average parameters autocorrelation function
Signal processing	DFT/FFT frequencies DFT/FFT amplitude cross-correlation energy

Table 1: Examples of descriptive features for biomedical system time series and spatio-temporal data

1.2.2 Distance functions

In mathematics, a metric is a function $g : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ which defines the distance between two points $x, y \in \mathcal{X}$ subject to the following four conditions:

$$I.\ g(x, y) \geq 0 \tag{1}$$

$$II.\ g(x, y) = g(y, x) \tag{2}$$

$$III.\ g(x, y) = 0 \iff x = y \tag{3}$$

$$IV.\ g(x, z) \leq g(x, y) + g(y, z) \tag{4}$$

known as the non-negativity, symmetry, identity of indiscernibles, and the triangle inequality conditions respectively. The first and third conditions together are required for positive definiteness. Although a metric must satisfy all four conditions, the notion of distance is maintained even if a subset of conditions is met, in which case the distance measure may be classified as a quasimetric and pseudometric for example. A quasimetric is a distance function which satisfies the triangle inequality but is not symmetric, and a

pseudometric relaxes the identity condition by allowing $g(x, y) = 0, x \neq y$. A list of commonly used distance functions are cataloged in Table 2.

Time series distance functions		
Name	**formula**	**metric conditions**
ℓ_0, discrete	$\sum_i \mathbb{1}_{X(i)=Y(i)}$	$I - IV$
ℓ_1, absolute, taxi cab, city block, Manhattan	$\sum_i \lvert X(i) - Y(i)\rvert$	$I - IV$
MAE: mean absolute error	$(1/n)\sum_i \lvert X(i) - Y(i)\rvert$	$I - IV$
ℓ_2, Euclidean	$\left(\sum_i (X(i) - Y(i))^2\right)^{1/2}$	$I - IV$
SSE: sum squared error	$\sum_i (X(i) - Y(i))^2$	$I - III$
MSE: mean square error	$(1/n)\sum_i (X(i) - Y(i))^2$	$I - III$
ℓ_p	$\left(\sum_i (X(i) - Y(i))^p\right)^{1/p}$	$I - IV$
Minkowski distance	$\left(\sum_i \lvert X(i) - Y(i)\rvert^p\right)^{1/p}$	$I - IV$
Cosine similarity	$X \cdot Y / \lVert X\rVert \lVert Y\rVert$	I, II, IV
(absolute) Pearson's distance	$1 - \lvert\rho_{X,Y}\rvert$	I, II
Mahalanobis distance	$\left((X - Y)^T \Sigma^{-1} (X - Y)\right)^{1/2}$	$I - IV$

Table 2: Commonly used distance functions for comparison of real-valued time series

In practice, the choice of the optimal distance metric is not always apparent, however, empirical tests [72] suggest various metrics outperform the often standard choice of Euclidean distance. Furthermore, Keough et al showed that these metrics may be enhanced by employing correction factors which account for the discrepancy in complexity of the two time series being compared [14]. By considering time series as stochastic signals, several statistical divergence measures [13] may also be employed as distance measures [115].These functions can either provide distances directly by acting on raw time series data, the corresponding derivatives [101], and any extracted features described in section 1.2, or be used in a nested scheme such as the DTW algorithm.

1.2.3 Dynamic time warping

Misaligned and temporally deformed time series cannot be compared directly using a distance function due to mismatch between local features, and classification or clustering pipelines built on such erroneous comparisons will fail to detect or group similar sources. When two time series are out of phase, a global alignment may reduce error in the direct application of a distance function, however, subsequence offsets pose a more complex problem which cannot be solved via a global alignment. Consequently, the non-

linear mapping learned from the DTW algorithm [18] leads to a substantial reduction in sensitivity to such misalignments. In the classic DTW algorithm the similarity operators in 1.2.2 act on a pair of points (x_i, y_j) from the two time series X and Y being compared, and a warped path of matching points is found by minimizing the total distance using dynamic programming, Fig.2A.

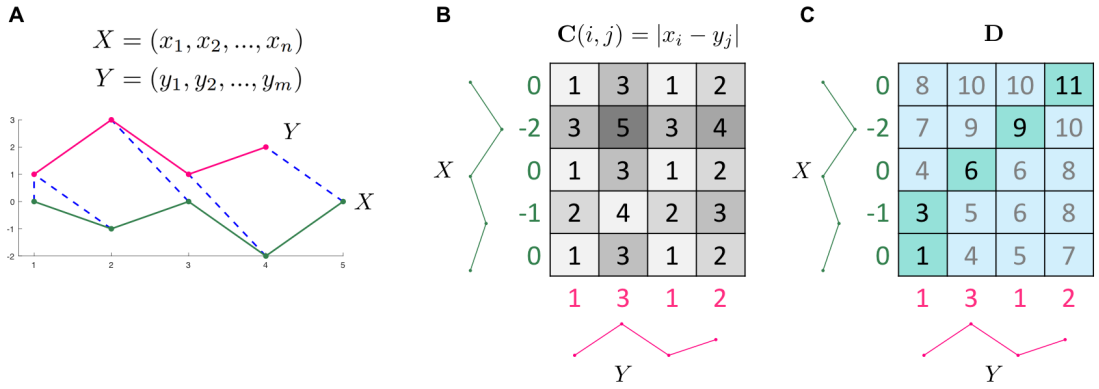


Figure 2: A) Two time series X and Y of lengths n and m respectively may be aligned using the dynamic time warping (DTW) algorithm. B) The cost matrix $\mathbf{C}$ based on the l_1 distance function. C) The accumulated cost matrix $\mathbf{D}$ created using eqs.14-16

The resulting DTW distance,

$$DTW(X,Y) = \sum_{(i,j)\in p} g(x_i, y_j) \tag{5}$$

is the sum of the distance function g over all pairs of matched points (x_i, y_j) in the optimal warping path p of length L [18]. The warping path must satisfy the i) boundary conditions,

$$p(1) = (x_1, y_1) \tag{6}$$

$$p(L) = (x_n, y_m) \tag{7}$$

the ii) monotonicity condition,

$$p(l) = (x_i, y_j) \tag{8}$$

$$i_1 \leq i_2 \leq ... \leq i_L \tag{9}$$

$$j_1 \leq j_2 \leq ... \leq j_L \tag{10}$$

$$\tag{11}$$

and iii) the step-size condition,

$$\Delta p = p(l+1) - p(l) \tag{12}$$

$$\Delta p \in \{(1,0),(0,1),(1,1)|l \in [1, L-1]\} \tag{13}$$

while assigning an element of X to one from Y. The boundary conditions prevent only a subsequence from one of the time series being used. The monotonicity conditions ensure that a feature from one time series is not repeatedly matched to the second time series, and the step-size conditions ensure that some features of one series are not entirely omitted in the warped path. Two optional, slope and window constraints, prevent matching between large and small sections and large warpings. Furthermore the search for the warping path may be altered by using local or global constraints which apply weights to certain path selections in the accumulated cost matrix.

The dynamic programming algorithm used to find the DTW distance and the warping path involves creating a cost matrix $\mathbf{C}$ of the pairwise distances (Fig.2B) and an accumulated cost matrix $\mathbf{D}$ (Fig.2C) [128], whose elements are calculated as

$$\mathbf{D}(a,1) = \sum_{i=1}^{a} g(x_i, y_1) \qquad a \in [1,n] \tag{14}$$

$$\mathbf{D}(1,b) = \sum_{j=1}^{b} g(x_1, y_j) \qquad b \in [1,m] \tag{15}$$

$$\mathbf{D}(a,b) = \min\{\mathbf{D}(a-1,b-1), \mathbf{D}(a-1,b), \mathbf{D}(a,b-1)\} + g(x_a, y_b) \qquad a \in (1,n], b \in (1,m] \tag{16}$$

where the final entry $\mathbf{D}(n,m) = DTW(X,Y)$ is the DTW distance between the time series and the warping path p is recovered by tracing back from $\mathbf{D}(n,m)$ to $\mathbf{D}(1,1)$, or, equivalently, keeping track of the choice made in the min function in eq.16. The DTW distance for two perfectly similar time series is zero, however, there is no upper bound or intrinsic scale to this measure and more dissimilar time series have larger DTW distances.

DTW satisfies the non-negativity, symmetry, and identity of indiscernibles metric conditions (eqs. 1-3) but not the triangle inequality condition (eq. 4), therefore it is not a metric. Nevertheless, it is a vastly popular similarity measure and several modified versions have been adapted to distinct applications [68].

Typically the distance metric used to find the warping path is also used to report the final distance, however, a secondary metric may be applied to the matched points such as the Malhabonis distance which in turn may be applied to neighboring subsequences of the matched points [217]. The time complexity of the dynamic programming algorithm is O(nm) when assessing similarity between two time series of lengths n

and m, however empirical results suggest O(n) complexity can be achieved using lower bounding techniques [161]. In addition to the broader tasks of clustering and classification, DTW may also be used for i) finding the average of set of time series [150], ii) sub-sequence alignment [169] which is the search of an alignment of a short time series to a much longer time series, and iii) query systems [218].

1.2.4 The multivariate case

Although unviariate time series may be compared to one another directly using the classic DTW algorithm, multivariate, or multidimensional time series require modification of the similarity measures and distance matrix, due not only to the higher dimensionality but also the correlation between those dimensions. GA ten Holt et al [89] introduced an intuitive generalization of the univariate DTW algorithm for K-dimensional multivariate time series involving the sum of the distances between all dimensions of a time series while constructing the accumulated cost matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$,

$$\mathbf{D}(a,1) = \sum_{i=1}^{a} \sum_{k=1}^{K} g(x_{i,k}, y_{1,k}) \qquad a \in [1,n] \qquad (17)$$

$$\mathbf{D}(1,b) = \sum_{j=1}^{b} \sum_{k=1}^{K} g(x_{1,k}, y_{j,k}) \qquad b \in [1,m] \qquad (18)$$

$$\mathbf{D}(a,b) = \min\{\mathbf{D}(a-1,b-1), \mathbf{D}(a-1,b), \mathbf{D}(a,b-1)\} + \sum_{k=1}^{K} g(x_{a,k}, y_{b,k}) \qquad a \in (1,n], b \in (1,m] \qquad (19)$$

and then searching for the warped path as usual, where $x_{a,k}$ and $y_{b,k}$ are the k^{th} dimensions of time series X and Y at time points a and b respectively. Here, X and Y may be of different lengths m and n, however all K dimensions of a time series must be of the same length. The K dimensions must be normalized to zero mean and unit variance due to the summation of distances between each dimension of the two time series.

This is not the only generalization of DTW for multidimensional time series. Shokoohi-Yekta et al [174] simulate and test two paradigms of DTW similarity for multivariate time series: dependent DTW and independent DTW. Dependent DTW, which is the technique proposed by GA ten Holt et al [89] and described in eqs.17-19, matches two multivariate signals by finding one warped path from a single accumulated cost matrix $\mathbf{D}$ for all the dimensions thereby generating a single distance, whereas independent DTW matches each pair of dimensions separately using K accumulated cost matrices,

$$DTW(X,Y) = \sum_{k=1}^{K} \mathbf{D}^k(n,m), \mathbf{D}^k \in \mathbb{R}^{m \times n} \qquad (20)$$

and aggregates the corresponding K pairwise warped path distances $\mathbf{D}^k(n,m)$ into a single resultant distance.

The underlying difference is that dependent DTW algorithms assume that all dimensions of the time series are warped by the same function, and independent DTW algorithms assume that each dimensions may be warped in a different way. Clearly, dependent DTW is appropriate to use when learning one nonlinear mapping between all the dimensions of two time series, and independent DTW ought to be used when a new mapping is responsible for misalignment between each dimension of two time series.

1.2.5 Dimensionality reduction and feature selection

The number of descriptive features of a biomedical system may be too large in certain cases and prove to be a hindrance towards achieving the ultimate goal of creating a simple quantified output. This high dimensionality of a system requires many samples for each combination of features to fully link the feature space to sample behavior and collecting the sufficient amount of data is nearly always impractical due to high cost of biomedical observations and experiments. Furthermore, a high dimensional scenario where the number of features in a dataset far exceeds the number of samples, leads to erroneous overfitting of prediction and classification models, and lower confidence predictions, thereby reducing clinically applicability of any analyses. These issues necessitate dimensionality reduction and feature selection techniques which reduce the total number of features required to describe a system. Although there is no optimal way of reducing the feature set, most methods either use a transformation of the entire original feature set and embed into a lower dimensional space, or build a smaller feature set using either exclusion or inclusion decision rules. The most widely used transformation technique is principal component analysis (PCA) which maximizes the variance of the data in the resulting dimensions by using a linear mapping. PCA can be performed using singular value decomposition (SVD) of a data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ with n samples and p features,

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{21}$$

to get a reduced data matrix $\mathbf{Y}_r$,

$$\mathbf{Y}_r = \mathbf{X}\mathbf{V} \in \mathbb{R}^{n \times r} \tag{22}$$

$$= \mathbf{U}\mathbf{S}\mathbf{V}^T\mathbf{V} \tag{23}$$

$$= \mathbf{U}\mathbf{S} \tag{24}$$

where $r \leq n$ is the number of principal components. Decision rule based feature selection methods use correlation metrics between features and target outcome variables to either select the most relevant features or remove the least informative features iteratively. The choice of correlation metrics depends on the nature

of the feature, for instance, categorical features may be selected using a the chi-squared test of independence and continuous variables may be ranked using Pearson's correlation.

1.3 Model based study of biomedical data

1.3.1 Modeling techniques

Mathematical modeling using biomedical data can be achieved using many different approaches as shown in Table 3. Some of these modeling techniques are tailored to very specific applications such as compartmental models for disease propagation, and rigid body kinematics for understanding biomechanical behavior, however, other more generic techniques such as time series analysis (Table 3) may be adopted to various datasets. Furthermore, statistical techniques for regression and classification such as support vector machines and random forest, require preprocessing of time series data. Whereas graphical models are more readily applicable to sequential data. Unlike feature based characterization of biomedical systems, where multiple disciplines are often combined in an analysis, model based approaches typically rely completely on a single type of a model for a given biomedical dataset. The choice of a modeling technique may be obvious if a confined system is being observed and an associated physical model exists, however, in general, model characterizations may assist this decision. Specifically, models may be characterized as i) linear or non-linear, ii) static or dynamic, iii) discrete or continuous, iv) deterministic or stochastic. Once a modeling technique is decided, the best set of model parameters can be identified using model selection techniques which optimize a selection criteria such as likelihood, error between generated and observed instances, and prediction or classification accuracy. In the following sections, statistical and graphical models which are employed in subsequent chapters of this thesis are introduced and discussed in more detail.

System model types	
Discipline	**Examples**
Statistical	regression, support vector machines, active shape model, random forest
Graphical	hidden Markov model, neural network, (switching) linear dynamical systems
Mechanics	rigid body kinematics, spring-mass models, hemodynamics
Time series analysis	ARIMA, ARCH, GARCH, EWMA
Ordinary/Parital differential equations	epidemiology compartmental models, mass action model

Table 3: Examples of models for biomedical system time series and spatio-temporal data

1.3.2 Point distribution model

Morphology is a direct indicator of phenotype in biology and can be measured using imaging technologies resulting in spatio-temporal data. In such situations, a point distribution model (PDM) [36, 37] is especially applicable as it requires only the discrete point sets which define the morphology and outputs statistical modes of geometric variation. Additionally, these models may even assist segmentation tasks when used as subroutines in active appearance models and active shape models. PDM can be used to study two-dimensional ($d = 2$) or three-dimensional ($d = 3$) shapes of length k,

$$\mathbf{Q} = \begin{cases} (\vec{x}, \vec{y})^T, & d = 2, \mathbf{Q} \in \mathbb{R}^{2k} \\ (\vec{x}, \vec{y}, \vec{z})^T, & d = 3, \mathbf{Q} \in \mathbb{R}^{3k} \end{cases} \tag{25}$$

to create a simulated model

$$\mathbf{Q}_{sim} = \bar{\mathbf{Q}} + \mathbf{P}\vec{b} \tag{26}$$

where $\bar{\mathbf{Q}}$ is the mean shape over a sequence of samples, $\mathbf{P} \in \mathbb{R}^{dk \times p}$ are the p principal modes of variation, and $\vec{b} \in \mathbb{R}^p$ are the coefficients of the principal modes. The mean shape is found by averaging M sample shapes,

$$\bar{\mathbf{Q}} = \frac{1}{M} \sum_{k}^{M} \mathbf{Q}_k, \tag{27}$$

and is used to create a covariance matrix,

$$\mathbf{S} = \frac{1}{M-1} \sum_{k}^{M} (\mathbf{Q}_k - \bar{\mathbf{Q}})(\mathbf{Q} - \bar{\mathbf{Q}})^T. \tag{28}$$

whose eigendecomposition defines the principal modes of geometric variation. Specifically the eigenvectors $\vec{u}_i$ are the principal modes of variation, and eigenvalues λ_i are the variance explained by the principal modes ($i = [1, dk]$). Using all $i = dk$ of the principal to generate new shapes $\mathbf{Q}_{sim}$ is not only numerically expensive but also not necessary as in practice a handful of modes create shapes which are sufficiently representative of the samples in the observation set $\mathbf{Q}_i, i = [1, M]$. Bounds for the principal mode coefficients can be determined from the set of observed shapes to generate shapes within an acceptable range. The benefit of using the eigendcomposition is that the principal modes of variation are orthogonal and $\mathbf{Q}_{sim}$ can represent novel combinations of these morphology dimensions resulting in shapes which may not necessarily be present

in the observation set.

Figure 3 shows an example of a point distribution model created an observation set of $M = 5$ concentric circles (Figure 3A) whose covariance matrix (Figure 3B) has one significant principal mode of variation (Figure 3C) with an eigenvalue $\lambda_1 = 1573$ and a second insignificant mode of variation with an eigenvalue $\lambda_2 = 0$. Using this single principal mode and a coefficient $\vec{b} = (b_1, 0) = (7, 0)$, a new simulated shape $\mathbf{Q}_{sim}$ is created in Figure 3D.

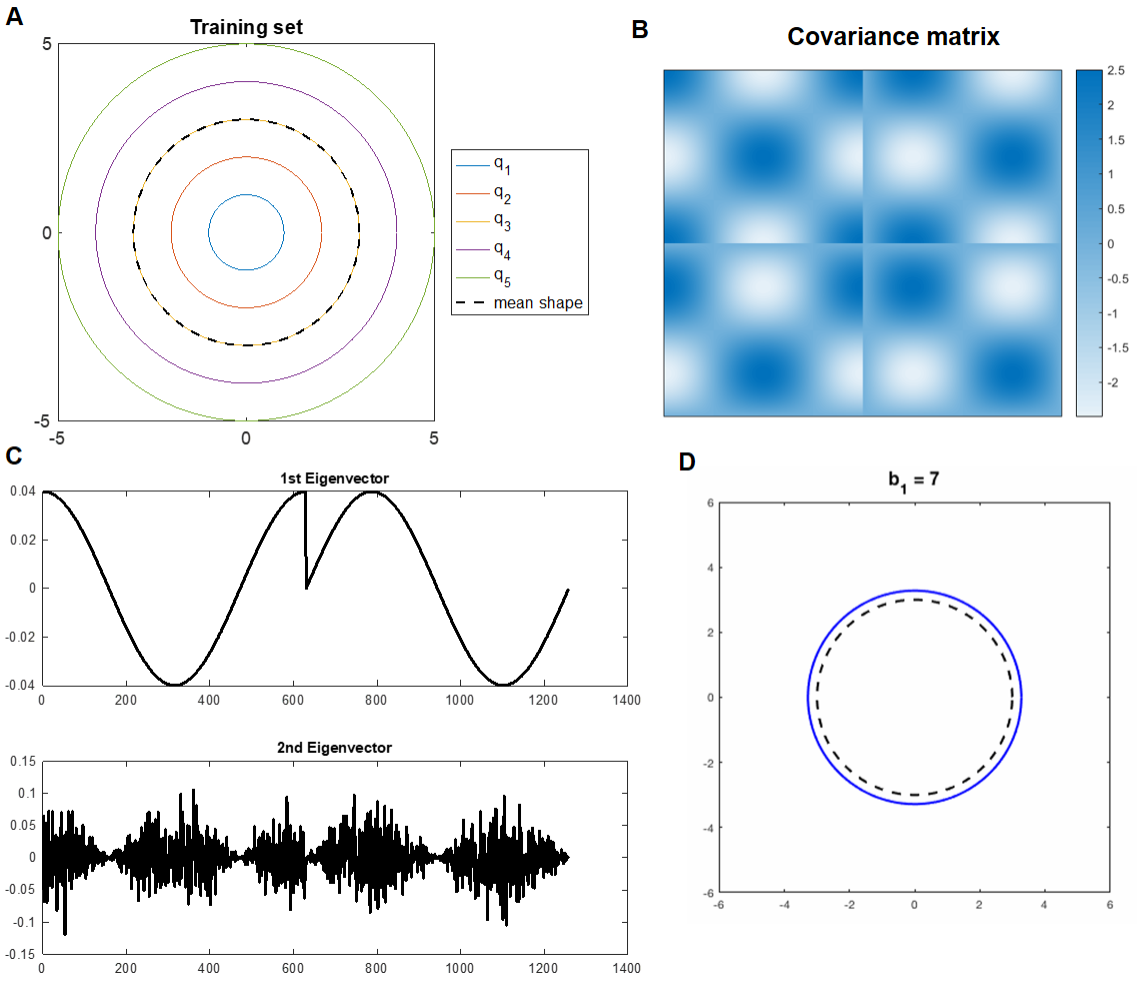


Figure 3: Example point distribution model. (A) A training set of five concentric circles. (B) The covariance matrix of the shapes. (C) Eigenvectors of the covariance matrix. (D) A simulated shape using the first eigenvector.

1.3.3 Hidden Markov model

The basic hidden Markov model [157, 156] architecture shown in Fig.4A is described by a discrete hidden random variable X_t which is an element from a set of Q hidden states, and an output observation vector Y_t

which may be discrete or continuous. Consequently, the hidden Markov model is a doubly stochastic process where the hidden generating sources evolve via a stochastic process and the observed sequence is generated via a second stochastic process conditioned on the hidden states[157]. The model parameters consist of a

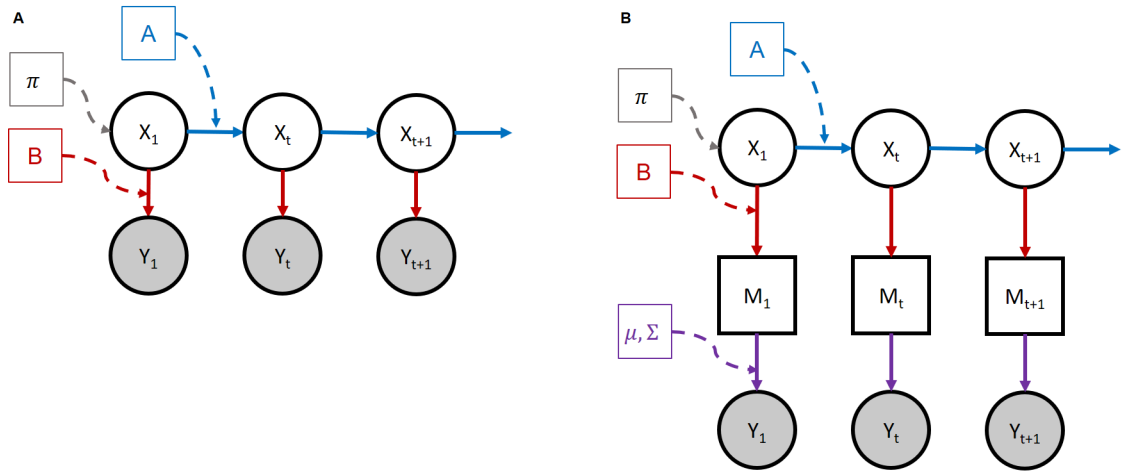


Figure 4: Hidden Markov model architecture. Hidden states X_t are shown in white circles, and observed variables Y_t are shown in solid gray circles. **A)** Standard hidden Markov model with parameters: transition matrix $\mathbf{A}$, emission distribution $\mathbf{B}$, and initial probability π. **B)** Hidden Markov model with mixture of Gaussians as emissions. In addition to Hidden Markov model parameters, the emission Gaussian mean values μ, and covariance matrices Σ are required.

stochastic matrix $\mathbf{A} \in \mathbb{R}^{Q \times Q}$ which dictates transition probabilities between Q hidden states,

$$\mathbf{A}(i,j) = P(X_t = j | X_{t-1} = i) \tag{29}$$

an emission or output probability distribution $\mathbf{B}$ which decides the output vector given a hidden state,

$$\mathbf{B} = P(Y_t | X_t) \tag{30}$$

and initial probability $\pi \in \mathbb{R}^Q$ which is a distribution over hidden states,

$$\pi = P(X_1 = j) \tag{31}$$

for a sequence of length T. The hidden state at time t depends only on the previous state X_{t-1}, and the observation at time t depends only on the hidden state at time t. Multinomial distributions are a common choice for $\mathbf{B}$ when the outputs are discrete, whereas Gaussian and exponential distributions are popular choices for $\mathbf{B}$ in the case of continuous outputs. More complex distributions, and consequently more robust

outputs can be modeled by using a mixture of Gaussians with different mean and covariance matrices as in Gaussian mixture models. A hidden Markov model with mixture of Gaussians (GMM-HMM) as emissions is shown in Fig.4B, where the basic hidden Markov model (Fig.4A) parameters π and $\mathbf{A}$ are augmented by a stochastic emission matrix $\mathbf{B} \in \mathbb{R}^{Q \times M}$

$$\mathbf{B}(i, j) = P(M_t = i | X_t = j) \tag{32}$$

which contains probabilities of Q hidden states emitting M Gaussians, and the MQ Gaussians' means μ and covariance Σ matrices. Therefore in a GMM-HMM, each hidden state has a unique mixture of M Gaussians from which it emits sequence values.

Regardless of the choice of $\mathbf{B}$, each hidden state has a unique emission distribution corresponding to a portion of the dynamical space spanned by the observation sequences if the observation sequences are time series. The transitions between the dynamical subspaces then is directly governed by the hidden layer Markov process, which can be as varied as the structure of the Markov chain. Fig.5 shows three examples of Markov chains used in hidden Markov models, where ergodic or non-absorbing models allow each hidden state to transition to any other state. Less flexible, yet more ordered models like the upper triangle (5), and more general left-to-right models are more suitable for time ordered sequences, as these models require higher numbered states to be preceeded by lower numbered states [157].

There are three functions or operations of a hidden Markov models and there exist three algorithms which are used to perform these functions[94]. The learning, training, or equivalently parameter estimation function is performed using the Baum-Welch algorithm [15] based on the Forward-Backward algorithm. Given any number of observation sequences, the hidden Markov model parameters can be learned using the Baum-Welch algorithm. The second function of a hidden Markov model is the evaluation or likelihood computation function which computes the likelihood of a given observation sequence, and this is performed using the Forward algorithm. Therefore, given a model the Forward algorithm can be used to find the likelihood of an observation sequence. Furthermore, classification can be done by finding the model which yields the highest likelihood for a test sequence. Finally, the hidden state assignment or decoding function is the task of assigning a sequence of hidden states to an observation sequence, and this is performed using the Viterbi algorithm.

Note that the structure of the hidden layer Markov chain and choice of emission parameters such as the number of Gaussians in GMM-HMM must be predetermined before training, and further model selection techniques are required to find the optimal parameter set. Pohle et al [154] outline best practices of model selection for hidden Markov models, and conclude that despite the direct applicability of selection criteria

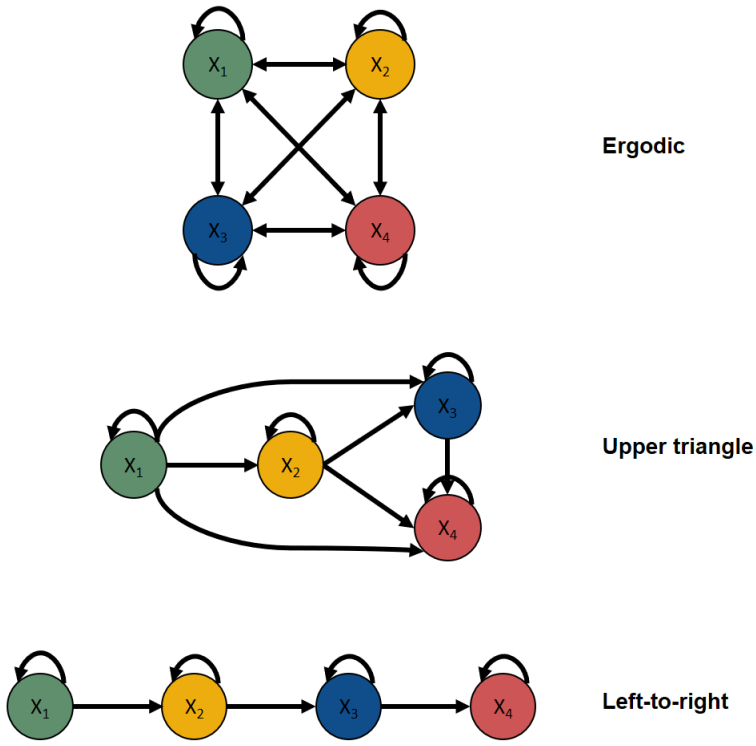


Figure 5: Examples of Markov chain structures for Hidden Markov model.

such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) care must be taken to avoid certain pitfalls. Therefore, to find the best parameters for hidden Markov models with mixture of Gaussians emissions, we use a combination of heuristics and BIC,

$$BIC = -2 \log \mathcal{L}(Y; \theta) + n_p \log n_o \tag{33}$$

$$\log \mathcal{L}(Y; \theta) : \text{Log-likelihood of observations given model} \tag{34}$$

where the parameter set $\theta = (\pi, \mathbf{A}, \mathbf{B}, \mu, \mathbf{\Sigma})$, n_p is the number of model parameters, and n_o is the number of observations used to train the model. The total number of parameters in the model depends on the choice of Markov chain structure,

$$n_p = Q + QM + QMd + QMd^2 + \begin{cases} 2Q - 1 & \text{left-to-right} \\ (1/2)Q(Q-1) & \text{upper-triangle} \\ 2Q - 1 & \text{bifurcated} \end{cases} \tag{35}$$

where d is the dimensionality of the multivariate observation time series.

1.3.4 Machine learning models

The feature based approach of extracting as many independent descriptions of a system feeds directly into machine learning models which use a set of features for i) prediction via regression or classification, and ii) detection via clustering or anomaly identification (Table 4). Therefore machine learning models can be trained to predict or detect by example if training samples with labeled outcomes are available, this is called supervised learning. However, in the absence of labeled target or outcome variables, unsupervised learning techniques can be used to cluster similar samples. There are other paradigms of machine learning such as reinforcement learning, but these do not have as direct applicability to biomedical problems as supervised and unsupervised techniques. For instance, supervised learning techniques can be used to predict probability of survival at the patient scale using demographic and human performance data, and classify pathological states at cellular and tissue scales using image data. In scenarios where a subset of the samples have labeled output data, e.g. if only a few patients' survival data is available in a longitudinal dataset, then semi-supervised techniques (Table 4) may be used to learn from partial knowledge of outcomes. Regardless of paradigm, machine learning models often employ a pipeline of steps including some or all of the following: i) preprocessing the data to replace missing data and to remove irrelevant features and samples, ii) a dimensionality reduction step to avoid curse of dimensionality ([132]), iii) calculating distance between features, particularly time series, iv) training, validation, and testing using one or a combination of algorithms. For instance, a set of time series features may be compared to one another or to a reference time series via dynamic time warping first and fed into a K-means or K-medoids based algorithm for clustering.

Machine learning paradigms			
	Supervised	**Semi-supervised**	**Unsupervised**
Paradigm	Ground truth is known	Mixed bag	Unknown ground truth
Data	(input, output)	(input, output) (input)	(input)
Tasks	Regression Classification	Regression Classification Clustering	Clustering Dimensionality reduction
Algorithms	Nearest neighbor Decision tree Naive Bayes Support vector machine (SVM) Random forest Artificial neural network	Generative models Self-training Transductive SVM Graph-based methods	K-means/medoids Mixture models Hierarchical clustering Principal component analysis Neural network/auto-encoders Graph model

Table 4: Three relevant paradigms of machine learning and examples of algorithms.

In biomedical applications of machine learning [16, 107, 216], model selection is performed by maximizing performance metrics such as area under the curve (AUC) of a receiver operating characteristic (ROC) curve

for classifiers, and minimizing root mean square error (RMSE) for regression models. Clustering models can be selected by comparing concordance between a given clustering and some benchmark expert clustering. Sensitivity and specificity, precision and recall, and F1 scores are also prevalent performance metrics in biomedical studies where it is most important to detect pathological cases while avoiding false positives. As such, developments in machine learning have lead to data pipelines which are highly successful at prediction and detection tasks in the biomedical field.

1.4 Chapter summaries

In Chapter 2 human performance motion capture data of from patients performing supervised exercises in a clinical setting are studied using a feature based pipeline of tools to classify patients by unexpected hospitalizations and low physical activity at home using statistical features extracted from the raw motion time series and subsequent kinematic feature time series. Kinematic feature extraction is explained, and a derivation of angular velocity of two-joint sections is shown. A short list of kinematic features which classify low and high risk hospitalization and physical activity patient groups are provided.

In Chapter 3 two model based pipelines are employed using the same human performance data from Chapter 2. The first model based pipeline uses kinematic features from the captured motion, however, this pipeline uses the dynamic time warping algorithm to compare time series in order to predict physician assigned fitness scores. The third pipeline also uses the dynamic time warping distances between kinematic time series features, but uses dimensionality reduction to learn a patient clustering and compares it with qualitative physician assigned clusterings of patients.

In Chapter 4 a clinical dataset containing preoperative and operative features is used to create classifiers of post-cystectomy survival and disease recurrence. Furthermore information theory based correlation techniques are used to understand dependencies among the feature set. Feature extraction is not performed here as the dataset is relatively high dimensional. The classifiers developed in this chapter make predictions for individual patients using the population level dataset.

In Chapter 5 a dataset of organoid morphogenesis images is studied using a feature based pipeline as well as a statistical model based pipeline. Contour time series are extracted from sequences of two-dimensional images of basal and growth factor treated organoids. Geometrical and signal processing based time series features are extracted from the contour time series and used to cluster organoids into groups of similar spatio-temporal behavior. Secondly, a point distribution statistical model of shape is learned from the organoid contours which is used to understand the behavior of the clusters detected using the feature based pipeline.

In Chapter 6 the same dataset as Chapter 5 is used to create a stochastic model of organoid area and

perimeter in order to discover various organoid subtypes and their dynamics. To do this, a hidden Markov model is used which is trained with the area and perimeter time series for basal and growth factor treated organoids. The goal of this strictly model based approach is to reveal the hidden genomic states responsible for the observed breadth of area and perimeter phenotypes.

Pre-thesis work is shown in Chapter 8 and Chapter 9, where a game theory model of cancer growth and tumor cell kinetics based on the prisoner's dilemma game is developed.

2 Quantified kinematics from in clinic movements differentiates patient risk for chemotherapy complications

2.1 Abstract

Cancer patients, particularly those treated with difficult chemotherapy regimens are at high risk for hospitalization during and between treatment cycles. Existing metrics based on physician observation of patients in the office are qualitative, subjective, and lack agreement among observers. Imaging tools have the potential to provide an objective and verifiable measurement of physician observations of patients in the office. To determine whether an in-office movement tracker, in cancer patients receiving highly emetogenic chemotherapy, can identify patients at risk for future unexpected hospital visits and low activity levels at home. Multicenter, single arm, observational clinical trial. A non-invasive motion capture system measured 1) patient movement from chair-to-table and 2) get-up-and-walk test. Activity levels were recorded using a wearable sensor over a two-month period with an additional 90-day follow-up. Kinematic signatures from motion capture data and subsequent statistics are tested for correlation with hospitalization and physical activity groups. A 15-hour (during study period) threshold was used to group patients into low and high activity groups. Secondly, patients were grouped if they had one or more unexpected healthcare encounters during the observation and follow-up period. Kinematic features from chair-to-table form the best classifiers for unexpected visits (AUC = 0.816) and physical activity (AUC = 0.735). Chair-to-table acceleration of non-pivoting knee (P = 0.001) and hip (P =0.002) are most correlated with unexpected hospitalizations. Angular velocity of the hip (P = 0.022) and non-pivoting leg (P = 0.024) are most correlated with physical activity. Physician observations of patient movement in the examination room can be quantified in a routine objective metric . Motion capture systems and wearable sensors can evaluate kinematic features that are correlated with unexpected hospital visits and physical activity.

2.2 Introduction

Each patient has specific and individual needs for optimal supportive care during cancer treatment. Predicting these needs and providing specific solutions has the opportunity to both improve outcomes and the experience during treatment. Poor patient outcomes, patient satisfaction, quality of life, and economic cost are associated with unexpected hospitalizations with patients actively receiving chemotherapy [164]. A recent survey of US oncology nurses found that 61% of nurses cared for patients who had to go to the emergency room or were hospitalized due to chemotherapy induced nausea and vomiting (CINV) [34]. These CINV hospitalization costs were estimated to be over $15,000 per occurrence [164, 34]. Readily available tools and

metrics such as ECOG performance status, Body Mass Index (BMI), Mini Mental State Exam (MMSE), and Charlson Comorbidity Index (CCI), are part of a comprehensive geriatric assessment, however few physicians perform the complete assessment, as they are time consuming. There is emerging data that a comprehensive geriatric assessment can predict complications and side effects from treatment [64, 57, 92, 58, 159, 79, 38]. Currently, the most routine assessment is the ECOG performance status. It is well known that in metastatic cancer such as lung origin, ECOG strongly predicts survival independent of treatment and usually guides if treatment should even be given if poor performance status [97, 76, 78, 114, 187, 148]. Clinical assessment of performance status and risk of toxicity from cancer therapy includes observation of patient movement as part of the physician examination within a clinic room environment. This has been routine practice for many years, and while it has been recognized for a long time, oncologists and patients substantially differ in their assessment of performance status with most oncologists being overly optimistic on the patient's performance status [6, 191]. Technological advances in low cost spatial cameras, such as Microsoft Kinect, have the potential to objectively define and categorize patients with varying levels of mobility at home or in the clinic [136]. Similarly, low cost activity trackers containing accelerometers, such as Microsoft Band, can capture daily movement in the clinic and at home, assessing dynamic changes related to exertion or to physical challenges such as the chemotherapy cycle. These consumer technologies have the capacity to bring objectivity to the assessment of mobility and performance status of patients on chemotherapy. The utility of activity trackers has been evaluated in areas outside of cancer medicine and demonstrated correlation with clinical outcomes in a wide variety of other disease settings [35, 28]. For example, in COPD, increasing additional steps correlates with reduced COPD hospitalizations [51] and formal exercise capacity evaluation such as the six-minute walk distance predicted COPD-related hospitalization [53]. After cardiac surgery, it was observed using an accelerometer that inpatient step count appears to predict repeat hospitalization [188]. In elderly hospitalizations it was found that mobility after hospital discharge could predict 30-day hospital readmissions [60]. To improve our understanding unexpected hospital visits in cancer patients receiving chemotherapy we conducted an observational study to evaluate the effect of physical activity as measured by a motion-capture system and wearable movement sensor and their relationship to unexpected healthcare encounters.

2.3 Methods

2.3.1 Trial Design

This study was a multicenter, single arm, observational trial conducted in the United States between July 2016 to July 2017. It was designed to compare kinematic signatures obtained from motion-capture systems

(e.g. Microsoft Kinect) and wearable motion sensors (e.g. Microsoft Band) and see if there was a correlation with unexpected hospital visits and physical activity at home and is described elsewhere [81, 120]. The study protocol was approved by the institutional review boards at all participating sites. Written informed consent was obtained from all participants.

2.3.2 Participants

Briefly, patients were eligible for the study if they were > 18 years of age, had a diagnosis of a solid tumor, and undergoing 2 planned cycles of highly emetogenic chemotherapy (see Appendix A), could ambulate without an assistive device, and had 2 separate kinematic evaluations successfully completed.

Patient demographics	
Number of patients	36
Age	
Median	48
Range	24-72
Gender	
Male	18
Female	18
Ethnicity	
Hispanic	22
Non-Hispanic	14
Goal of treatment	
Curative	30
Palliative	6

Table 5: Baseline characteristics of participants

2.3.3 Clinical exercises and motion capture

Patients underwent two clinically supervised tasks including chair-to-table (CTT) and get-up-and-walk (GUP) [129, 203]. CTT task begins with patients standing up from a chair while rotating the hip and left leg and pivoting on the right leg. Therefore, the CTT task design requires larger range of motion from the left lower extremities . The GUP task requires patients to stand up and walk to a marker 8 feet away, turn, and walk back to the starting position. We analyze the entire CTT task and the walking portion of GUP using the motion capture system.

The two tasks are performed by the cohort of cancer patients once pre-treatment (visit-1) and once post-treatment (visit-2). The Microsoft Kinect, a depth-sensing motion capture camera is used record the exercises, and three-dimensional positions of 25 anatomical sites (Figure 7) are extracted [136], from which six types of kinematic features are calculated: 1) velocity, 2) acceleration, 3) specific kinetic energy, 4) specific potential energy, 5) sagittal angle, 6) angular velocity (Appendix B). We exclude wrist, hand, ankle, and

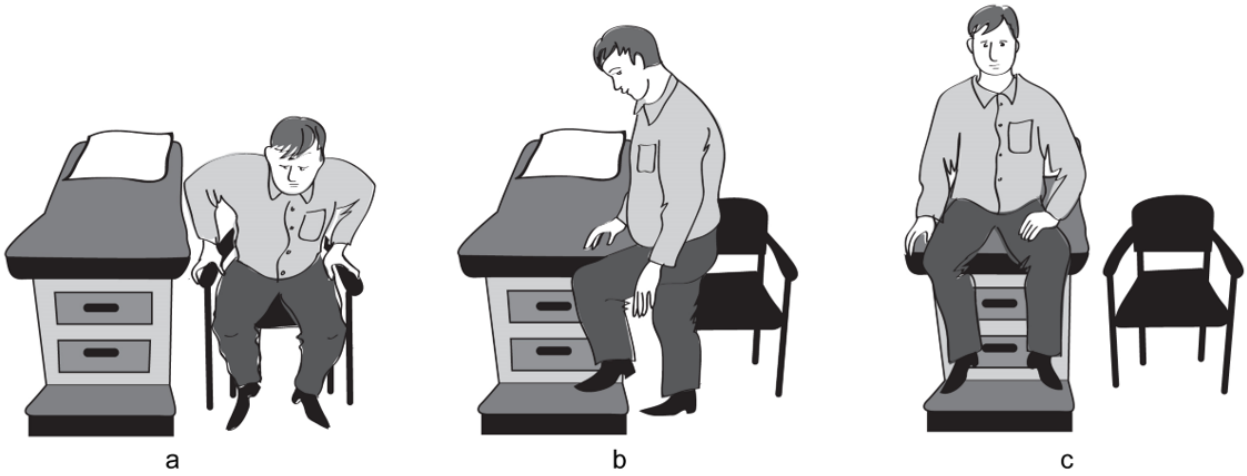


Figure 6: Illustration of chair-to-table exam maneuver demonstrating (A) sitting at rest (B) standing, rotating hip, pivoting on right leg, and using left (non-pivoting) leg to move onto exam table (C) sitting on exam table

foot joints (Figure 7) from statistical analysis as the motion capture signal for these joints is less reliable. The combination of selected joints and kinematic features capture the underlying biomechanics of patient movement and are therefore selected for inter-patient comparison.

Each patient has a pre- and post-treatment pair of samples of each feature, and four statistics (minimum, maximum, mean, median) from each visit's time series kinematic feature are averaged (mean) over the two samples. For the remainder of the paper, we refer to the mean-(minimum, maximum, mean, median) over the two visits simply as the minimum, maximum, mean, and median.

2.3.4 Physical Activity measure

Patient outcomes were grouped by activity level and unexpected hospital visits. During the study period that spanned for 60 days while receiving chemotherapy and a 90-day follow-up period, patients wore a wrist motion sensor to track their overall daily physical activity. We recorded the number of hours spent above low physical activity (LPA) for each patient over this period. Patients were considered high activity, rather than low activity, if they met greater than a 15-hour during the study period physical activity threshold. Patients with more than 15 hours of activity above LPA (HALPA = 0) and patients with 15 hours or less active time than LPA form the two HALPA groups.

Likewise, patients were grouped if they had one or more unexpected hospital visits compared to those that did not have any. Four types of unexpected hospital visits were tracked including: 1) Unplanned triage/infusion center visits, 2) urgent office visits, 3) urgent hospitalizations, and 4) ER visits. Patients with zero unexpected hospitalizations (UHV = 0) and patients with one or more unexpected hospital visits

are (UHV = 1) form the UHV groups.

2.3.5 Statistical Analysis

We differentiate patients by the average of visit-1 and visit-2 statistics for the set of kinematic features and correlate to two binarized clinical outcome UHV and HALPA. The Welch's t-test is used to test whether the mean value of the four averaged statistics is different for the UHV or HALPA groups, thereby revealing kinematic features which distinguish between UHV = 0 and UHV = 1 patients, and similarly HALPA = 0 and HALPA = 1 patients. The Welch's t-test also known as the unequal variance t-test allows the central tendency of two groups of unequal sizes and unequal variance to be tested for equivalence [167]. Secondly, we calculate the receiver operating characteristic (ROC) curve and use the corresponding area under the curve (AUC) as a metric of a feature's ability to classify patients into risk groups.

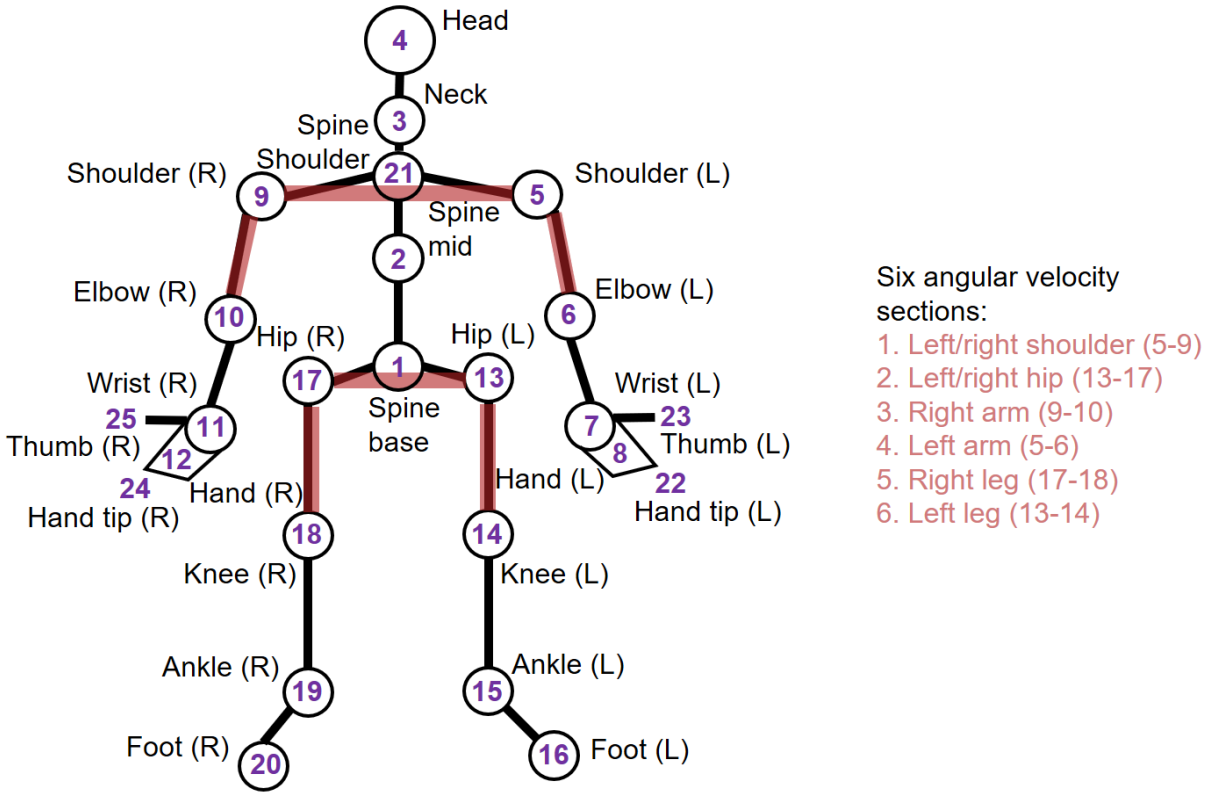


Figure 7: Schematic of anatomical sites detected in motion capture, along with six two-joint sections (red) whose angular velocities are extracted from raw kinematic data.

2.4 Results

2.4.1 Patient cohort

Of the 60 persons screened and agreed to participate in the study, 36 persons completed the study without drop out and had associated unexpected hospital visits and physical activity results. Overall the mean age of participates were 47.8 years old, and 50% were men. Breast, testicular, and head and neck cancer, comprised most of study participants . Chemotherapy was primarily of curative intent for most patients . Presumed reasons for higher than expected study drop out were likely due to a large proportion of persons being recruited from the Los Angeles County Hospital uninsured patient population combined with a large proportion being young males receiving chemotherapy for testicular cancer. These factors may explain why there was not a higher percentage of patients could complete the five-month study period. There are 16 UHV = 0 patients and 20 UHV = 1 patients for a total of N = 36 patients for whom hospitalization data is collected. Similarly, there are 17 HALPA = 0 patients and 18 HALPA = 1 patients for a total of N = 35 patients for whom physical activity data is collected.

2.4.2 Unexpected hospitalizations

We report the kinematic features that correlate most with unexpected hospital visits according to i) t-test and ii) ROC analysis in Table 6. CTT features dominate the list of UHV differentiating kinematic features and GUP features were less associated with the two outcomes. The full list of 55 features with significant t-test scores (p-value < 0.05) are listed in Appendix C.

Hip and left side joints are the top UHV features due to the pivot on the right side, and resulting large left side motion of CTT (see Figure 6). Figure 8A shows the ROC curves for the features with the highest AUC values for UHV where the maximum left leg angular velocity about the y-axis during CTT forms the best classifier of UHV (AUC = 0.816). The top three UHV differentiating features according to the t-test are plotted in Figure 8B, which shows the left knee, left hip, and the spine base mean accelerations during CTT are all generally higher for patients with no unexpected hospitalizations compared to patients with one or more unexpected hospitalizations.

2.4.3 Physical activity

Kinematic features that correlate most with physical activity according to i) t-test and ii) ROC analysis in Table 7. Unlike UHV, both CTT and GUP features appear in the list of HALPA differentiating kinematic features. The full list of 15 features with significant t-test scores (p-value < 0.05) are listed in Appendix D. Angular velocities, particularly those of the hip, differentiate HALPA groups the most. Nevertheless,

Welch's t-test				**ROC analysis**		
	Feature	t-test	p-value		Feature	AUC
1	Left knee: mean CTT acc	3.735	0.001	1	Left leg: max CTT av-y	0.816
2	Left hip: mean CTT acc	3.398	0.002	2	Left knee: mean CTT acc	0.806
3	Spine base: mean CTT acc	3.258	0.003	3	Left elbow: max CTT pe	0.781
4	Left knee: mean CTT vel	3.177	0.003	4	Left hip: max CTT acc	0.781
5	Left knee: mean CTT ke	3.14	0.004	5	Spine base: mean CTT acc	0.775
6	Left elbow: max CTT pe	2.988	0.005	6	Left hip: mean CTT acc	0.775
7	Right hip: mean CTT acc	2.928	0.006	7	Left knee: mean CTT ke	0.775
8	Left hip: max CTT acc	2.925	0.006	8	Right leg: min CTT av-x	0.759
9	Left hip: mean CTT ke	2.921	0.006	9	Hip: min CTT av-z	0.756
10	Right arm: mean GUP av-y	2.91	0.006	10	Left hip: mean CTT ke	0.753

Table 6: Top ten kinematic features from Welch's t-test (ranked by absolute value of two-sample t-test scores) and top ten kinematic features with highest AUC for differentiating between patients with no unexpected hospitalizations (UHV = 0) and patients with one or more unexpected hospitalizations (UHV = 1). (vel: velocity; acc: acceleration; pe: potential energy; ke: kinetic energy; sa: sagittal angle; av-x, av-y, av-z: angular velocity about x,y, or z axes).

kinematic features from the clinical exercises are less correlated with HALPA groups than UHV groups as both t-test scores and AUC values are generally lower in Table 7 compared to Table 6.

Figure 9A shows the ROC curves for the features with the highest AUC values for HALPA where the mean hip angular velocity about the vertical axis during CTT forms the best classifier of HALPA (AUC = 0.735). Mean hip and minimum left leg angular velocities during GUP are both larger (absolute value) for higher activity patients as seen in Figure 9B.

Welch's t-test				**ROC analysis**		
	Feature	t-test	p-value		Feature	AUC
1	Hip: mean GUP av-x	-2.414	0.022	1	Hip: mean CTT av-z	0.735
2	Left leg: min GUP av-x	-2.379	0.024	2	Hip: mean CTT av-y	0.729
3	Back: mean CTT sa	-2.331	0.026	3	Left arm: mean GUP av-y	0.725
4	Left arm: min GUP av-y	-2.328	0.032	4	Left knee: median GUP ke	0.722
5	Right leg: mean GUP av-z	2.224	0.033	5	Left leg: min GUP av-x	0.722
6	Left hip: mean CTT acc	2.221	0.033	6	Spine mid: mean CTT acc	0.719
7	Back: median CTT sa	-2.219	0.034	7	Right leg: median CTT av-y	0.719
8	Hip: mean CTT av-x	-2.193	0.035	8	Back: mean CTT sa	0.716
9	Left knee: median GUP ke	2.185	0.039	9	Shoulder: median CTT av-x	0.712
10	Right leg: median CTT av-y	-2.184	0.037	10	Hip: mean CTT av-x	0.706

Table 7: Top ten kinematic features from Welch's t-test (ranked by absolute value of two-sample t-test scores) and top ten kinematic features with highest AUC for differentiating between patients with more than 15 hours of activity above LPA (HALPA = 0) and patients with 15 hours or less activity above LPA (HALPA = 1). (vel: velocity; acc: acceleration; pe: potential energy; ke: kinetic energy; sa: sagittal angle; av-x, av-y, av-z: angular velocity about x,y, or z axes).

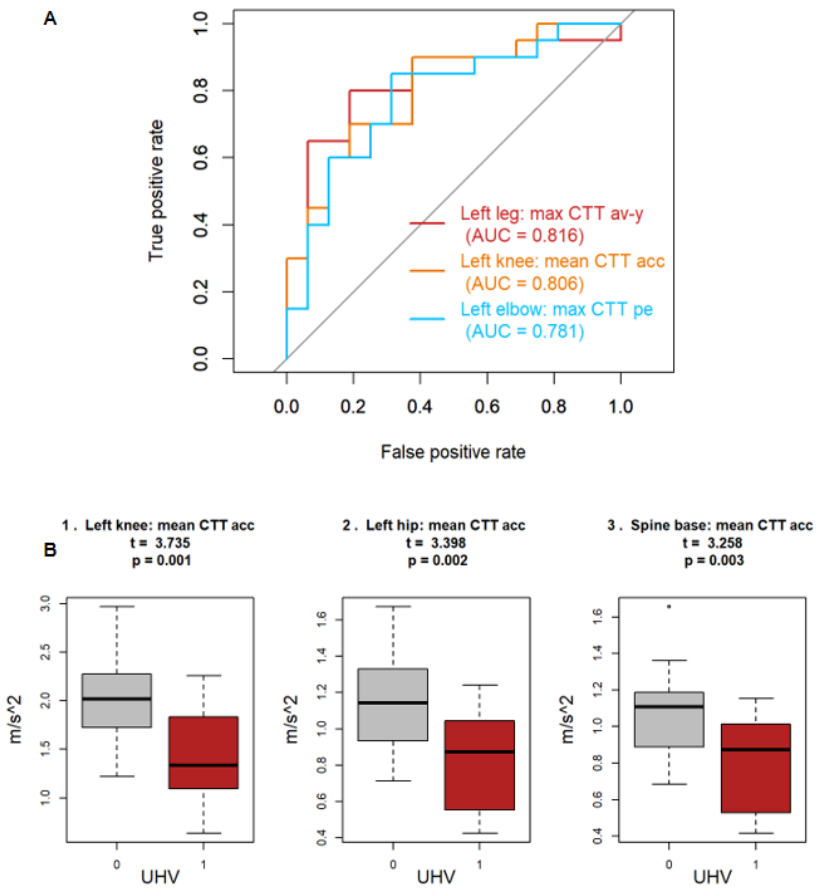


Figure 8: Kinematic features which differentiate patients with zero unexpected hospitalizations from patients with one or more hospitalizations. (A) ROC curves for features with the highest AUC. (B) Boxplots for features with the highest t-test scores (UHV = 0: gray, UHV = 1: red). (vel: velocity; acc: acceleration; pe: potential energy; ke: kinetic energy; sa: sagittal angle; av-x, av-y, av-z: angular velocity about x,y, or z axes).

2.5 Discussion

This study corroborates other studies by demonstrating that using a motion capture system and wearable motion sensor is feasible and can give kinematic data that can correlate and possibly be predictive of important clinical outcomes such as unexpected healthcare encounters [24, 77]. However, it seems to be clear that not all kinematic features are relevant while performing chair-to-table and get-up-and-go exams. Clinical assessment is of value, and what this study attempted to accomplish was to put a quantitative metric on the clinical assessment. As mentioned earlier, the kinematic features were based off of 25 anatomical sites that include head, arms, spine, hips, knees, and feet. The kinematic features of the get-up-and-go exam did not correlate with unexpected hospital visits however five kinematic features of the chair-to-table exam did. Interestingly, the anatomic sites that were statistically significant were left (non-pivoting) knee and hip, as

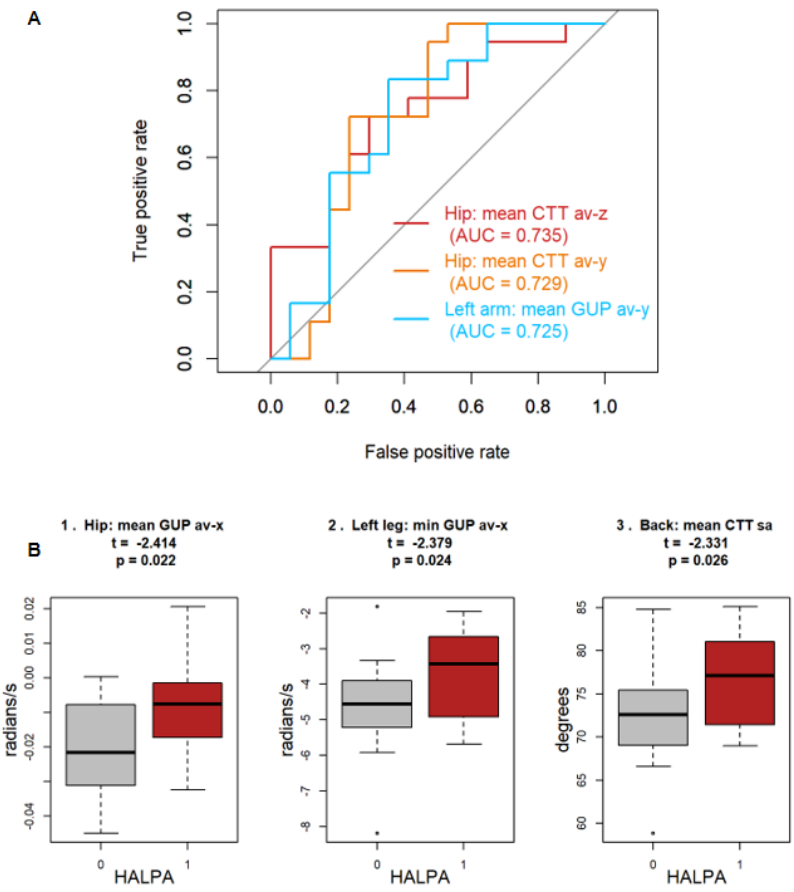


Figure 9: Top three kinematic features which differentiate patients with 15 hours or more of activity above LPA from patients with 15 hours or less of activity above LPA. A) ROC curves for features with the highest AUC. B) Boxplots for features with the highest t-test scores (HALPA = 0: gray, HALPA = 1: red). (vel: velocity; acc: acceleration; pe: potential energy; ke: kinetic energy; sa: sagittal angle; av-x, av-y, av-z: angular velocity about x,y, or z axes).

well as the spine base. The spine base velocity reflects the movement of a majority of the patient's mass that is not subject to high variability such as the distal hands or feet. Logically it would be safe to assume that patients can have a wide range of how they move their hands or feet during an exam maneuver which would not correlate with their overall physical fitness and hence not correlate with hospitalization risk .

The association between high physical activity level and kinematic features also seem to revolve around leg, knee, hip, and back movement. Similarly to above, these areas of the body intuitively carry the majority of a patient's mass and lower extremities generally are a more predictive measurement of a persons overall physical activity. This is supported by the calculated kinematic features (Table 6) noted above.

The mean hip and minimum left leg angular velocities about the x-axis during get-up-and-go are the two best differentiators of HALPA groups (Figure 9), and both these angular velocities are greater for patients

with higher physical activity compared to patients grouped in the low activity group. Mean sagittal angle during CTT is generally lower for patients with higher physical activity, which may be due to the increased ability of more active patients to crouch lower in the seated position before standing up and after reaching the medical table.

Ultimately the main goal of many recent studies including this one is to identify high-risk patients in a quantitative manner thus increasing reproducibility and reliability. Other modalities such as an Actigraph, a triaxial accelerometer, or using patient reported outcomes, have also recently been studied to better understand this issue [153, 184]. A range of different metrics have been evaluated that usually comprise of age, ECOG performance status, cognitive function (mini-mental state exam), nutritional status, comedications, comorbidities, autonomy, creatinine clearance, albumin, or other lab abnormalities [64, 57, 92, 58, 159, 79, 38]. Identifying high-risk patients is postulated as only one component of reducing costly preventable hospitalizations in cancer patients. Other proven strategies include enhancing access and care coordination, standardize clinical pathways for symptom management, availability of urgent cancer care, and early use of palliative care [184, 204, 26, 80]. Interestingly, although performance status likely still has value in predicting side effects and hospitalization risk that in the era of targeted therapy and immunotherapy performance status is less important in affecting response to novel therapies [33].

This study has limitations due to the small sample size and observational design compared to a traditional large randomized controlled trial. Confirmation of the finding would best be accomplished by integrating the tool into randomized clinical trials. The eventual goal would be to identify high-risk patients that have low activity levels or high risk for hospitalizations that appropriate interventions could be performed to reduce poor clinical outcomes. These interventions could include more frequent outpatient visits, IV hydration, or closer nursing/clinical navigator support.

In conclusion, patient performance and physical activity can be reliably quantified using camera based kinematic analysis. Modern sensor technology makes such as assessment rapid and low cost. The current platform requires a separate sensor and computer, and incorporation of these into a single hand-held device would be necessary to garner adoption of these algorithms for predicting adverse outcomes. Having such a tool available that quantifies what the physician sees during the in-clinic examination has the potential to harmonize findings among different physicians, specialists, researchers and families who all rely on a uniform assessment of patient fitness for receiving difficult cancer treatments.

2.6 Acknowledgement

The project described was supported in part by award number P30CA014089 from the National Cancer Institute, and by Noetic Contract FFP 012016 (USA 051) Collaborative Initiative on Human Performance Optimization. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

2.7 Appendix A

2.7.1 Calculating the Emetogenicity of Multiple Agent Chemotherapy/Biotherapy Regimens

Steps and Guidelines: 1. List each agent contained within the multiple agent regimen. 2. Identify the agent with the highest emetogenic level 3. Determine the contribution of the remaining agents using the following guidelines:

a. Level 1 agents do not contributor to emetogenicity in combination regimens.

Examples: level

$$1 + 1 = 0$$

$$2 + 1 = 2$$

$$3 + 1 = 3$$

$$4 + 1 = 4$$

b. Adding one or more level 2 agents increases the highest level by 1 in combination regimens.

Examples: level

$$2 + 2 = 3$$

$$3 + 2 = 4$$

$$2 + 2 + 2 = 3$$

$$3 + 2 + 2 = 4$$

c. Adding level 3 or 4 agents increase the highest level by 1 per each agent in combination regimens.

Examples: level

$$3 + 3 = 4$$

$$3 + 3 + 3 = 5$$

$$4 + 3 = 5$$

2.8 Appendix B

Details of kinematic feature extraction from the raw three-dimensional position motion capture data are described here. Anatomical site position vectors $\vec{r}_i = \langle \vec{x}_i, \vec{y}_i, \vec{z}_i \rangle$ are three-dimensional time series constructed from position at each time point, $r_i(t) = (x_i(t), y_i(t), z_i(t))$ for $i = 25$ anatomical sites. The position vectors are used to calculate velocity magnitude,

$$\vec{v}_i = (\dot{r_x}^T \dot{r_x} + \dot{r_y}^T \dot{r_y} + \dot{r_z}^T \dot{r_z})^{1/2} \tag{36}$$

and acceleration magnitude

$$\vec{a}_i = (\ddot{r_x}^T \ddot{r_x} + \ddot{r_y}^T \ddot{r_y} + \ddot{r_z}^T \ddot{r_z})^{1/2} \tag{37}$$

of each anatomical site using the mean-value theorem. Due to the lack of distribution of mass information, specific kinetic energy

$$\vec{T}_i = \vec{v}_i^T \vec{v}_i / 2 \tag{38}$$

and specific potential energy

$$\vec{U}_i = g \Delta \vec{z}_i = g(\vec{z}_i - \vec{z}_i(t = t_1)) \tag{39}$$

We define sagittal angle as the angle formed between $v_{1,m}$ the vector originating at the spine base and pointing in the direction of motion, and $v_{1,3}$ the vector connecting anatomical site 1 (spine base) and 3 (neck) at each time point.

2.8.1 Sagittal angle calculation

We define sagittal angle as the angle formed between $v_{1,m}$ the vector originating at the spine base and pointing in the direction of motion, and $v_{1,3}$ the vector connecting anatomical site 1 (spine base) and 3 (neck) at each time point. The sagittal angle is calculated using the inverse tangent of the ratio of the cross product and dot product of $v_{1,m}$) and $v_{1,3}$,

$$\theta_s = \tan^{-1}\left(\frac{\|v_{1,m} \times v_{1,3}\|}{v_{1,m} \cdot v_{1,3}}\right) \tag{40}$$

2.8.2 Angular velocity calculation

The angular velocity of the sections defined in Figure 7 are calculated using three-dimensional rigid body kinematic equations for relative motion. A section (Figure 7) is treated as a rigid bar and is defined by two anatomical points (e.g. left and right hips define the hip section) and we refer generically to these two ends as point A and point B. We calculate the velocities of these two points from the position vectors using the mean-value theorem as mentioned previously. Therefore, using these two velocities, the angular velocity of the section ω_{AB} can be isolated in the relative velocity vector equation

$$\vec{v}_B - \vec{v}_B = \vec{\omega}_{AB} \times \vec{r}_{AB} = (\Delta v_x, \Delta v_y, \Delta v_z) \tag{41}$$

where $\vec{r}_{AB}$ is the vector from point A to point B,

$$\vec{r}_{AB} = \vec{r}_B \vec{r}_A = (r_{AB,x}, r_{AB,y}, r_{AB,z}) \tag{42}$$

This vector equation has three components corresponding to the three directions and require an additional equation to solve for the three components of the angular velocity. Consequently, we use a kinematic restriction equation,

$$\vec{\omega}^{AB} \cdot \vec{r}_{AB} = 0 \tag{43}$$

because the angular motion of the section along the axis of the section does not affect its action. This allows for a solution to the three components of the angular velocity vector $\omega_{AB} = (\omega_x, \omega_y, \omega_z)$:

$$\omega_x = \frac{\Delta v_z r_{AB,y} - \Delta v_y r_{AB,z}}{r_{AB,x}^2 + r_{AB,y}^2 + r_{AB,z}^2} \tag{44}$$

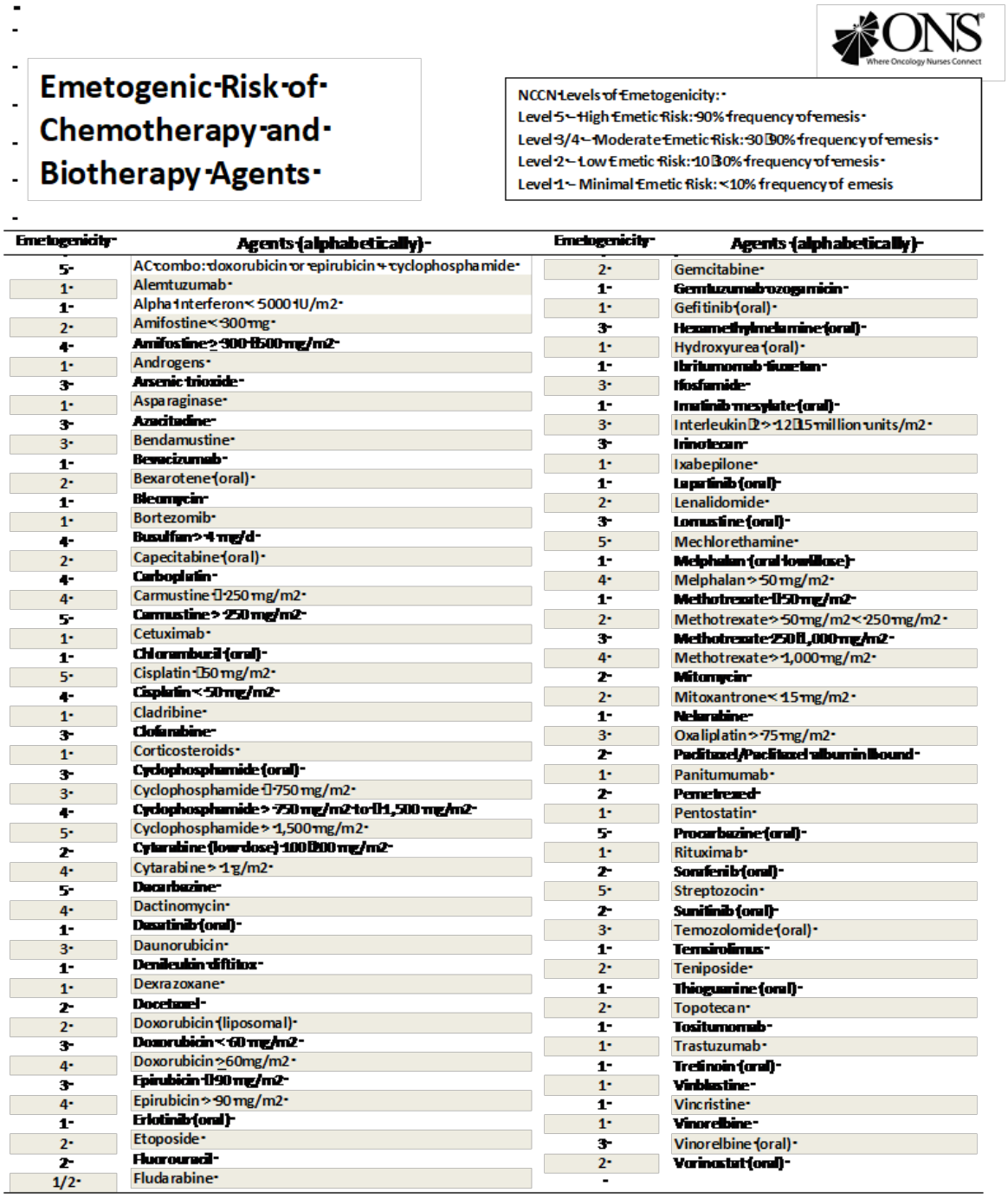
$$\omega_y = \frac{1}{r_{AB,x}} \left(r_{AB,y} \omega_x - \Delta v_z \right) \tag{45}$$

$$\omega_z = \frac{1}{r_{AB,y}} \left(r_{AB,z} \omega_y - \Delta v_x \right) \tag{46}$$

These equations are solved at each time point to get the time series of angular velocities for each section in Figure 7.

2.9 Appendix C

Two-sample t-tests are done to determine if mean values of kinematic features are different for patients with zero unexpected hospitalizations (UHV = 0) and patients with one or more hospitalizations (UHV = 1), and the distribution of the resulting t-test scores and significance values for the entire set of 526 features is shown in Appendix Figure 11. Full list of 55 significant (p-value < 0.05) t-test scores is shown in Appendix Table 8, and boxplots of these significantly differentiating kinematic features is shown in Appendix Figures 12, 13, and 14.



Emetogenic Risk of Chemotherapy and Biotherapy Agents

ONS Where Oncology Nurses Connect

NCCN Levels of Emetogenicity:
Level 5 – High Emetic Risk: >90% frequency of emesis
Level 3/4 – Moderate Emetic Risk: 30–90% frequency of emesis
Level 2 – Low Emetic Risk: 10–30% frequency of emesis
Level 1 – Minimal Emetic Risk: <10% frequency of emesis

Emetogenicity	Agents (alphabetically)
5	AC combo: doxorubicin or epirubicin + cyclophosphamide
1	Alemtuzumab
1	Alpha interferon < 5000 IU/m2
2	Amifostine < 300 mg
4	**Amifostine ≥ 300–500 mg/m2**
1	Androgens
3	**Arsenic trioxide**
1	Asparaginase
3	**Azacitadine**
3	Bendamustine
1	**Bevacizumab**
2	Bexarotene (oral)
1	**Bleomycin**
1	Bortezomib
4	**Busulfan > 4 mg/d**
2	Capecitabine (oral)
4	**Carboplatin**
4	Carmustine 1–250 mg/m2
5	**Carmustine > 250 mg/m2**
1	Cetuximab
1	**Chlorambucil (oral)**
5	Cisplatin ≥ 50 mg/m2
4	**Cisplatin < 50 mg/m2**
1	Cladribine
3	**Clofarabine**
1	Corticosteroids
3	**Cyclophosphamide (oral)**
3	Cyclophosphamide 1–750 mg/m2
4	**Cyclophosphamide > 750 mg/m2 to 1,500 mg/m2**
5	Cyclophosphamide > 1,500 mg/m2
2	**Cytarabine (low dose) 100–200 mg/m2**
4	Cytarabine > 1 g/m2
5	**Dacarbazine**
4	Dactinomycin
1	**Dasatinib (oral)**
3	Daunorubicin
1	**Denileukin diftitox**
1	Dexrazoxane
2	**Docetaxel**
2	Doxorubicin (liposomal)
3	**Doxorubicin < 60 mg/m2**
4	Doxorubicin ≥ 60 mg/m2
3	**Epirubicin 1–90 mg/m2**
4	Epirubicin > 90 mg/m2
1	**Erlotinib (oral)**
2	Etoposide
2	**Fluorouracil**
1/2	Fludarabine

Emetogenicity	Agents (alphabetically)
2	Gemcitabine
1	**Gemtuzumab ozogamicin**
1	Gefitinib (oral)
3	**Hexamethylmelamine (oral)**
1	Hydroxyurea (oral)
1	**Ibritumomab tiuxetan**
3	**Ifosfamide**
1	**Imatinib mesylate (oral)**
3	Interleukin 2 > 12–15 million units/m2
3	**Irinotecan**
1	Ixabepilone
1	**Lapatinib (oral)**
2	Lenalidomide
3	**Lomustine (oral)**
5	Mechlorethamine
1	**Melphalan (oral-lowdose)**
4	Melphalan > 50 mg/m2
1	**Methotrexate 1–50 mg/m2**
2	Methotrexate > 50 mg/m2 < 250 mg/m2
3	**Methotrexate 250–1,000 mg/m2**
4	Methotrexate > 1,000 mg/m2
2	**Mitomycin**
2	Mitoxantrone < 15 mg/m2
1	**Nelarabine**
3	Oxaliplatin > 75 mg/m2
2	**Paclitaxel/Paclitaxel albumin bound**
1	Panitumumab
2	**Pemetrexed**
1	Pentostatin
5	**Procarbazine (oral)**
1	Rituximab
2	**Sorafenib (oral)**
5	Streptozocin
2	**Sunitinib (oral)**
3	Temozolomide (oral)
1	**Temsirolimus**
2	Teniposide
1	**Thioguanine (oral)**
2	Topotecan
1	**Tositumomab**
1	Trastuzumab
1	**Tretinoin (oral)**
1	**Vinblastine**
1	Vincristine
1	**Vinorelbine**
3	Vinorelbine (oral)
2	**Vorinostat (oral)**

Figure 10: Chemotherapy Emetogenicity Table.

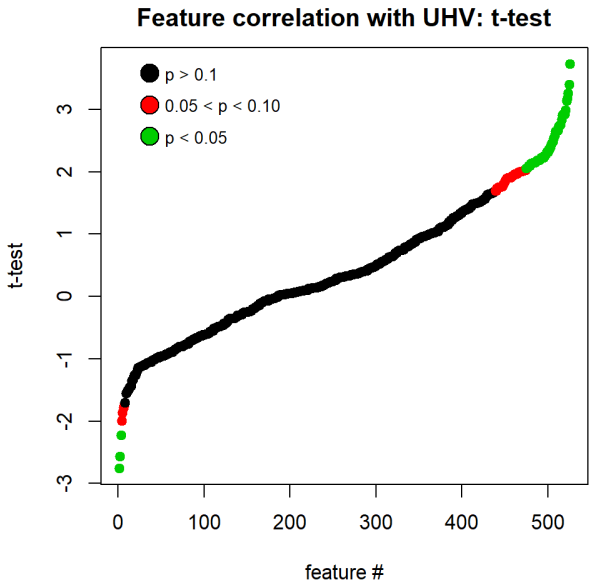


Figure 11: Distribution of t-test scores and significance values from two-sample t-tests for differences in mean values of kinematic features between patients with no unexpected hospitalizations (UHV = 0) and patients with one or more unexpected hospitalizations (UHV = 1).

	Feature	**t-test**	**p-value**		**Feature**	**t-test**	**p-value**
1	Left knee: mean CTT acc	3.735	0.001	29	Spine base: max CTT vel	2.322	0.027
2	Left hip: mean CTT acc	3.398	0.002	30	Neck: max CTT pe	2.315	0.027
3	Spine base: mean CTT acc	3.258	0.003	31	Shoulder: median GUP av-z	2.29	0.028
4	Left knee: mean CTT vel	3.177	0.003	32	Left elbow: mean CTT pe	2.26	0.031
5	Left knee: mean CTT ke	3.14	0.004	33	Left hip: mean CTT pe	2.257	0.031
6	Left elbow: max CTT pe	2.988	0.005	34	Spine base: median CTT acc	2.233	0.033
7	Right hip: mean CTT acc	2.928	0.006	35	Right elbow: median CTT acc	2.232	0.032
8	Left hip: max CTT acc	2.925	0.006	36	Spine base: mean CTT pe	2.229	0.033
9	Left hip: mean CTT ke	2.921	0.006	37	Left shoulder: mean CTT pe	2.228	0.033
10	Right arm: mean GUP av-y	2.91	0.006	38	Left leg: median GUP av-x	-2.227	0.033
11	Left knee: median CTT acc	2.844	0.008	39	Left knee: max CTT acc	2.195	0.037
12	Spine base: mean CTT ke	2.764	0.01	40	Left hip: median CTT acc	2.19	0.036
13	Left leg: min CTT av-x	-2.759	0.011	41	Right hip: mean CTT pe	2.186	0.036
14	Spine base: max CTT pe	2.745	0.01	42	Right elbow: mean CTT vel	2.186	0.036
15	Right hip: max CTT pe	2.725	0.01	43	Right leg: max CTT av-x	2.181	0.037
16	Left hip: mean CTT vel	2.671	0.012	44	Right knee: mean CTT vel	2.161	0.038
17	Spine base: max CTT acc	2.658	0.012	45	Right shoulder: max CTT pe	2.151	0.04
18	Left shoulder: max CTT pe	2.654	0.013	46	Spine mid: mean CTT acc	2.15	0.039
19	Left hip: max CTT pe	2.65	0.012	47	Left elbow: mean CTT vel	2.149	0.039
20	Spine base: mean CTT vel	2.591	0.014	48	Left shoulder: median CTT pe	2.143	0.04
21	Right leg: min CTT av-x	-2.566	0.017	49	Left elbow: median CTT acc	2.137	0.041
22	Right arm: max GUP av-y	2.542	0.02	50	Right hip: max CTT acc	2.13	0.04
23	Right hip: mean CTT ke	2.486	0.019	51	Left hip: max CTT vel	2.103	0.043
24	Spine mid: max CTT pe	2.456	0.02	52	Head: max CTT pe	2.095	0.044
25	Right hip: mean CTT vel	2.442	0.02	53	Left elbow: median CTT vel	2.078	0.046
26	Hip: median CTT av-m	2.396	0.023	54	Spine mid: mean CTT pe	2.071	0.046
27	Shoulder: median CTT av-m	2.363	0.024	55	Right hip: median CTT acc	2.062	0.047
28	Spine shoulder: max CTT pe	2.356	0.025				

Table 8: Full list of kinematic features which significantly (p-value < 0.05) differentiate between patients with no unexpected hospitalizations (UHV = 0) and patients with one or more unexpected hospitalizations (UHV = 1). Ranked by absolute value of two-sample t-test scores. (vel: velocity; acc: acceleration; pe: potential energy; ke: kinetic energy; sa: sagittal angle; av-x, av-y, av-z: angular velocity about x,y, or z axes).

2.10 Appendix D

Two-sample t-tests are done to determine if mean values of kinematic features are different for patients with 15 hours or more of activity above LPA (HALPA = 0) from patients with 15 hours or less of activity above LPA (HALPA = 1), and the distribution of the resulting t-test scores and significance values for the entire set of 526 features is shown in Appendix Figure 15. Full list of 28 significant (p-value <0.05) t-test scores is shown in Appendix Table 9, and boxplots of these significantly differentiating kinematic features is shown in Appendix Figures 16 and 17.

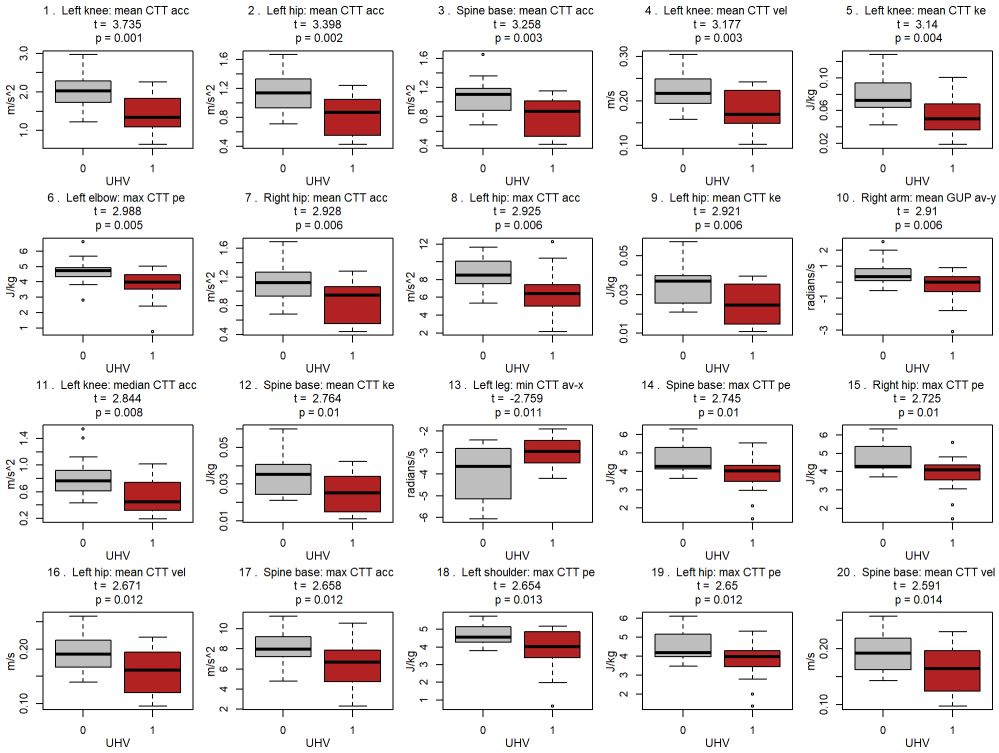


Figure 12: Box plots of kinematic features which significantly differentiate between patients with no unexpected hospitalizations (UHV = 0, gray) and patients with one or more unexpected hospitalizations (UHV = 1, red). Kinematic features 1-20.

	Feature	**t-test**	**p-value**		**Feature**	**t-test**	**p-value**
1	Hip: mean GUP av-x	-2.414	0.022	15	Spine base: mean CTT acc	2.039	0.05
2	Left leg: min GUP av-x	-2.379	0.024	16	Right hip: mean CTT acc	1.987	0.055
3	Back: mean CTT sa	-2.331	0.026	17	Hip: mean CTT av-y	1.96	0.065
4	Left arm: min GUP av-y	-2.328	0.032	18	Right leg: median GUP av-x	-1.96	0.06
5	Right leg: mean GUP av-z	2.224	0.033	19	Head: mean CTT acc	1.879	0.071
6	Left hip: mean CTT acc	2.221	0.033	20	Left arm: max GUP av-x	1.838	0.076
7	Back: median CTT sa	-2.219	0.034	21	Hip: max CTT av-y	1.837	0.084
8	Hip: mean CTT av-x	-2.193	0.035	22	Shoulder: mean CTT av-x	-1.805	0.083
9	Left knee: median GUP ke	2.185	0.039	23	Right arm: median CTT av-x	-1.775	0.086
10	Right leg: median CTT av-y	-2.184	0.037	24	Left leg: median GUP av-m	-1.775	0.086
11	Spine mid: mean CTT acc	2.181	0.037	25	Left knee: median GUP vel	1.763	0.091
12	Spine shoulder: mean CTT acc	2.136	0.042	26	Right leg: mean GUP av-x	-1.742	0.091
13	Neck: mean CTT acc	2.125	0.043	27	Spine mid: max CTT acc	1.727	0.094
14	Shoulder: median CTT av-x	-2.115	0.042	28	Left hip: mean CTT ke	1.702	0.098

Table 9: Full list of kinematic features which (feature 1-15: p-value < 0.05, feature 16-28: $0.05 <$ p-value < 0.10) differentiate between patients with no unexpected hospitalizations (UHV = 0) and patients with one or more unexpected hospitalizations (UHV = 1). Ranked by absolute value of two-sample t-test scores. (vel: velocity; acc: acceleration; pe: potential energy; ke: kinetic energy; sa: sagittal angle; av-x, av-y, av-z: angular velocity about x,y, or z axes).

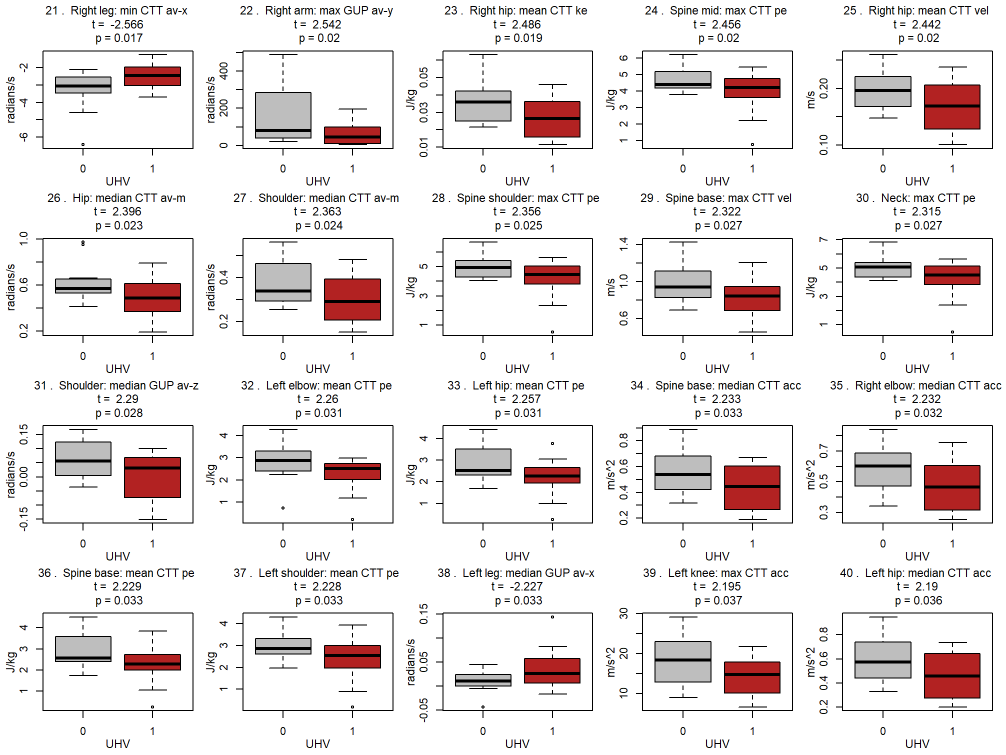


Figure 13: Box plots of kinematic features which significantly differentiate between patients with no unexpected hospitalizations (UHV = 0, gray) and patients with one or more unexpected hospitalizations (UHV = 1, red). Kinematic features 1-20.

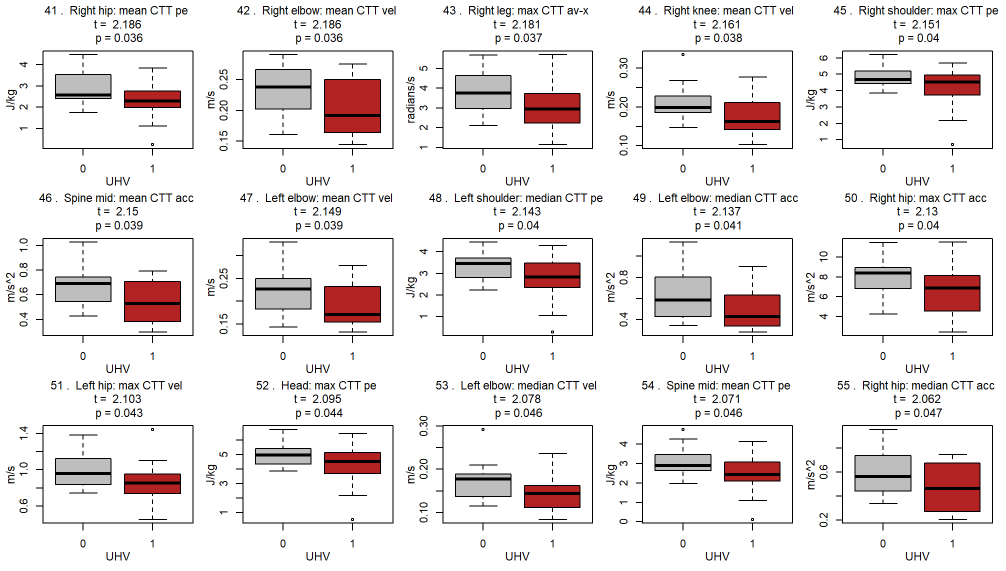


Figure 14: Box plots of kinematic features which significantly differentiate between patients with no unexpected hospitalizations (UHV = 0, gray) and patients with one or more unexpected hospitalizations (UHV = 1, red). Kinematic features 1-20.

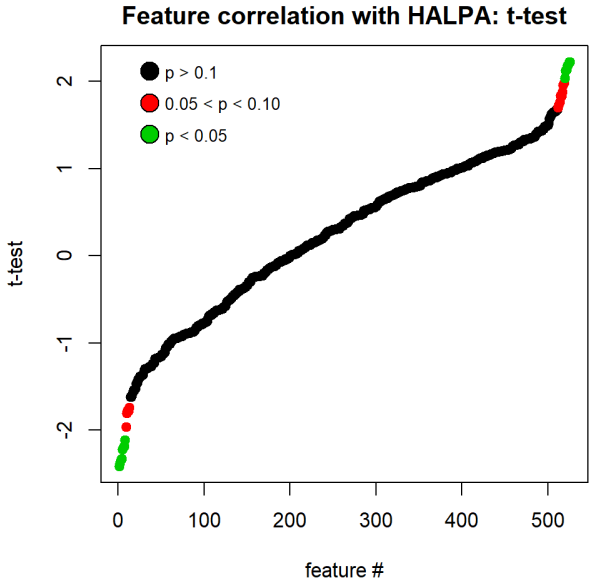


Figure 15: Distribution of t-test scores and significance values from two-sample t-tests for differences in mean values of kinematic features between patients with 15 hours or more of activity above LPA (HALPA = 0) from patients with 15 hours or less of activity above LPA (HALPA = 1).

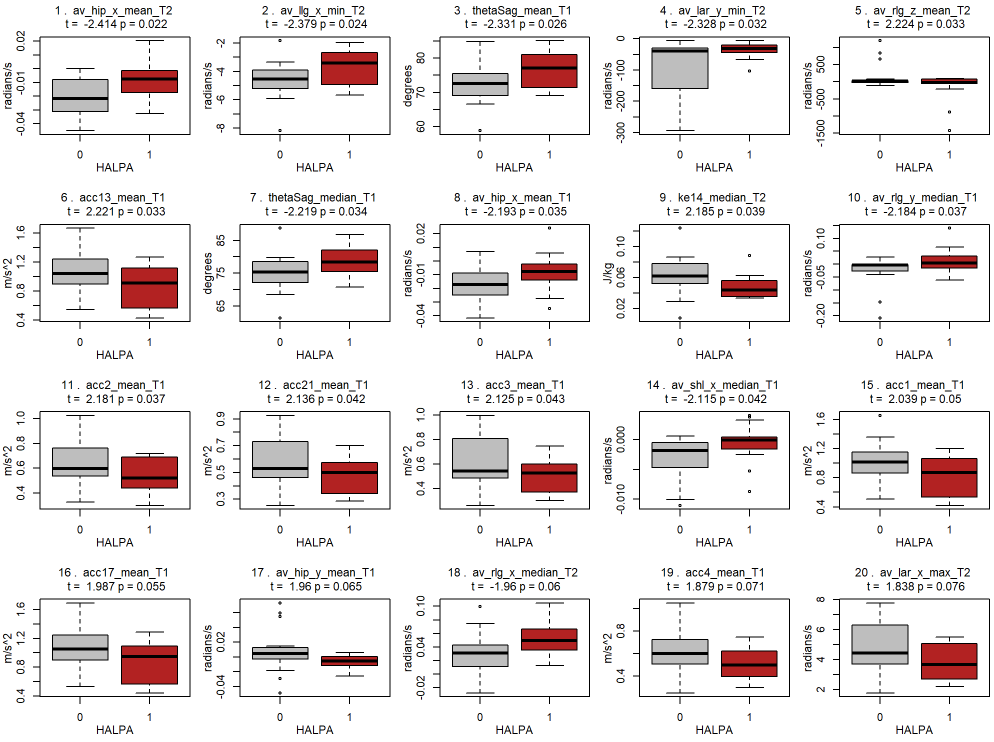


Figure 16: Box plots of kinematic features which significantly differentiate between patients with 15 hours or more of activity above LPA (HALPA = 0, gray) from patients with 15 hours or less of activity above LPA (HALPA = 1, red). Kinematic features 1-20.

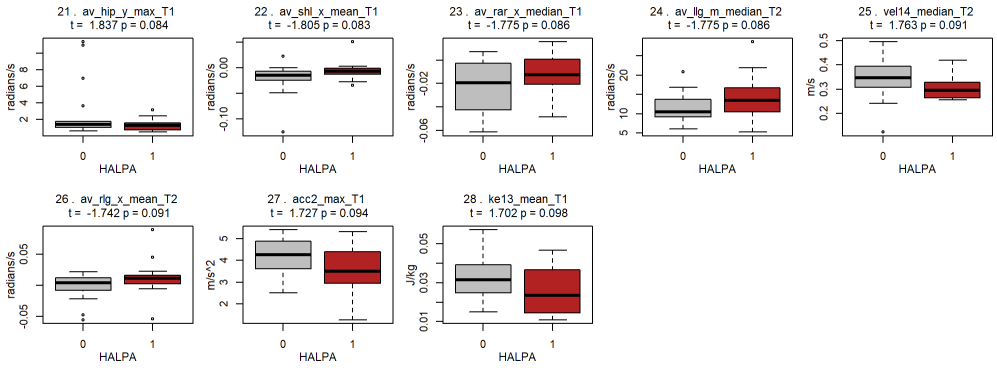


Figure 17: Box plots of kinematic features which significantly differentiate between patients with 15 hours or more of activity above LPA (HALPA = 0, gray) from patients with 15 hours or less of activity above LPA (HALPA = 1, red). Kinematic features 21-28

3 Low-dimensional characterization of human performance of cancer patients using motion data

3.1 abstract

Biomechanical characterization of human performance with respect to fatigue and fitness is relevant in many settings, however is usually limited to either fully qualitative assessments or invasive methods which require a significant experimental setup consisting of numerous sensors, force plates, and motion detectors. Qualitative assessments are difficult to standardize due to their intrinsic subjective nature, on the other hand, invasive methods provide reliable metrics but are not feasible for large scale applications. Presented here is a dynamical toolset for detecting performance groups using a non-invasive system based on the Microsoft Kinect motion capture sensor, and a case study of 37 cancer patients performing two clinically monitored tasks before and after therapy regimens. Dynamical features are extracted from the motion time series data and evaluated based on their ability to i) cluster patients into coherent fitness groups using unsupervised learning algorithms and to ii) predict Eastern Cooperative Oncology Group performance status via supervised learning. The unsupervised patient clustering is comparable to clustering based on physician assigned Eastern Cooperative Oncology Group status in that they both have similar concordance with change in weight before and after therapy as well as unexpected hospitalizations throughout the study. The extracted dynamical features can predict physician, coordinator, and patient Eastern Cooperative Oncology Group status with an accuracy of approximately 80%. The non-invasive Microsoft Kinect sensor and the proposed dynamical toolset comprised of data preprocessing, feature extraction, dimensionality reduction, and machine learning offers a low-cost and general method for performance segregation and can complement existing qualitative clinical assessments.

3.2 Introduction

In oncologic practice, clinical assessments of performance stratify patients into subgroups and inform decisions about the intensity and timing of therapy as well as cohort selection for clinical trials. The Karnofsky performance status (KPS) [96] and the ECOG/World Health Organization (WHO) performance status [147] are two equally prevalent measures of the impact of disease on a patient's physical ability to function. The Karnofsky score is an 11-tier measure ranging from 0 (dead) to 100 (healthy) whereas the ECOG score is a simplified 6-tier score summarizing physical ability, activity, and self-care: 0 (fully active), 1 (ambulatory), 2 (no work activities), 3 (partially confined to bed), 4 (totally confined to bed), 5 (deceased) [147] .

Although these metrics have been employed for many decades due the practicality, standardization of

patient stratification, and speed of assessment, prospective studies have revealed inter- and intra-observer variability [151], gender discrepancies [20], sources of subjectivity in physician assigned performance assessments [151], and a lack of standard conversion between the two different scales [25]. Nevertheless performance status provides clinical utility because it is able to differentiate patient survival [158, 98]. Consequently, the existing protocol of assigning a performance status based on an inherently subjective assessment must be refined to achieve a more objective classification of a patient's physical function.

In contrast to the qualitative and relatively practical nature of physician assessments in the clinic, laboratory based invasive methods have been developed to biomechanically quantify elements of human performance. Many of these efforts have conducted gait analysis using accelerometer, gyroscope and other types of wearable sensors and motion capture systems [190] to detect and differentiate conditions in patients with osteoarthritis [198], neuromuscular disorders [67], and cerebral palsy [46]. The shortcomings of more extensive assessments such as gait analysis include high cost, time required to perform tests, and general difficulty in interpreting results [176]. The need for new technologies has been emphasized, particularly in the oncology setting [99], to bridge the gap between subjective prognostication using KPS or ECOG performance status and objective, yet cumbersome assessments of performance.

To this end, we propose a non-invasive motion-capture based performance assessment system which can (i) characterize performance groups using solely kinematic data and (ii) be trained to predict ECOG scores by learning from various physicians in order to reduce bias and intra-observer variability. The Microsoft Kinect is used as the motion-capture device due to its low cost, and ability to extract kinematic information without the need of invasive sensors. We describe and test a data processing and analysis pipeline using a cohort of 40 cancer patients who perform two clinically supervised tasks before and after therapy at USC Norris Comprehensive Cancer Center, Los Angeles County+USC Medical Center, and MD Anderson Cancer Center.

3.3 Methods

A set of dynamical analysis and machine learning tools is developed to gather kinematic information from recordings of patients performing tasks (Fig.18) with the goal of validating the experiment design by performing unsupervised classification of performance categories (Fig.18, step 4a), as well as supervised learning of physician assigned ECOG performance status (Fig.18, step 4b). Although we illustrate the use of the toolset by exploring its application to an oncology cohort, the following methods are general and may be used to characterize patient performance in other settings.

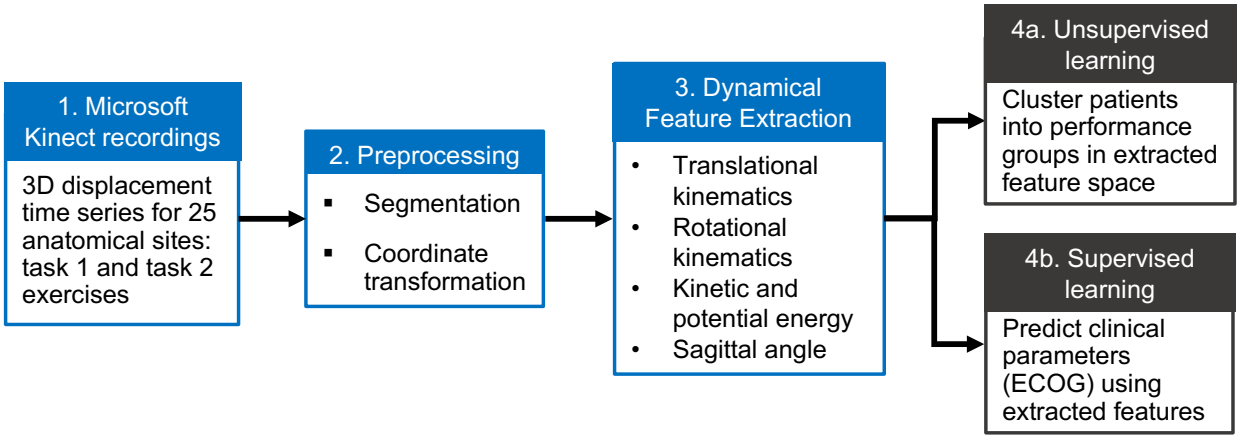


Figure 18: Schematic of dynamical and machine learning analysis pipeline. Raw skeletal displacement data (Step 1) from two clinically monitored tasks are preprocessed (Step 2) before feature extraction (Step 3) and two mutually exclusive machine learning analyses are performed. Unsupervised clustering (Step 4a) of patients in a low dimensional space reveals the degree to which performance groups can by stratified using solely motion data. Supervised classification (Step 4b) tests the ability of motion data to evaluate patients similar to physician ECOG performance status.

3.3.1 Experimental setup

The Kinect depth sensor employs an infrared laser projector to detect a representative skeleton composed of 25 anatomical points (Fig.19A) and recordings are post-processed using Microsoft Kinect SDK (v2.0) to extract 3-dimensional displacement time series data for the 25 points. The Microsoft Kinect sensor is used in the clinical setting to record patients performing two tasks: (i) task-1 requires patients, who start from a sitting a position, to stand up and sit down on an adjacent elevated medical table (Fig.19B), (ii) task-2 requires patients to walk 8 ft towards the Kinect sensor, turn, and return to the original position (Fig.19C). Both tasks are performed by each patient before and after a therapy cycle, providing two samples for each task for a total of four time series per patient. In both tasks the Kinect camera is secured to a tripod on a table, and oriented so as to capture the entire figure. Details about the data collection, skeletal data extraction, and experimental setup are described by Nguyen and Hasnain in [135].

3.3.2 Data preprocessing

Due to irregularities in the positioning of the Kinect camera across different experiments, time series for task-2 is distorted such that a level plane (e.g. clinic floor) appears sloped in the recordings. To resolve this, an automated element rotation about the x-axis is performed. The angle of distortion θ ranges between 5-20° in the time series studied. The second preprocessing step involves manually segmenting the series to trim irrelevant data in the beginning and end of each task while the patient is stationary.

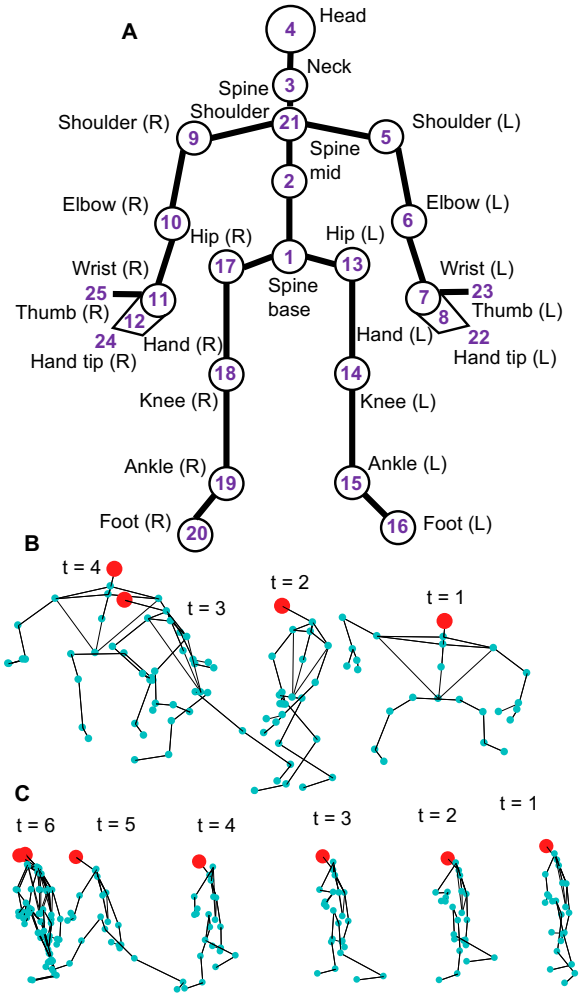


Figure 19: (A) Kinect recordings are post-processed using Microsoft Kinect SDK (v2.0) to extract displacement time series data for a set of 25 anatomical joints and sites. (B) Task-1 requires a patient to stand up from a chair and to sit at a medical table, a sample time series is shown. (C) Task-2 sample time series, patient starts from a standing position (t = 1) and walks to a mark 8 ft away (t = 6) and returns to original position (not shown).

3.3.3 Feature extraction

The position vector, $\vec{r}_i(t) = \langle x_i(t), y_i(t), z_i(t) \rangle$ for an anatomical joint i is used to calculate its velocity magnitude,

$$v_i(t) = \left\| \dot{\vec{r}}_i(t) \right\| \tag{47}$$

and acceleration magnitude,

$$a_i(t) = \left\| \ddot{\vec{r}}_i(t) \right\| \tag{48}$$

using the mean-value theorem. In the absence of distribution of mass information, specific kinetic energy,

$$ke_i(t) = \frac{1}{2}v_i^2(t) \tag{49}$$

and specific potential energy,

$$pe_i(t) = g\Delta z_i = g(z_i(t) - z_i(t=1)) \tag{50}$$

quantities are used to describe the energy signature of each anatomical joint. The sagittal angle, $\theta_s(t)$, is defined as the angle formed between the vector originating at the spine base and pointing in the direction of motion, and the vector connecting points 1 and 3 (Fig.19A) at each time point t.

Time series corresponding to the hand (7, 8, 11, 12, 22, 23, 24, 25 in Fig.19A) and feet (15, 16, 19, 20 in Fig.19A) joints are relatively noisy therefore these time series are precluded from analysis, yielding 13 joints of interest in Fig.19. In summary, the list of extracted features for each task performed by a patient during a single visit includes: $v_i, a_i, ke_i, pe_i, \theta_s$ for $i = 1, ..., 13$ anatomical joints resulting in 53 time series features per task, and $K = 106$ time series features per visit.

3.3.4 Time series similarity

For a given patient and task the before- and after-therapy time series of each feature are compared using a Euclidean metric dynamic time warping (DTW), which assigns a distance of zero for completely identical series and larger distances for more dissimilar series. Although DTW provides a distance which does not satisfy the triangle inequality and therefore is not a metric, it has been used extensively for time series clustering and classification [160]. In the present work, DTW is used to describe changes in the extracted features where it is necessary to detect similar series despite noise and distortions which are intrinsic to the Kinect sensor and subsequent skeleton extraction. Consequently, the pair of before and after-therapy time series are assigned a DTW distance, $d_{DTW}(p,k)$, for each patient p and feature k:

$$\mathbf{D}(p,k) = d_{DTW}(p,k) = \text{DTW}\left(X_{1,k}^p, X_{2,k}^p\right) \tag{51}$$

where $X_{1,k}^p$ and $X_{2,k}^p$ are the time series of patient p's feature k for visits 1 and 2 respectively. Calculating the DTW distance between before- and after-therapy visits for P patients and K features results in a matrix $\mathbf{D} \in \mathbb{R}^{P\times K}$ of DTW distances. This matrix captures the changes in the dynamical feature set before and after therapy. Feature distance vectors $\vec{d}_{k'} \in \mathbb{R}^P$ for $k' = 1, ..., K$, whose entries are $d_{DTW}(p,k')$, are columns of $\mathbf{D}$.

3.3.5 Dimensionality reduction

In practical applications, and the clinical case study presented here the number of patients who completed both visits, $P = 37 << K$, therefore further dimensionality reduction is required before implementing learning algorithms based on the matrix $\mathbf{D}$ in order to avoid overfitting and the curse of dimensionality [50]. Here, we use principal component analysis (PCA) to recast $\mathbf{D}$ into a lower dimension space while still maintaining most of the variance in the data. The scale of DTW distances are feature dependent, therefore column-wise standardization of $\mathbf{D}$ is performed prior to PCA. This process results in a reduced distance matrix $\mathbf{D}_r \in \mathbb{R}^{P \times N}$, comprised of N principal components, where $N \leq P$.

3.3.6 Unsupervised clustering

Performance groups are detected in reduced principal component space by employing the K-medoids algorithm, where number of clusters, k, corresponds to the number of performance groups detected. The K-medoids algorithm is chosen as the unsupervised algorithm for its insensitivity to outliers and fast implementation for the small dataset studied.

The overall quality of the resulting clusters is assessed by varying (i) the number of clusters in the K-medoids algorithm, and (ii) the number of principal components N in the low dimensional distance matrix $\mathbf{D}_r$ and measuring the silhouette s of the resulting clusterings as well as the concordance between a given learned clustering and three clinical clusterings based on changes in weight before and after therapy,

$$\Delta\text{weight} = \text{weight}_{\text{After therapy}} - \text{weight}_{\text{Before therapy}} \tag{52}$$

and change in ECOG performance status before and after therapy,

$$\Delta\text{ECOG} = \text{ECOG}_{\text{After therapy}} - \text{ECOG}_{\text{Before therapy}} \tag{53}$$

and the number of unexpected hospital visits (UHV) where patients are grouped into 0,1,$>$ 1 UHV over the course of the entire study. Changes in physician and coordinator assigned ECOG scores are used in this comparison.

The similarity between the ΔECOG, Δweight, UHV and Kinect based unsupervised clusterings is measured using the Rand index (RI):

$$RI(C, C') = \frac{a+b}{\binom{n}{2}} \tag{54}$$

where C and C' are two clusterings of n objects, a is the number of objects in the same clusters in C and C', and b is the number of objects in separate clusters in C and C'. $RI = 0$ when there is complete disagreement between two clusterings, and $RI = 1$ for identical clusterings.

3.3.7 Supervised classification

Instead of comparing a given patient's before- and after-therapy time series samples directly to each other as described in section section 3.3.4, a reference time series from a prototypical sample can be used to compare to the before- and after-therapy series separately. This approach allows for the construction of a distance matrix from a single patient visit, and enables subsequent machine learning models of the corresponding physician, coordinator, and patient assigned ECOG performance status. Three healthy subjects perform tasks 1 and 2 to generate the prototypical samples which serve as the reference points for patient performance, and DTW distances between a patient's time series data and the prototypical samples offers a standardized measure of performance. The DTW distances to each of the $L = 3$ prototypical healthy samples are averaged for patient p's extracted feature k from visit v, $X^p_{v,k}$,

$$d^v_{DTW}(p,k) = \frac{1}{L}\sum_{l=1}^{L} \mathrm{DTW}\left(X^p_{v,k}, X^l_k\right) \tag{55}$$

where X^l_k is the l'th prototypical sample's feature k, and visit $v = 1$ is the before-therapy sample and visit $v = 2$ is the after-therapy sample. Subsequently a standardized DTW distance matrix, $\mathbf{D}_s \in \mathbb{R}^{2P\times K}$, is formulated in which each patient contributes a total of two rows for the two visits. $\mathbf{D}_s$ represents the task-1 and task-2 average DTW distance between a patient's performance and the three prototypical samples. Along with a $\mathbb{R}^{2P}$ vector of a ground-truth target variable, $\mathbf{D}_s$ can be used to develop supervised learning models. Here, we use the physician, coordinator, and patient assigned ECOG performance status as the target variables in three separate models.

3.4 Results & Discussion

3.4.1 Current clinical parameters

37 patients completed the before and after therapy visits, and the corresponding physician and coordinator ECOG scores were recorded for a total of 74 visits, while only 31 patients reported ECOG scores (Table 1).

For the subset of 57 cases of patient reported ECOG scores, from either one or both visits, the mutual information (MI) association between physician ECOG and patient ECOG scores is $MI = 0.0653$, while the association between coordinator ECOG and patient ECOG is $MI = 0.1661$. Consequently, there is a

ECOG score distribution					
	0	1	2	3	Total
Physician	35	37	2		74
Coordinator	37	35	2		74
Patient	19	27	9	2	57

Table 10: ECOG scores assigned to patients by physicians, coordinators, and patients themselves for the before and after therapy visits.

larger agreement between the coordinator and patient scores in the current experiment, however, more data needs to be collected to verify this trend.

Fig.20A shows the relationship between physician ΔECOG (Eq.53) and change in weight over therapy, where change in weight follows a normal distribution (Fig.20B). However, due to the large spread of change in weight for ΔECOG = 0 group in Fig.20C, there is no clear relation between ΔECOG and Δweight, suggesting either the patients are unhindered even when undergoing large weight change (e.g. patients 5, 9, 11 Fig.20A) or that physicians consider other physical and expression cues more heavily while assigning ECOG scores.

Binning patients by the percent change in weight into groups of those who lose weight after therapy (Δweight $< 2\%$), maintain weight ($-2\% \leq$ Δweight $\leq 2\%$), and gain weight (Δweight $> 2\%$) results in a weight based clustering of the patients, which has a $RI = 0.509$ (n = 37 patients) with the physician ΔECOG clustering. The UHV clusters are comprised of 16, 9, and 11 patients in the 0, 1, and > 1 UHV groups, and has a $RI = 0.498$ (n = 36) with the physician ΔECOG. Although the time points of the physician ECOG scores correspond to the before and after therapy visits, the UHV events are summed over the entire course of the study. The level of concordance between physician ECOG and existing clinical parameters serves as a benchmark for the unsupervised clustering in step 4a (Fig.18).

3.4.2 Validation of preprocessing and choice of DTW

A hierarchical Euclidean distance based clustering of the the task-1 and task-2 feature distances $\vec{d}_{k'}$ (Fig.22A, feature names in Fig.22B) reveals that changes detected by DTW in most features are related mainly to other features of the same task, with the exception of a few features which correlate across tasks. These cross-task associated features include potential energies of the knee joints, and task 2 sagittal angle and left elbow potential energy to a lesser extent (purple, Fig.22A). The smaller clusters within the larger task 1 (blue, Fig.22A) and task 2 (red, Fig.22A) clusters validate the preprocessing and DTW calculations of across-therapy time series feature because anatomically related sites appear in coherent subclusters. For instance the potential energy of the left and right hips (joints 13 and 17, Fig.19A) and spine base (joint 1, Fig.19A) appear in the same subclusters for both tasks respectively. Furthermore, the velocities for the

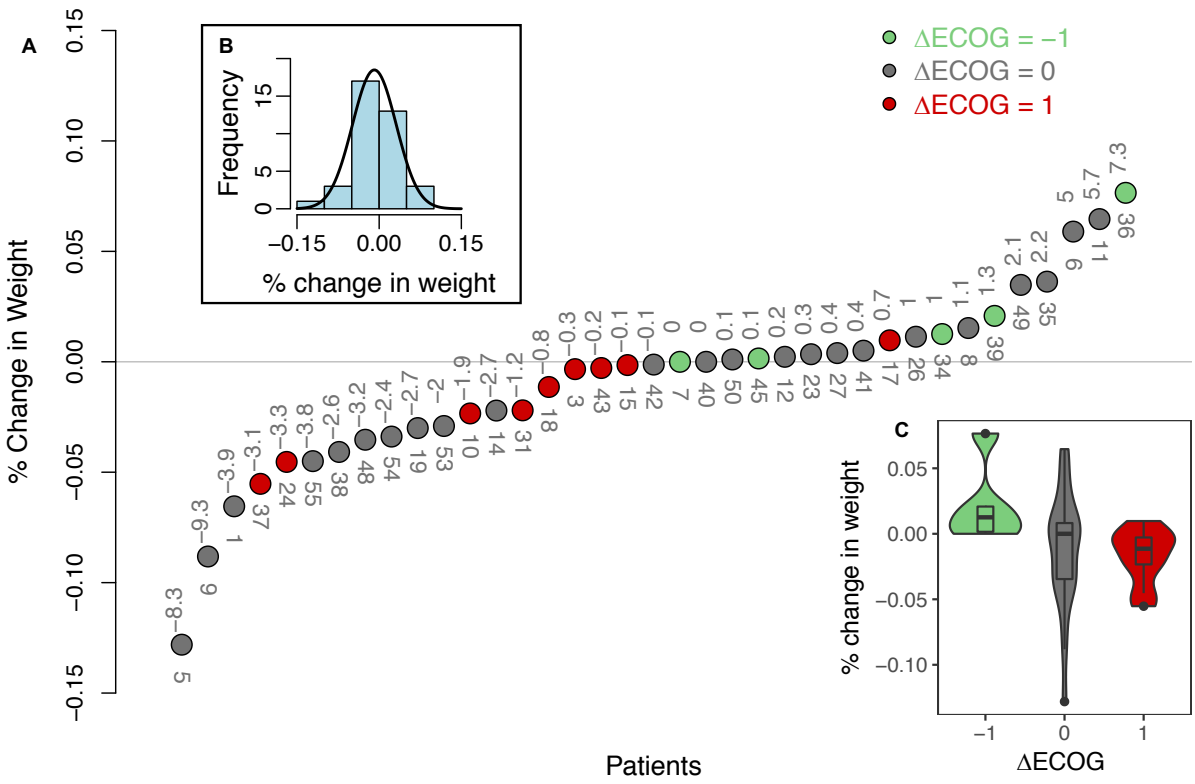


Figure 20: (A) Relation between change in physician assigned ECOG (eq.53) and percent change in weight before and after therapy. The absolute change in weight in kg is annotated above the circles, and patient ID is annotated below. (B) Histogram and normal distribution fit to percent change in weight. (C) Boxplot of change in weight by ΔECOG groups.

knee joints 14 and 18 are more closely related in the task-2 subcluster than the task-1 subcluster, which makes sense intuitively because the knees synchronously oscillate while walking in task-2, but perform unique functions in the task-1 twisting motion of turning towards and climbing the medical table. Therefore, the choice of using DTW, despite its dependence on the underlying scale of the time series being compared, is suitable for the subsequent unsupervised and supervised learning analyses.

3.4.3 Low dimension representation

The distance matrix $\mathbf{D}$ consisting of 106 feature distance vectors $\vec{d}_{k'}$ (56 per task) shown in Fig.22A constitutes a high dimensional representation of changes in the biomechanical performance of tasks 1 and 2 before and after therapy because the number patient samples is much less than the number of features: $P = 37 << K = 106$.

A low dimension representation is achieved by performing PCA on distance matrices $\mathbf{D}$ consisting of task-1, task-2, and both task features to generate the reduced matrices $\mathbf{D}_r$ for comparison (Fig.22C). In each

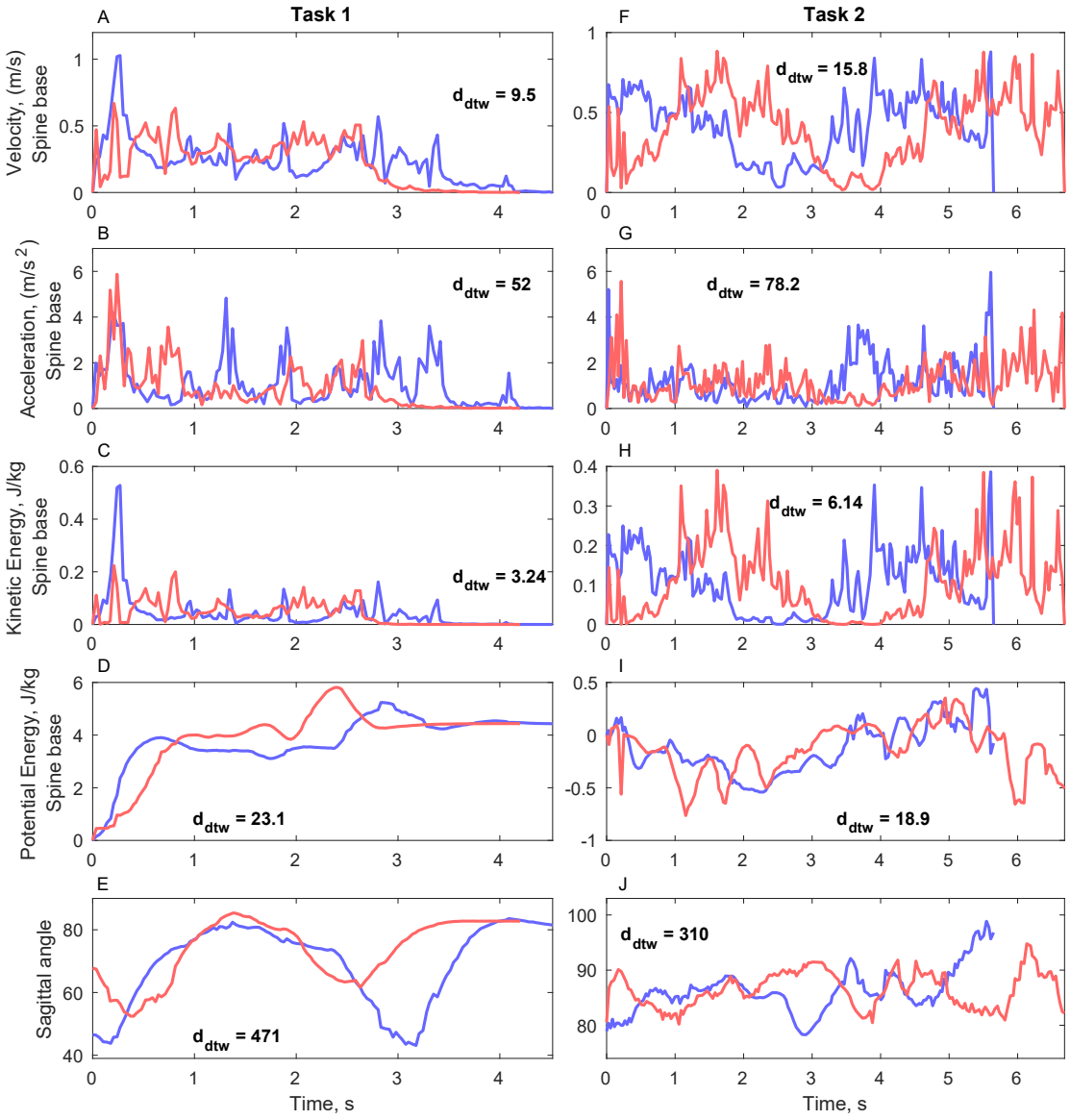
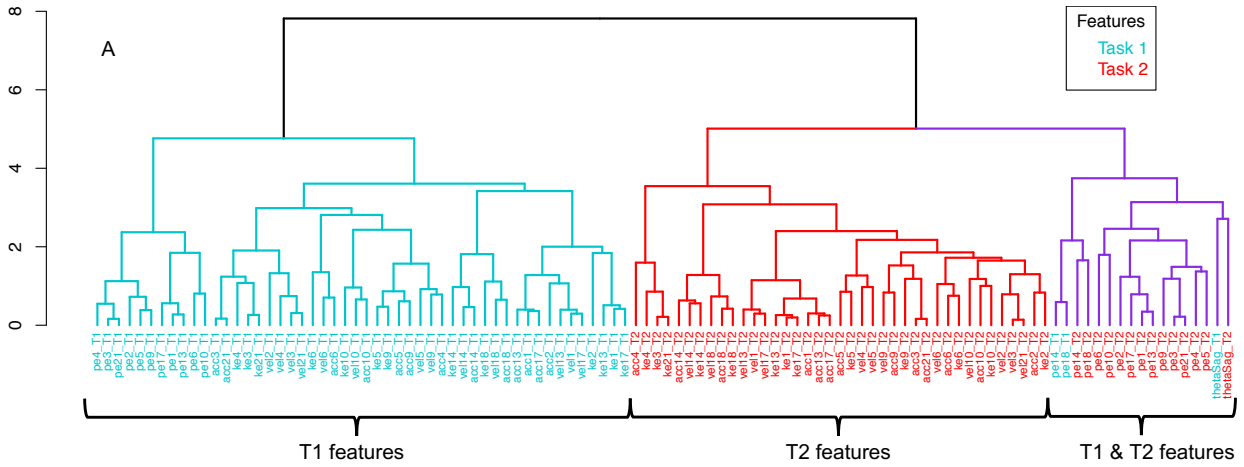


Figure 21: Example of extracted dynamical time series features (patient ID = 36). Before therapy (blue) and after therapy (red) feature time series are compared using DTW and the distances are annotated on the corresponding plots. (A-E) Task-1 features. (F-J) Task-2 features.

case, a small number of reduced dimensions can explain a significant portion of the variance in the high-dimensional space. Specifically, reduced matrices for task-1 and task-2 each require 8 principal components to describe nearly 90% of variance in the corresponding high dimensional distance matrices. 13 principal components are required to capture a similar amount of variance when distance features from both tasks are included in $\mathbf{D}$ due to the relative lack of cross-task association between distance features shown in Fig.22A. Nevertheless, the subsequent results are based on $\mathbf{D}$ which contains features from both tasks so as to prevent loss of information, and the additional task features do not adversely affect the learning algorithms.



B

Feature labels: (variable)_(point)_(task)		
Variable	Point	Task
acc: acceleration	25 anatomical points	T1: task 1, chair to table
vel: velocity		
ke: kinetic enregy		T2: task 2, get up and go
pe: potential energy		
thetaSag: sagittal angle		

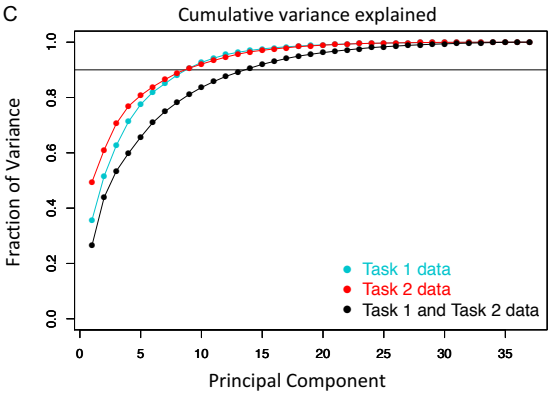


Figure 22: (A) Hierarchical clustering dendrogram of task-1 and task-2 feature distance vectors $\vec{d}_{k'}$. All but a few features (highlighted in purple) cluster primarily by task. (B) Feature label nomenclature for the dendrogram in A. (C) Fraction of variance explained by principal components of (i) distance matrix $\mathbf{D}$ comprised of task-1 and task-2 (K = 106) features (black), (ii) distance matrix $\mathbf{D}$ comprised of task-1 (K = 53) features (blue), (iii) distance matrix $\mathbf{D}$ comprised of task-2 (K = 53) features (red).

The feature clustering and dimensionality reduction analyses in Fig.22 illustrate the fact that both clinical tasks provide unique information and to use one test in the absence of the other would incur a loss of biomechanical information.

3.4.4 Detecting performance clusters

Changes in the before- and after-therapy performance of tasks 1 and 2 are captured in $\mathbf{D}_r$, and the number of performance clusters detected in $\mathbf{D}_r$ is a latent variable derived by selecting the number of clusters, k, in the K-medoids clustering algorithm which minimizes the distance to a representative cluster patient, or medoid, in the reduced low-dimensional space spanned by the N principal components. Therefore the choice of the number of performance clusters, from strictly the machine learning perspective, is dependent on the balance between N and the corresponding quality of the clustering which is measured by the silhouette $s \in [-1, 1]$. Higher values of s indicate higher intra-cluster cohesion and lower inter-cluster cohesion for a given patient.

This balance is shown in Fig.23, where greater number of dimensions in $\mathbf{D}_r$ generally correspond to lower average s. For instance, no matter the choice of k, a $N = 2$ distance matrix $\mathbf{D}_r$ cannot be clustered with a higher silhouette than a $N = 1$ distance matrix.

Here we seek the number of performance clusters to be much less than the number of patients in order to validate the unsupervised clustering with the three cluster Δweight, ΔECOG, and UHV clusterings, however, in general, larger values of k result in a higher resolution performance clustering of the patient group which in turn can be compared to higher resolution clinical categorizations.

From Fig.23A, we select the $N = k = 3$ clustering and visualize it on the first two principal components in Fig.23B, and compare all of the $k = 3$ clusterings to the physician ΔECOG, ΔWeight, and UHV clusterings in Fig.23C. The number of principal components used in $\mathbf{D}_r$ to segregate patients in to $k = 3$ clusters is varied and the corresponding concordance, as measured by RI is shown on the left axis and the average silhouette on the right axis of Fig.23C. Although the average silhouette of the K-medoid clusterings decreases with N, RI(K-medoid, ΔECOG), RI(K-medoid, ΔWeight), and RI(K-medoid, UHV) are maximized by N = 4, 17, and 19 principal components respectively.

As the number of principal components in the low-dimension space is varied, the concordance between the K-medoids clusters and the physician ΔECOG (black), ΔWeight (yellow), and the UHV (green) clusterings also changes (Fig.23C). The K-medoids clustering has a higher RI with the ΔWeight clustering for most choices of N compared to the benchmark $RI = 0.509$ between physician ΔECOG and ΔWeight (Fig.23C). As illustrated in Fig.23C, the ΔECOG, ΔWeight, UHV, and the unsupervised clusterings all have a similar concordance, therefore, the distance matrix $\mathbf{D}$ of before- and after-therapy performance of task-1 and task-2 offers an objective platform to stratify patients, and, as shown by the RI metric, may achieve concordance with existing clinical measurements.

The RI between the unsupervised K-medoid clusterings and the clinical clusterings including the physician and coordinator assigned ΔECOG clusterings is shown in Table 2. The K-medoid clustering has the highest RI with the UHV clusters, and the second highest association with the ΔWeight. The coordinator ΔECOG clusters are more associated with the ΔWeight and UHV clusters than the physician ΔECOG clusters. Although these trends are particularly interesting,larger datasets are required to validate these RI values and to fully detect statistically significant disparities. Nevertheless, Table 2 shows that the K-medoid unsupervised clusterings based on kinematic changes in task-1 and task-2 across therapy offers a unique but useful patient clustering.

The added utility of the motion analysis based unsupervised clustering method shown here is the ability to achieve higher resolution clusterings of patients compared to physician or coordinator ΔECOG by increasing k in the K-medoids algorithm. Fig.23D shows a comparison between $N = 3$ K-medoids and ΔWeight

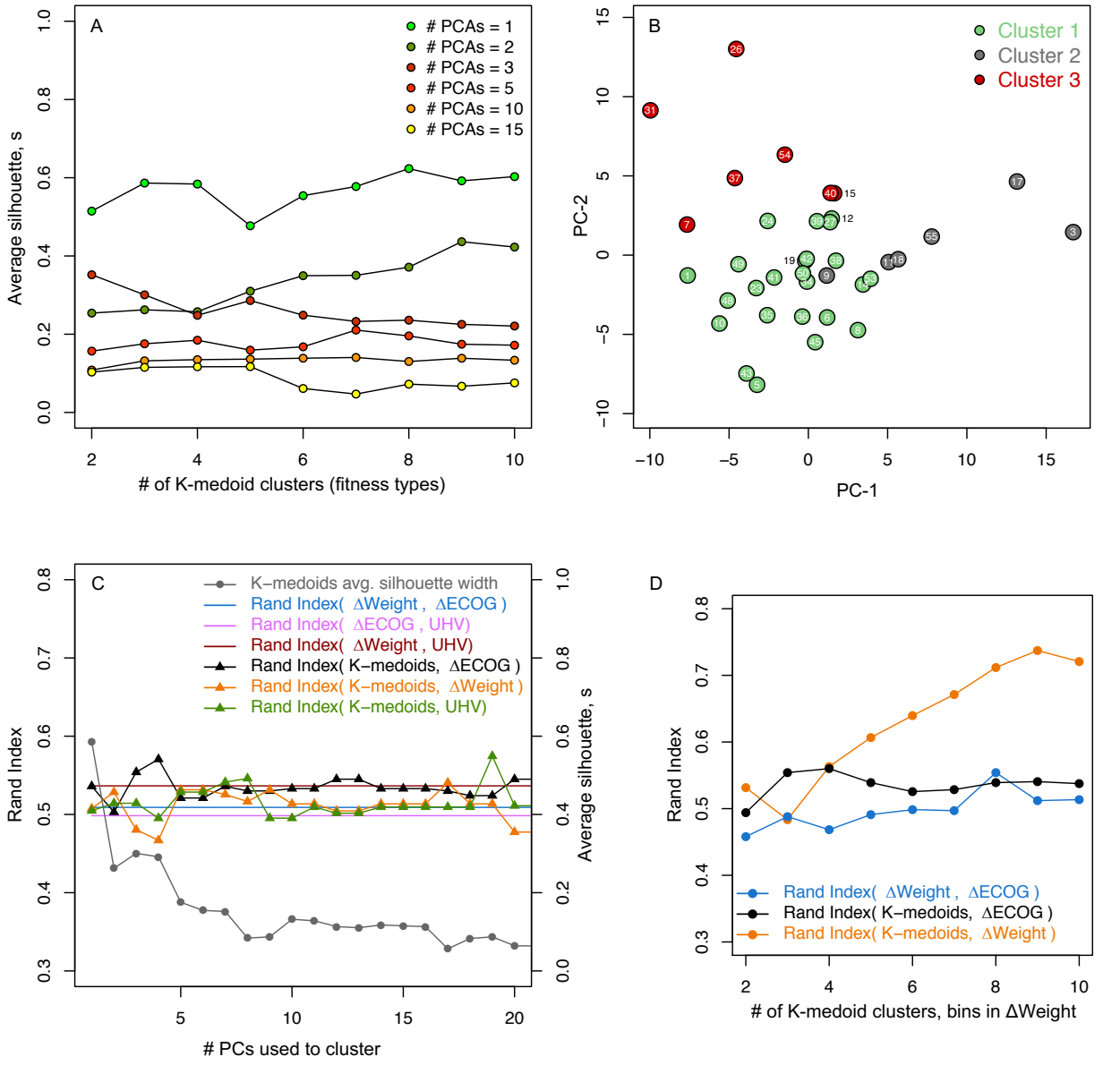


Figure 23: (A) Quality of K-medoids patient clusterings measured by average silhouette s compared to the number of clusters k for different numbers of principal components N in the reduced distance matrix $\mathbf{D}_r$. (B) $N = 3$ and $k = 3$ K-medoids unsupervised clustering of patients shown in the plane formed by the first two principal components of $\mathbf{D}_r$. (C) The RI concordance between the $k = 3$ K-medoids clusterings and physician ΔECOG (black), ΔWeight (yellow), and UHV (green) compared to benchmark RI associations among the ΔECOG, ΔWeight, and UHV based clusterings (solid lines). Quality of the K-medoids clusterings (gray) is shown on the right axis. (D) Concordance between clusterings where the number of clusters in K-medoids and bins in ΔWeight are increased.

clusterings where k and the number of bins in the ΔWeight clustering are increased which leads to an increasing RI between the two clusterings and a maximum $RI = 0.737$ is reached at $k = 9$ clusters. This further demonstrates the potential clinical utility of the pipeline of analytical tools developed.

Rand Index between clusterings				
ΔWeight	0.541			
UHV	0.575	0.537		
ΔECOG_P	0.571	0.509	0.498	
ΔECOG_C	0.550	0.571	0.530	0.497
	K-medoids	*ΔWeight*	*UHV*	*ΔECOG_P*

Table 11: Association between patient clusterings based on: 1) unsupervised K-medoids clustering (k = 3 clusters) of $\mathbf{D}_r$ 2) ΔWeight: change in weight before and after therapy 3) UHV: three clusters based on 0, 1, and >1 unexpected hospital visits over the course of the study 4) ΔECOG_P: change in physician ECOG scores before and after therapy 5) ΔECOG_C: change in coordinator ECOG scores before and after therapy

3.4.5 Learning physician ECOG performance status

A natural application of the Kinect motion capture system is to use the extracted kinematic signature of a patient's task-1 and task-2 performance to learn the associated physician, coordinator, and patient assigned ECOG performance status, particularly if patients are examined by different physicians in order to reduce bias of the resulting model.

To learn the ECOG scores in the cancer patient cohort, we use the standardized distance matrix $\mathbf{D}_s$ (section 3.3.7) and perform dimensionality reduction via PCA with scaling and centering. As in the unsupervised model, dimensionality reduction is required due to the relatively small number of patients compared to the number of features.

In the 74 physician and coordinator assessments of the $P = 37$ patients there are 2 ECOG = 2 cases. Since the majority of cases were ECOG = 0 or 1 scores, the two ECOG = 2 samples are excluded, and a binary classifier is trained to predict a 0 or 1 ECOG status using the two visits from $P = 37$ patients (excluding two ECOG = 2 cases from two separate patients), which results in a distance matrix for the physician and coordinator ECOG classifiers. For the patient classifier a $\mathbf{D}_s \in \mathbb{R}^{57\times 106}$ matrix is used and the ECOG = 1,2, and 3 categories are combined due to limited data.

The association of the original features of $\mathbf{D}_s$ and its principal components with the ECOG status is measured by the point biserial correlation coefficient r_{pb}, where positive r_{pb} values indicate larger values of the feature are associated with ECOG = 1, and vice versa (Fig.24).

For the physician ECOG, age has a $r_{pb} = 0.186$, as older patients were more likely to receive higher

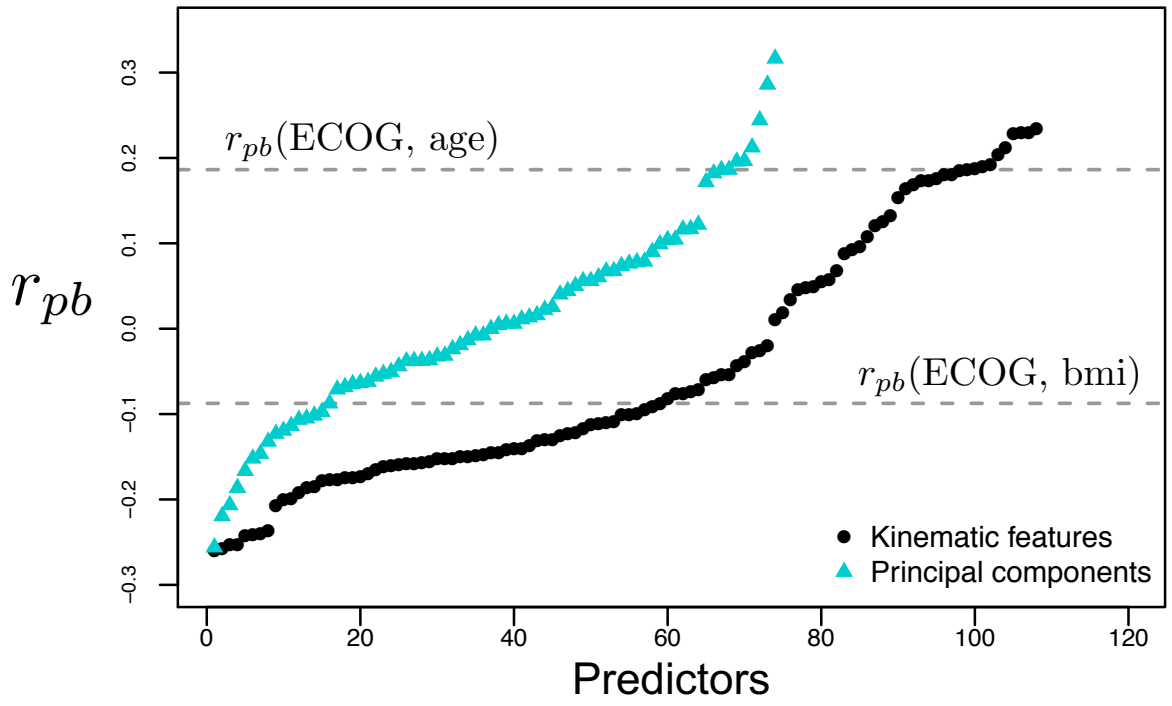


Figure 24: Point biserial correlation between physician assigned ECOG and kinematic features of $\mathbf{D}_s$ (black) and reduced space principal components of $\mathbf{D}_s$ (blue). Larger r_{pb} values indicate association with ECOG = 1, and vice versa. The correlations of age and BMI with ECOG (gray dashed lines) serve as a comparison for the r_{pb} of the kinematic features and principal components.

ECOG scores and in comparison, 10 principal components in the reduced $\mathbf{D}_s$ have a higher absolute r_{pb}. Principal components with the largest $\|r_{pb}\|$ are used to create and cross-validate classifiers by leaving one patient's two samples out as the test set. 10 principal components are used in a mixture-of-experts model comprised of a SVM, logistic regression, and a KNN model to predict physician ECOG with an average cross-validated test set accuracy of 84.7% accuracy. The same accuracy was achieved using the same number of principal components to predict coordinator ECOG with a SVM model. Top 5 principal components were used to train a logistic regression model to predict patient ECOG $= 0$ or > 0 which performed at an accuracy of 80.7%.

3.5 Conclusions

A non-invasive motion capture system is proposed to measure the kinematic signature of clinically supervised patient assessments of performance. A toolset to pre-process and extract dynamical features from skeletal displacement data is combined with complimentary unsupervised and supervised learning schemes. The unsupervised clusters reveal a new and valuable grouping of patients in a cancer cohort undergoing therapy. Additionally, the supervised learning model is able to predict physician, coordinator, and patient assigned

ECOG scores using the kinematic signature with a high level of accuracy. In comparison to the low-resolution ECOG scale, the present toolset provides a pipeline to develop a high resolution performance grading. In general, the dynamical characterization toolset may be used for prognostication in various applications where biomechanical signatures are reasonably correlated with existing clinical measures. The present work is a proof of concept of a low-cost non-invasive method for objectively assessing human performance in the clinic.

4 Machine learning models for long-term outcome prediction in bladder cancer

4.1 Abstract

Currently in patients with bladder cancer, various clinical evaluations (imaging, operative findings at transurethral resection and radical cystectomy, pathology) are collectively used to determine disease status and prognosis, and recommend neoadjuvant, definitive and adjuvant treatments.We analyze the predictive power of these measurements in forecasting two key long-term outcomes following radical cystectomy, i.e., cancer recurrence and survival. Information theory and machine learning algorithms are employed to create predictive models using a large prospective, continuously collected, temporally resolved, primary bladder cancer dataset comprised of 3503 patients (1971-2016). Patient recurrence and survival one, three, and five years after cystectomy can be predicted with greater than 70% sensitivity and specificity. Such predictions may inform patient monitoring schedules and post-cystectomy treatments. The machine learning models provide a benchmark for predicting oncologic outcomes in patients undergoing radical cystectomy and highlight opportunities for improving care using optimal preoperative and operative data collection.

4.2 Introduction

Bladder cancer (BCa) is the 6th most common cancer in the U.S, with an estimated 79,030 new cases and 16,870 deaths in 2017 [175] and has a 5-year relative survival rate of 79% [30]. BCa staging is based on the TNM system (tumor, nodes, metastasis). In BCa, the "T" stage is dictated by how deep the tumor invades into the various layers of the bladder wall. Ta represents a noninvasive papillary tumor, while T1, T2, T3 and T4 stages represent more aggressive cancers invading the sub-epithelial tissue, muscle, perivesical fat and adjacent organs, respectively. Radical surgery is the primary treatment method for invasive cancer and may be augmented with other forms of therapy such as chemotherapy to treat more advanced and aggressive cancers [47]. Radical cystectomy, the recommended method for treating invasive BCa [183], is surgical removal of the bladder, regional lymph nodes and adjacent organs (prostate, uterus, etc.) which may contain cancer. Technical precision of this surgical operation can dictate long-term oncologic outcomes, for instance, post-cystectomy survival is higher when negative surgical margins are obtained and more than ten pelvic lymph nodes removed during radical cystectomy [87]. Conversely, cancer recurrence rates are higher with positive margins and removal of less than ten nodes.5 Furthermore, patients with organ-confined disease are less likely to relapse beyond 5 years, and unlikely beyond 10 years after cystectomy, even without adjuvant treatment [11].

These trends are derived from focused studies with disparate cohorts. The large size of the current dataset offers a chance to confirm and refine these relations. Beyond knowledge discovery, larger and electronically managed medical databases lend to predictive tool development. Consequently, machine learning techniques have been applied extensively on clinical, epidemiological, and molecular data to predict prognosis and outcome in various cancers. Cruz and Wishart [43], and more recently, Kourou et al. [108] offer a review of some of these studies which predict of susceptibility, recurrence and survival, where the merit of techniques and the quality of the data are quantified by prediction accuracies. In BCa, the most relevant existing study used a multi-institution dataset of 9000 patients, including 980 data points from the present dataset, to construct a nomogram for predicting 5-year recurrence which achieved a concordance index of 0.75 [21].

The present work focuses on: (i) using preoperative and operative BCa data to uncover patterns of long term outcomes and (ii) assessing the predictive power of BCa-specific factors in elucidating overall survival (OS) and recurrence. We employ the information theory concept of mutual information (MI) to uncover correlated parameters. We then stratify the set of predictors by correlation with recurrence and OS to quantify their relative importance. The prognostic power of these variables is assessed by developing a machine-learning classification pipeline to predict recurrence and survival after radical cystectomy, urinary diversion and extended lymphadenectomy, the standard-of-care for high-risk, muscle-invasive BCa. The models presented deliver a quantitative method for stratifying patients into higher resolution risk groups than is possible with current methods.

4.3 Methods

4.3.1 Data summary

The original dataset (details in Table 14) comprised of 3503 patients is pruned to 3499 (mean age 67.8 years) patients by removing 4 cases with missing survival data. All patients underwent radical cystectomy at the USC Institute of Urology from 1971 to 2016. Statistical results based on this dataset up to 1997 were published by Stein et al. in 2001 [182] on a subset of 1054 urothelial carcinoma patients. Presently, this is one of the largest known single-institute datasets of BCa cystectomy patients in terms of sample size and the 45-year timespan over which the data were prospectively and continuously collected with institutional review board (IRB) approval. Consequently, the evolution of preoperative and operative assessments is also explored. In addition to information pertinent to BCa, comorbidity data were also collected to study the effect of preexisting diseases on progression of BCa. Remainder of the data is comprised of demographics, clinical diagnostic information prior to cystectomy, tumor markers prior to cystectomy, and pathologic and surgical data at time of cystectomy including adjuvant therapy treatment information. In the context of machine

learning, these preoperative and operative measurements are called predictors, and the target variables are binary indicator variables for recurrence and survival after a given number of years post-cystectomy.

4.3.2 Statistics and information theory

We perform survival analysis using the Kaplan-Meier estimator to differentiate OS by various predictors. However, to develop an understanding of system-wide patterns between all the predictors, recurrence, and OS, a network approach is more suitable. Relevance, or correlation networks [29, 127, 90, 189] can be created using a similarity measure. Therefore, we create a mutual information (MI) network and subsequently a Euclidean distance based complete-linkage agglomerative hierarchical clustering of the most closely associated variables. Here, we use normalized MI which ranges from 0 to 1 for entirely unrelated to maximally related pairs of variables. Larger values of MI correspond to higher dependence between two variables. We normalize MI by the maximum entropy of the two variables being compared [121], and normalized MI will be abbreviated as MI. The set of all pairwise MI relations make up the adjacency matrix of the MI network, which is visualized as a clustered heat-map.

The predictors are ranked by their association with the two long term outcomes, recurrence and OS, using the chi-squared test of independence which measures the association between two categorical variables. OS, age, and other continuous variables are discretized to perform the chi-squared test. The composite assessment identifies higher fidelity variables and encapsulates the clinical relevance of the measurements. A composite predictor ranking,

$$rank_i = \sqrt{(\bar{\chi}^2_{Rec,i})^2 + (\bar{\chi}^2_{OS,i})^2} \tag{56}$$

based on the chi-squared values for recurrence (χ^2_{Rec}),

$$\bar{\chi}^2_{Rec,i} = \frac{\chi^2_{Rec,i}}{\sigma_{\chi^2,\mathrm{Rec}}} \tag{57}$$

and the OS chi-squared values (χ^2_{OS}),

$$\bar{\chi}^2_{OS,i} = \frac{\chi^2_{OS,i}}{\sigma_{\chi^2,\mathrm{OS}}} \tag{58}$$

is used to identify predictor importance. The chi-squared values for both outcomes are normalized by their respective standard deviations (Eq. 1-3) to weigh the effect of both outcomes equally.

4.3.3 Machine learning approach

The performance of multivariate predictive models is compared to univariate logistic regression models. To create the multivariate models, a series of base predictive models are employed, subsequently mixture-of-experts and stacking based ensemble models are trained using these base models. The base models consist of: support vector machines (SVM), bagged SVM, K-nearest neighbor (KNN), adaptive boosted trees (AdaBoost), random forest (RF), and gradient boosted trees (GBT). The mixture-of-experts models are based on hard-voting among the base models, whereas the stacking ensemble models perform dimensionality reduction of the base model predictions before employing a second logistic regression or support vector machine (SVM) model.

For each prediction task, a different triplet of models forms the final meta-classifier, which is constructed by combining one each of the best base, mixture-of-experts, and stacking classifiers using hard-voting. This method of combining various models is achieves the highest performance metrics.

Patients who leave the study before the target year of the survival models are removed from the dataset, resulting in n = 3201, 3066, and 2780 patients for the 1-, 3-, and 5-year survival datasets respectively. However, only patients who have no recurrences and leave the study before the target year of the recurrence models are removed from the corresponding models' datasets; resulting in n = 3071, 2955, and 2695 patients in the 1-, 3-, and 5-year recurrence datasets respectively. To avoid class imbalance while training, the subset of patients who recur are randomly oversampled to yield an equal count of recurring and non-recurring patients in the training sets. Similarly, for the survival classifiers, the fraction of surviving and non-surviving patients is balanced by random oversampling.

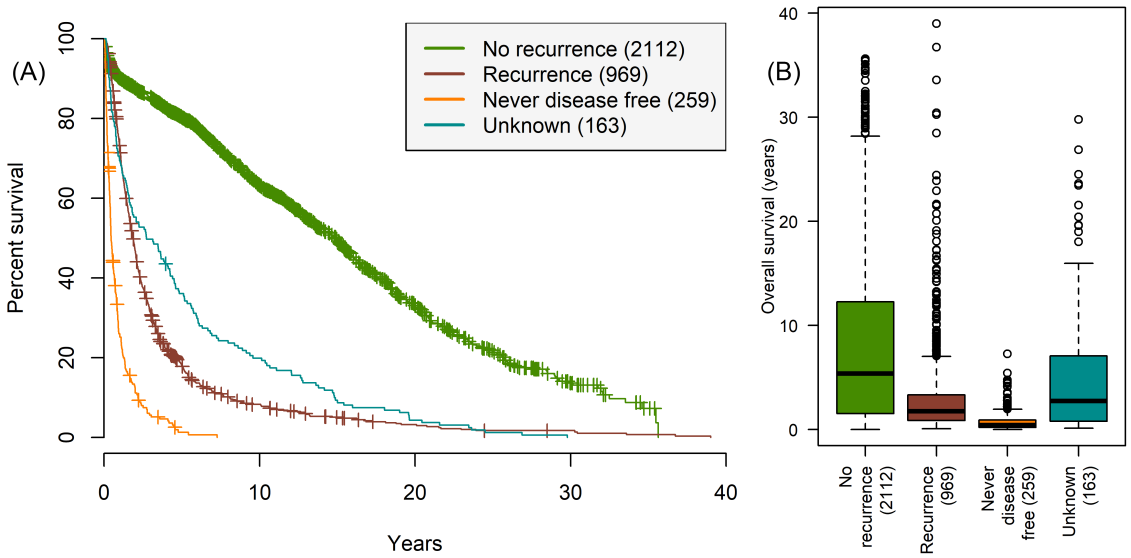
The procedure for feature selection consists of two steps: removal of irrelevant predictors and removal of redundant predictors. To remove redundant predictors, the hierarchical clustering of the 73 predictors is used to define 60 predictor clusters, and the predictor with the highest MI with the target variable is selected from each cluster. Subsequently, to remove irrelevant predictors from the dataset, predictors with low MI with the target variables are excluded from the dataset (MI$<$0.006 for predicting recurrence, MI$<$0.003 for predicting survival). These two feature selection steps yield a set of 52, 54, and 51 predictors for the 1-, 3-, and 5- year recurrence models respectively, and 42, 45, and 45 predictors for predicting 1-, 3-, and 5-year survival respectively.

Final performance scores are found using nested cross-validation with ten outer folds and five inner folds in which the SVM, RF, GBT, and AdaBoost hyper-parameters are tuned. The Scikit-Learn platform [149] is used to implement the models.

4.4 Results

4.4.1 Survival statistics

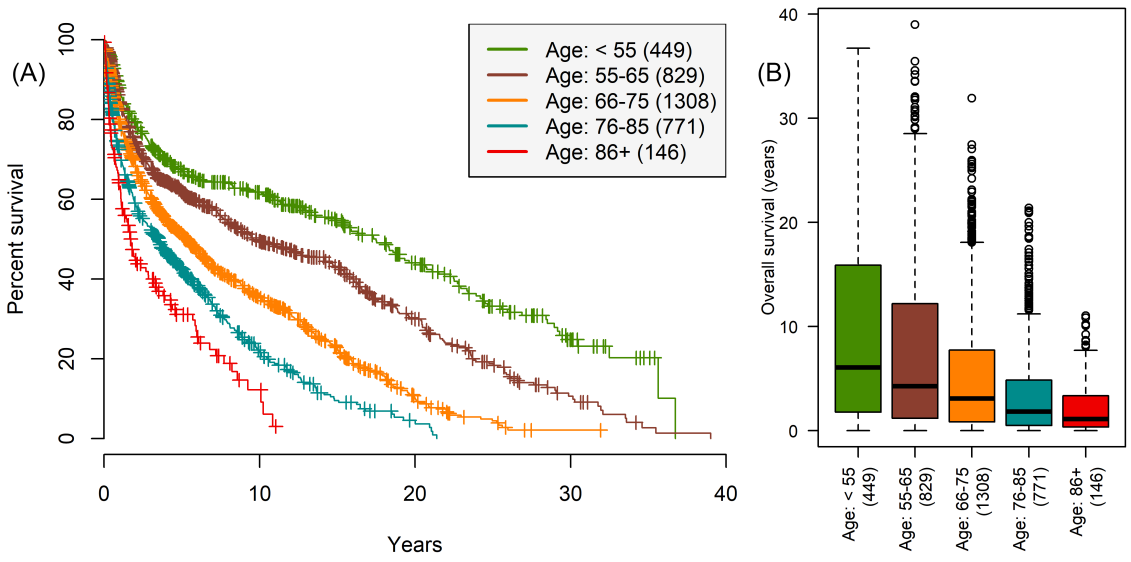
3503 patients' OS is outlined in Fig.25 and patients with unknown recurrence status are excluded from analysis pertaining to recurrence in the rest of the study. There is an exponential decay in survival by age groups in the five-year period post-cystectomy, which suggests the burden of BCa diminishes significantly within five years for patients undergoing radical cystectomy (Fig.26). Consequently, our prediction tasks focus on 1-, 3-, and 5-year survival and recurrence.



at risk (# censored)

No recurrence	2112 (0)	741 (843)	190 (1120)	28 (1198)	0 (1215)
Recurrence	969 (4)	54 (87)	13 (101)	5 (103)	0 (103)
Never disease free	259 (0)	0 (19)	0 (19)	0 (19)	0 (19)
Unknown	163 (0)	32 (1)	7 (1)	0 (1)	0 (1)

Figure 25: Relationship between disease status and survival. (A) Kaplan-Meier survival for patients who never recurred (green, n = 2112, 60.3%) have the highest mean OS 7.93 years (SD = 7.69), patients who were initially disease free but then had recurrence (red, n = 969, 27.7%) have mean OS 3.00 years (SD = 4.14), patients who were never free (orange, n = 259, 7.4%) of the disease have the lowest mean OS 0.81 years (SD = 1.03), and patients whose post-cystectomy progression is unknown (teal, n = 163, 4.7%). (B) Boxplot of OS by disease progression of patients. (C) Histogram of OS in cohort show a dramatic reduction in survival rates up to five years post-cystectomy.



at risk (# censored)

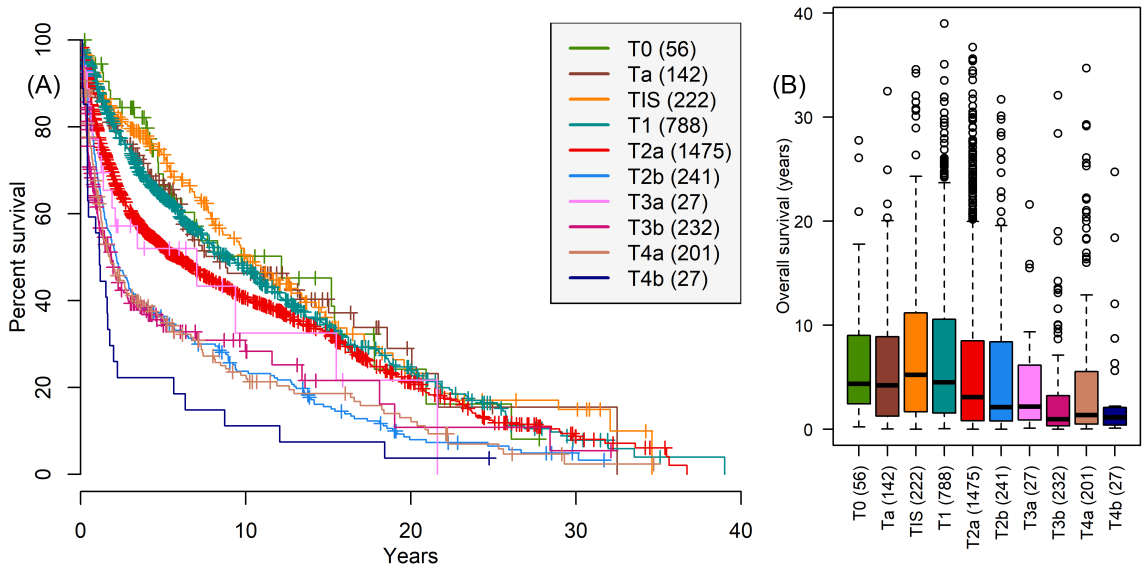
Age: < 55	449 (0)	190 (103)	83 (169)	18 (207)	0 (221)
Age: 55-65	829 (1)	243 (233)	87 (317)	14 (345)	0 (349)
Age: 66-75	1308 (0)	257 (349)	35 (436)	1 (450)	0 (451)
Age: 76-85	771 (2)	66 (250)	5 (272)	0 (272)	0 (272)
Age: 86+	146 (1)	5 (45)	0 (47)	0 (47)	0 (47)

Figure 26: Relationship between age and survival. patients are stratified into five age groups, according to the Surveillance, Epidemiology, and End Results (SEER) program age standards for survival [4]. (A) Boxplot of OS for patients by age group. Age < 55 (n = 449 patients, 12.8%), $55 \leq$ age < 65 (n = 829 patients, 23.7%), $65 \leq$ age < 75 (n = 1308 patients, 37.3%), $75 \leq$ age < 85 (n = 771 patients, 22.0%), $85 \leq$ age (n = 146 patients, 4.2%). (B) Kaplan-Meier survival by disease progression shows disease burden greatly diminishing five years after surgery.

Comparing survival for clinical staging prior to surgery (Fig.27) and pathologic staging (pT staging: TNM 5th edition staging) at time of cystectomy (Fig.28) reveals the higher fidelity of pathologic staging. Clinical staging fails to separate staging as clearly as pathologic staging, for example clinical staging does not separate T2b and T3a patients as clearly as pathologic staging P2b and P3a patients.

During the study (1971-2016) there is a 24% (811/3417) agreement between the two staging measures, with clinical staging over-estimating pathologic stage by 25% (865/3417 patients). Since 2010, there is a surge in stage over-estimation, with a corresponding decrease in under-estimation; however, overall concordance between the two staging measures has remained relatively constant over the decades studied (Fig.31). A graver consequence of the inferior resolution of clinical staging is that it underestimated pathologic stage in 51% of patients (1741/3417).

Patients with organ-confined (OC), extra-vesical (EV) and node-positive (N+) BCa had 5-year survival rates of 0.750 (95% CI [0.729, 0.772]), 0.413 (95% CI [0.377, 0.452]), and 0.213 (95% CI [0.186, 0.243]),



at risk (# censored)

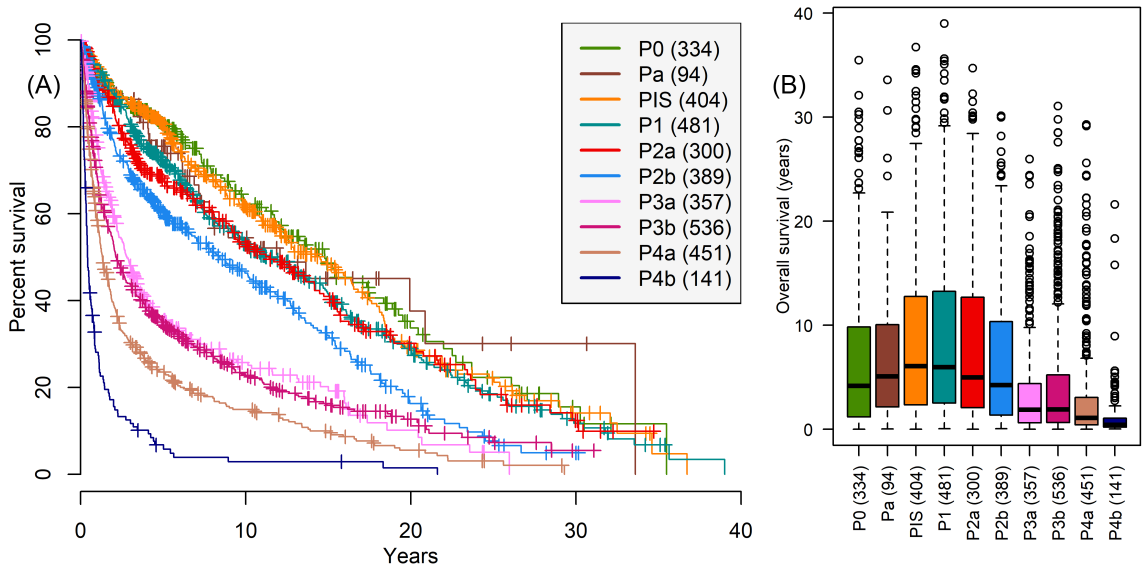
T0	56 (0)	12 (24)	3 (29)	0 (30)	0 (0)
Ta	142 (0)	35 (49)	5 (73)	1 (75)	0 (75)
TIS	222 (0)	63 (74)	15 (99)	7 (102)	0 (107)
T1	788 (1)	212 (244)	59 (316)	7 (340)	0 (344)
T2a	1475 (2)	324 (423)	94 (533)	12 (573)	0 (579)
T2b	241 (0)	48 (14)	12 (21)	3 (26)	0 (28)
T3a	27 (0)	3 (10)	1 (11)	0 (11)	0 (11)
T3b	232 (0)	11 (91)	2 (96)	1 (96)	0 (97)
T4a	201 (0)	30 (33)	14 (38)	1 (42)	0 (43)
T4b	27 (0)	3 (0)	1 (0)	0 (1)	0 (1)

Figure 27: Relationship between survival and clinical T stage assigned prior to cystectomy. Based on a combination of imaging and transurethral resection. (A) Kaplan-Meier survival shows some degradation with tumor staging, but many stages overlap significantly. T3a and T2b patients in the (B) boxplot of OS by T stage are not differentiated. P-values in Table 19.

respectively (p<0.001) (Fig.32). Patients with and without lympho-vascular invasion had 5-year survival rate of 0.299 (95% CI [0.271, 0.330]) and 0.637 (95% CI [0.617, 0.658]), respectively (p<0.001) (Fig.33).

Patients with a negative soft tissue surgical margin had 5-year survival of 0.572 (95% CI [0.553, 0.591]) compared to 0.358 (95% CI [0.285, 0.448]) for patients with positive ureteral/urethral margins and 0.063 (95% CI [0.034, 0.117]) for patients with soft-tissue margins ($p < 0.001$) (Fig.34).

Patients with carcinoma in situ had no discernable difference in OS as well as 5-year probability of survival compared to other patients.



at risk (# censored)

T0	56 (0)	12 (24)	3 (29)	0 (30)	0 (0)
Ta	142 (0)	35 (49)	5 (73)	1 (75)	0 (75)
TIS	222 (0)	63 (74)	15 (99)	7 (102)	0 (107)
T1	788 (1)	212 (244)	59 (316)	7 (340)	0 (344)
T2a	1475 (2)	324 (423)	94 (533)	12 (573)	0 (579)
T2b	241 (0)	48 (14)	12 (21)	3 (26)	0 (28)
T3a	27 (0)	3 (10)	1 (11)	0 (11)	0 (11)
T3b	232 (0)	11 (91)	2 (96)	1 (96)	0 (97)
T4a	201 (0)	30 (33)	14 (38)	1 (42)	0 (43)
T4b	27 (0)	3 (0)	1 (0)	0 (1)	0 (1)

Figure 28: Relationship between survival and pT stage (TNM 5th edition) assigned at time of cystectomy. (A) Kaplan-Meier survival shows a steadier degradation with tumor staging than clinical staging. (B) Boxplot of OS by pT stage at time of cystectomy clearly differentiates P3a and P3b patients from stage two patients. P-values in Table 20.

4.4.2 Correlations among predictors

Fig.29 shows the MI network adjacency matrix as a heat-map and the hierarchical clustering of variables in the dataset, where a four-cluster division is highlighted. Mean MI within the four clusters (purple, green, blue, red) is 0.629, 0.233, 0.0237, and 0.653, respectively, therefore the blue cluster is comprised of significantly less correlated variables than the other clusters. The purple cluster consists of pathologic staging variables which form the most correlated set of BCa-specific predictors (1-6 in Fig.29). BCa variables (purple, green and blue clusters) have an average MI of 0.0337 with each other, and an average MI of 0.00575 with the set of comorbidity variables (red cluster), marking the first division in the clustering. This suggests that patients' preexisting comorbidities are not strongly related to BCa variables.

In contrast to comorbidity factors which are strongly associated with each other, the mean MI between

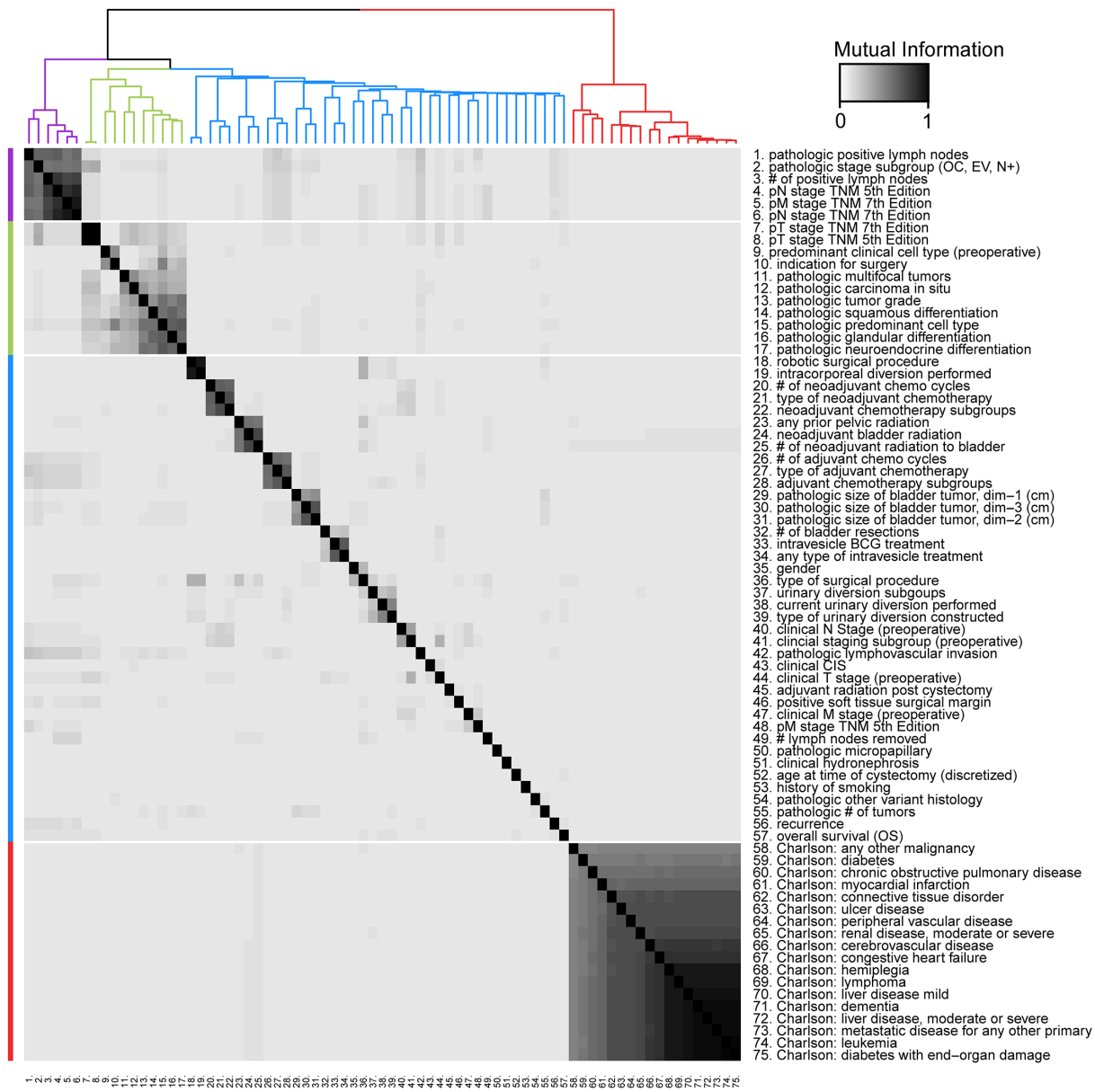


Figure 29: MI between the set of predictors as well as recurrence and OS. Predictors are clustered into four groups using a hierarchical clustering algorithm to discover associated predictor groups. BCa predictors and long-term outcomes are contained in the purple (anatomic staging), green (histologic staging), and blue (treatment and OS and recurrence) clusters, whereas the comorbidity factors comprise a solitary (red) cluster. Correlations within the purple and red clusters are high, but correlations between the comorbidity cluster and other clusters is low.

predictors and the binary recurrence target variable (56 in Fig.29) is 0.01028, and 0.01229 between predictors and discretized OS (57 in Fig.29), highlighting the difficulty in predicting long term BCa outcomes by using only preoperative and operative data. The association between discretized OS and recurrence, MI = 0.0482, is much higher than the mean MI between the predictors and either long term outcome. Both long-term

outcomes are in the blue cluster (Fig.29) where there is a lack of strong associations among the variables aside from three sets of variables related to neoadjuvant chemotherapy (20-22 in Fig.29), adjuvant chemotherapy (26-28 in Fig.29), and radiation (23-25 in Fig.29) which are highly related because they are clinical re-classifications or sub-groupings of each other within the original variable's domain. Aside from TNM 7th edition staging (7 in Fig.29) which is nearly identical to pT staging (TNM 5th edition, 8 in Fig.29), the highest associations with pT staging are pathologic stage subgroup (2 in Fig.29, MI = 0.255) and presence of pathologic carcinoma in situ (12 in Fig.29, MI = 0.1825). In contrast, clinical T stage (44 in Fig.29) does not have equally high MI with any of the other predictors except for the regrouped clinical staging variable (41 in Fig.29). The MI based heat-map and clustering in Fig.29 provides a system-wide view of the entire medical database for BCa patients and correlations between the predictors can be used to assess the quality of clinical measurement techniques.

4.4.3 Correlations with long-term outcomes

The chi-squared test of independence for recurrence (vertical axis) and OS (horizontal axis) in Fig.30 and Table 22 shows the relative importance of each predictor. The variance in chi-squared values is computed by singular value decomposition [5] and is shown by an ellipse whose axes are the standard deviations (SD) along the first (SD = 195.4) and second (SD = 76.8) principal components (green lines) in Fig.30A. Some predictors intrinsically contain more information about survival than recurrence, and vice versa, because the variances along the principal components are of comparable size. For instance, urinary diversion (rank 10) and age (rank 12) are strongly correlated with OS but not recurrence, nevertheless, they rank high due to large effect on survival.

Pathologic stage subgroup (rank 3), which indicates whether patients have OC, EV, or N+ disease at time of cystectomy, has the largest χ^2_{Rec}, and its χ^2_{OS} is the fourth highest. Pathologic stage (rank 1) is even more strongly associated with OS and comparing it to clinical staging (rank 17) reinforces the superiority of pathologic staging in differentiating patients by outcome as observed in the Kaplan-Meier curves (Fig.27A and Fig.28A). The number of positive lymph nodes removed at time of cystectomy (rank 7) is significantly more correlated with recurrence than the total number of lymph nodes removed (rank 16), however, further data may be required for clarification due to the high p-value corresponding to χ^2_{Rec} for total number of removed lymph nodes. The predictors ranked 4-8 in Fig.30 have similar correlations with either outcome, and these predictors are part of the highly associated (purple) cluster in Fig.29. The Cramer's V effect size of pathologic positive lymph nodes (rank 8) and lymphovascular invasion (rank 14) are particularly high for both outcomes (Fig.36).

2757 patients who did not receive adjuvant chemotherapy had 5-year survival rate of 0.563 (95% CI [0.543,

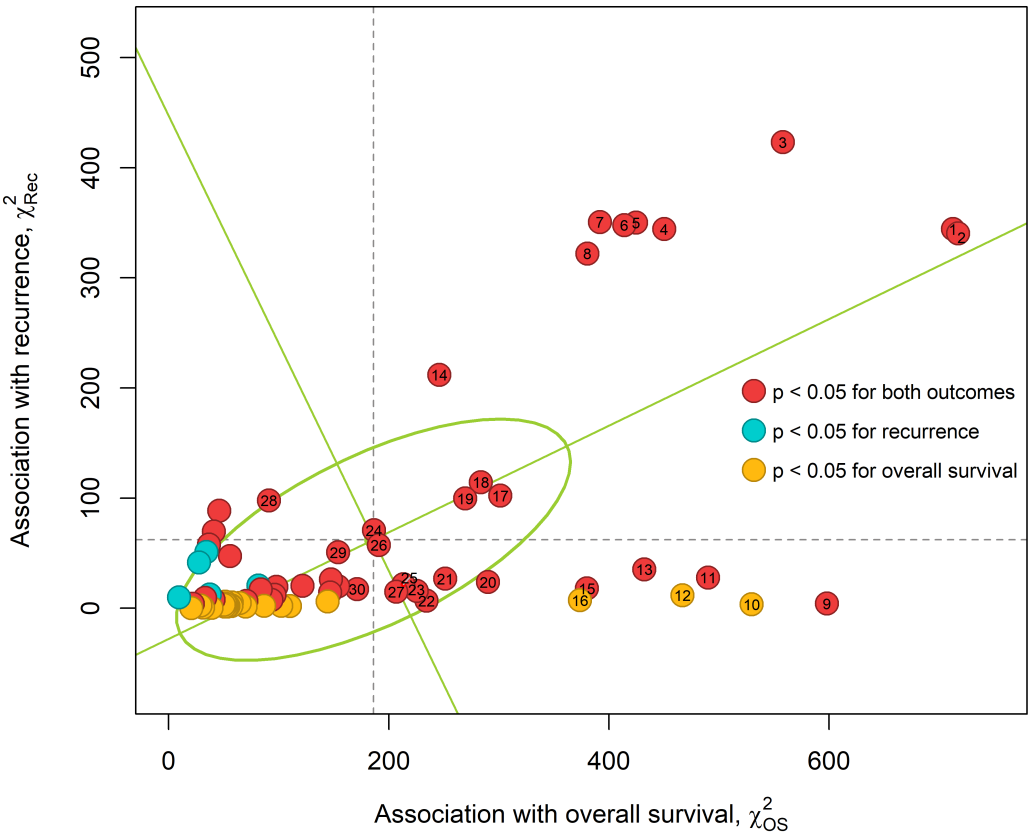


Figure 30: Association between predictors and the two long term outcomes, recurrence and OS as measured by the chi-squared test of independence. The predictors which have a statistical significance of $p < 0.05$ for both outcomes (43 predictors) are colored in red. Predictors of significance ($p < 0.05$) with recurrence (5 predictors) or OS (24 predictors) are shown in blue and yellow, respectively. The solid (green) lines are the two principal axes computed using singular value decomposition and the singular values are used to define the semi-major and semi-minor axes of the solid (green) ellipse. The horizontal and vertical dashed (gray) lines indicate the mean chi-squared for recurrence and OS respectively.

0.583]), and 633 patients who did receive adjuvant chemotherapy had a 5-year survival rate of 0.450 (95% CI [0.411, 0.493]) (Fig.35). Since the prescription of adjuvant chemotherapy is limited to a homogeneous set of patients who are node positive at time of cystectomy, the association of adjuvant chemotherapy with recurrence and OS may be artificially high in this dataset.

Although the predictors with below average chi-squared values are less important than the others, they may still differentiate patients who are similar in the higher ranked variables. The lowest correlates of OS and recurrence are gender ($\chi^2_{OS} = 9.53$) and intravesical treatment ($\chi^2_{Rec} = 0.1046$) respectively. Overall the predictors have a weaker association with recurrence than OS. Since the MI between BCa specific predictors is small, even the predictors with small chi-squared values add new information about a patient. However, this new information may not necessarily inform OS and recurrence.

		Recurrence		**Overall survival**	
Rank	Predictor	χ^2	p-value	χ^2	p-value
1	pT stage TNM 7th Edition	344.3	0.000	713.0	0.000
2	pT stage TNM 5th Edition	340.3	0.000	717.0	0.000
3	pathologic stage subgroup (OC, EV, N+)	423.1	0.000	558.0	0.000
4	pM stage TNM 7th Edition	344.2	0.000	450.0	0.000
5	pN stage TNM 7th Edition	350.0	0.000	425.0	0.000
6	pN stage TNM 5th Edition	347.9	0.000	414.0	0.000
7	# of positive lymph nodes	350.6	0.000	392.0	0.000
8	pathologic positive lymph nodes	322.0	0.000	380.0	0.000
9	type of urinary diversion constructed	4.2	0.041	598.0	0.000
10	current urinary diversion performed	3.4	0.064	530.0	0.000
11	pathologic # of tumors	27.8	0.000	490.0	0.000
12	age at time of cystectomy (discretized)	11.7	0.230	467.0	0.000
13	type of surgical procedure	35.2	0.000	432.0	0.000
14	pathologic lymphovascular invasion	211.9	0.000	246.0	0.000
15	urinary diversion subgoups	17.9	0.000	380.0	0.000
16	# lymph nodes removed	7.2	0.407	374.0	0.000
17	clinical T stage (preoperative)	102.2	0.000	301.0	0.000
18	adjuvant chemotherapy subgroups	114.4	0.000	284.0	0.000
19	type of adjuvant chemotherapy	99.6	0.000	270.0	0.000
20	neoadjuvant chemotherapy subgroups	23.7	0.000	290.0	0.000
21	positive soft tissue surgical margin	26.8	0.000	251.0	0.000
22	pM stage TNM 5th Edition	6.8	0.034	234.0	0.000
23	intracorporeal diversion performed	15.6	0.000	225.0	0.000
24	# of adjuvant chemo cycles	70.8	0.000	187.0	0.000
25	# of neoadjuvant chemo cycles	21.7	0.003	213.0	0.000
26	clincial staging subgroup (preoperative)	57.3	0.000	191.0	0.000
27	robotic surgical procedure	14.9	0.000	207.0	0.000
28	pathologic predominant cell type	97.8	0.000	91.2	0.001
29	clinical N Stage (preoperative)	50.7	0.000	154.0	0.000
30	type of neoadjuvant chemotherapy	17.3	0.000	171.0	0.000

Table 12: The ranked list of predictors by importance. Association between predictors and the two long term outcomes, recurrence and OS as measured by the chi-squared test of independence. Predictors are ranked by Eq.56.

4.4.4 Predicting post-cystectomy recurrence

We evaluate the performance of machine learning models to predict post-cystectomy disease recurrence using preoperative and operative data as well as the type and number of adjuvant therapy cycles administered (Table 13). Both univariate (logistic regression) and more complex multivariate models (meta-classifiers in Table 13) are used to predict 1-, 3-, and 5-year recurrence. Pathologic stage subgroup (rank 3 in Fig.30) and pT stage (rank 2 in Fig.30) are used to create the univariate models and these have lower precision and F1 scores than the meta-classifiers. Furthermore, the single predictor models tend to suffer from imbalance between sensitivity and specificity. In contrast, all recurrence meta-classifiers have sensitivities and specificities over 70%. F1 scores improve with year perhaps due to a more even number of positive and negative

recurrence cases in the corresponding datasets.

Post-cystectomy outcome	Model	Year	Test set performance metrics			
			Sensitivity	Specificity	Precision	F1
Recurrence	Meta-classifier	1	0.739	0.714	0.388	0.508
	pT stage TNM 5th Edition	1	0.761	0.653	0.349	0.478
	pathologic stage subgroup	1	0.826	0.593	0.332	0.473
	Meta-classifier	3	0.720	0.708	0.535	0.613
	pathologic stage subgroup	3	0.774	0.631	0.493	0.602
	pT stage TNM 5th Edition	3	0.670	0.694	0.503	0.574
	Meta-classifier	5	0.700	0.702	0.588	0.636
	pathologic stage subgroup	5	0.744	0.611	0.537	0.623
	pT stage TNM 5th Edition	5	0.619	0.698	0.553	0.583
Survival	Meta-classifier	1	0.741	0.770	0.473	0.577
	pT stage TNM 5th Edition	1	0.739	0.672	0.387	0.506
	pathologic stage subgroup	1	0.805	0.602	0.362	0.499
	Meta-classifier	3	0.722	0.788	0.700	0.711
	pathologic stage subgroup	3	0.762	0.691	0.628	0.688
	pT stage TNM 5th Edition	3	0.696	0.739	0.646	0.670
	Meta-classifier	5	0.741	0.768	0.780	0.760
	pathologic stage subgroup	5	0.730	0.717	0.742	0.735
	pT stage TNM 5th Edition	5	0.664	0.766	0.759	0.708

Table 13: Performance of machine learning models for predicting recurrence and survival. Single predictor (pT stage and pathologic stage subgroup classifiers) and multiple predictor (Meta-classifier) models for predicting 1-, 3-, 5-year recurrence and survival after cystectomy. The performance of all models for a given year is ranked per F1 scores (2*precision*recall/(precision+recall)) as well as mean sensitivity, specificity, and precision on test sets from a 10-fold cross validation.

4.4.5 Predicting post-cystectomy survival

Like the recurrence predictions, the meta-classifiers outperform the univariate models in predicting survival (Table 13), however the disparity is greater as the meta-classifiers have considerably higher performance metrics for all year predictions. Additionally, and unlike the recurrence models, the survival meta-classifiers have comparable precision and probability of detection, except for the 1-year survival models. The combination of high precision and sensitivity leads to significantly higher F1 scores for the 3- and 5-year survival meta-classifiers.

4.5 Discussion

Although recurrence and OS are highly associated, preoperative and operative measurements generally do not relate equally to recurrence and OS, and the two outcomes should be assessed separately. The primary predictors of long-term outcomes are pathologic stage and its subgrouping into localized or metastatic conditions. However, the machine learning pipeline developed here can leverage less powerful predictors to improve accuracy of long term predictions. The benefit of having low MI between variables means that each

variable offers unique information however the drawback is that each patient needs to be described by many variables and thus the prediction task becomes a higher dimensional problem, for which lack of data can greatly limit predictions of long-term outcomes. Clinical T stage offers a lower resolution signal than the true pathologic T stage, and this loss of information can be particularly impactful in cases where there is an underestimation of disease severity prior to surgery [185].

The sensitivity and specificity of all the survival meta-classifiers, and the 1-year recurrence meta-classifiers are considerably higher than 70%. Recurrence meta-classifiers are less accurate, perhaps, because of undetected metastatic disease at the time of cystectomy. 1-year meta-classifier predictions for both outcomes offer a better combination of sensitivity and specificity than the 3- and 5-year meta-classifiers. However, the later year models may also be used in the clinical setting to differentiate lower- and higher-risk patients due to higher precision scores.

In current clinical practice, post-radical cystectomy prognostication in the individual patient is informed by the best-evidence found in the literature [182, 155] which reflect probable outcomes in cohorts not the individual, or the prognostic nomogram which only calculates a 5-year outcome [21]. To improve upon this, we employ machine-learning algorithms to construct novel, patient prognostication models for survival and recurrence. Presently the international bladder cancer nomogram has proven to be a validation of multivariate approaches in predicting long term outcomes in the clinical setting[21, 200], and the models developed here offer higher resolution predictions which can assist post-cystectomy treatment and screening decisions.

Despite several machine learning research efforts in predicting outcomes of cancer patients there is a low penetration of such models in clinical practice [108]. There are two specific hurdles before current models can be deployed in a clinical setting, first, because the performance reported here reflects the quality of data collected at one center and to ensure the generalizability of the models, data from other institutions should also be studied.

Combining additional datasets, such as the international BCa dataset [21], may also improve the performance of the algorithms due to general sparsity and low frequency of certain combinations of predictors in the present data. Secondly, the recurrence and survival models use a total of 42-54 predictors, therefore the standardized collection of these parameters must be ensured before the machine learning models can be deployed successfully in a clinical setting.

The accuracy of predicting cancer recurrence, which may depend on several evolutionary steps beyond cystectomy, can undoubtedly be improved by combining genomic and molecular data, and this would be fruitful direction to pursue. The quality of the dataset, coupled with machine learning models in the present work, offers a benchmark of the value of current preoperative and operative patient assessment standards

with respect to forecasting long term outcomes during the most vulnerable 5-year timespan in BCa treatment post-cystectomy. Furthermore, due to the absence of widely recognized biomarkers for BCa [32], clinicopathological-based predictions of clinical outcomes as shown here set the standard for long-term personalized predictions in BCa. If deployed correctly, machine learning models can transform preoperative and operative data into accurate predictions and mitigate post-cystectomy burden of BCa.

4.6 Supporting information

	All (3503)	**P0** (334)	**Pa** (94)	**PIS** (404)	**P1** (481)	**P2a** (300)	**P2b** (389)	**P3a** (357)	**P3b** (536)	**P4a** (451)	**P4b** (141)
Age (yr)											
Median	68	66	67	67	66	67	67	70	69	70	68
Range	23-95	25-92	34-87	27-94	33-95	31-89	30-92	38-91	31-92	38-94	23-95
Gender (n)											
Female	757	66	16	62	91	52	99	97	171	70	30
Male	2746	268	78	342	390	248	290	260	365	381	111
Histology (n)											
Adeno	40	3	1	0	4	1	6	7	7	7	4
Neuro	42	7	0	0	0	3	5	5	12	10	0
Sarcoma	19	2	0	0	1	5	3	3	1	3	1
Squamous	90	3	0	2	5	5	16	7	27	13	11
Urothelial	3301	315	92	402	470	286	358	335	486	417	125
Other	11	4	1	0	1	0	1	0	3	1	0
pN stage (n)											
Nx	151	14	5	11	16	8	8	5	16	30	32
N0	2477	309	88	381	435	248	293	206	284	181	49
N1	220	3	0	5	14	18	29	41	59	37	11
N2	639	7	1	7	16	25	59	101	176	197	46
N3	16	1	0	0	0	1	0	4	1	6	3
Pathologic subgroup (n)											
OC	1812	323	93	391	450	254	299	0	0	0	0
EV	760	0	0	0	0	0	0	208	293	201	58
N+	931	11	1	13	31	46	90	149	243	250	83
Radiation (n)											
Neoadjuvant	325	35	7	34	50	26	33	20	35	54	28
Adjuvant	54	0	0	0	2	2	1	5	14	8	20
Chemotherapy (n)											
Neoadjuvant	519	93	7	59	32	32	39	52	62	89	52
Adjuvant	633	8	3	12	26	37	80	106	178	142	36
Smoker (n)											
Current	674	52	19	61	88	71	86	61	127	74	35
Never	807	84	18	82	117	57	92	94	132	94	34
Previous	1891	188	55	249	255	157	195	189	264	267	63
Metastasis (n)											
Total Patients	1061	36	16	67	96	68	100	154	242	231	48
Bone	332	6	1	14	27	18	26	54	91	80	14
Pelvis	308	8	1	15	13	14	33	48	83	70	23
Lung	288	11	4	19	29	24	29	43	70	49	10
Liver	255	12	2	20	24	8	19	35	68	56	10
Retro	208	6	0	21	19	12	14	42	44	43	7
LN (reg)	134	3	2	9	16	7	9	23	27	30	7
LN (dist)	95	4	0	5	5	12	9	19	14	22	5
Peritoneum	93	7	0	6	2	5	8	10	18	28	9
Urethra	88	5	7	15	20	11	5	2	8	14	1
Brain	65	1	0	4	6	2	10	11	14	14	3
UT	61	3	5	11	12	11	2	2	8	7	0
Adrenal	34	0	0	3	2	2	2	6	10	6	3
Other	174	7	2	11	15	14	13	29	37	36	10

Table 14: Dataset description. Patient and tumor characteristics by pathologic T stage.

S1 Appendix. Machine learning process for finding best model. The feature selection process for generating datasets for machine learning is described here. MI is calculated between p = 73 predictors and used to cluster variables into K = 60 clusters. MI is also calculated between p = 73 predictors and RecX (x = 1, 3, 5 years) and SurvX (x = 1, 3, 5 years). Correlated predictors are removed by selecting a single predictor from K = 60 clusters whose MI with the outcome is the highest. Irrelevant predictors are removed by pruning all predictors whose MI with the outcome is below a threshold (MI = 0.006 for RecX, MI = 0.003 for SurvX). These thresholds are selected heuristically, with some trial and error. The choice of number of

clusters is another decision which can be explored with further trial and error computations, however this task was found to be too demanding. Therefore, it may be possible to achieve similar performance with more strict removal of correlated predictors by picking a smaller number of correlated clusters, K to prune from. All p = 73 predictors are categorical except for age, therefore discretized (10 bins with equal frequency) and continuous versions of age were tried separately.

Class-imbalance was overcome by random oversampling to achieve a ratio of minority to majority class k = 0.8, 0.9, or 1.0. Number of patients with recurrences (1) and no recurrences (0) in the 1-, 3-, and 5-year recurrence datasets are: (0: 2469, 1: 602) for 1-year, (0: 2020, 1: 935) for 3-year, (0: 1683, 1: 1012) for 5-year. Number of surviving (0) and not surviving (1) patients in the 1-, 3-, and 5-year survival datasets are: (0: 2503, 1: 698) for 1-year, (0: 1821, 1: 1245) for 3-year, (0: 1318, 1: 1462) for 5-year.

Thirteen base models include: SVM (polynomial, sigmoid, radial basis function kernels), KNN, AdaBoost, gradient boosted trees, random forest, pathologic stage based logistic regression, pathologic stage subgroup based logistic regression, and bagged versions of SVM and KNN. The main challenge in this study was to overcome the skewed performance of base models towards high sensitivity or high specificity. Consequently, a series of ensemble learning techniques are employed to take advantage of the base models. Nine different majority voting based mixture-of-expert (MOE) ensemble models, and 13 different logistic regression and SVM based stacking ensemble models are created using the results of the base models. Finally, the meta-classifiers are created by using a triplet of: a base model, a MOE ensemble model, and a stacking ensemble model. There are 1287 such triplets, and we find the performance of these models using nested k-fold cross validation, and the best models from this list are reported. These ensemble methods prove successful in overcoming the skewed base model performance. Due to the medial nature of the dataset, the predictors are categorical and the addition of numerical predictors directly from imaging and biopsy procedures may improve performance.

S2 Appendix. Clinical vs Pathologic staging trend since 1970s. Clinical staging (staging based on a combination of imaging and transurethral resection) and pathological staging (staging done at time of cystectomy) do not always classify patients in the same stages, and clinical staging may underestimate (shown in blue), agree (shown in green) or overestimate (shown in red). Clinical T stages are: T0, Ta, TIS, T1, T2a, T2b, T3a, T3b, T4a, T4b. Pathologic stages are: P0, Pa, PIS, P1, P2a, P2b, P3a, P3b, P4a, P4b. TX and PX indicate patients whose stage was declared unknown, and these patients are removed for this analysis resulting in a dataset of N = 3417 patients for plementary Figure 1 and all concordance analyses between clinical and pathologic staging.

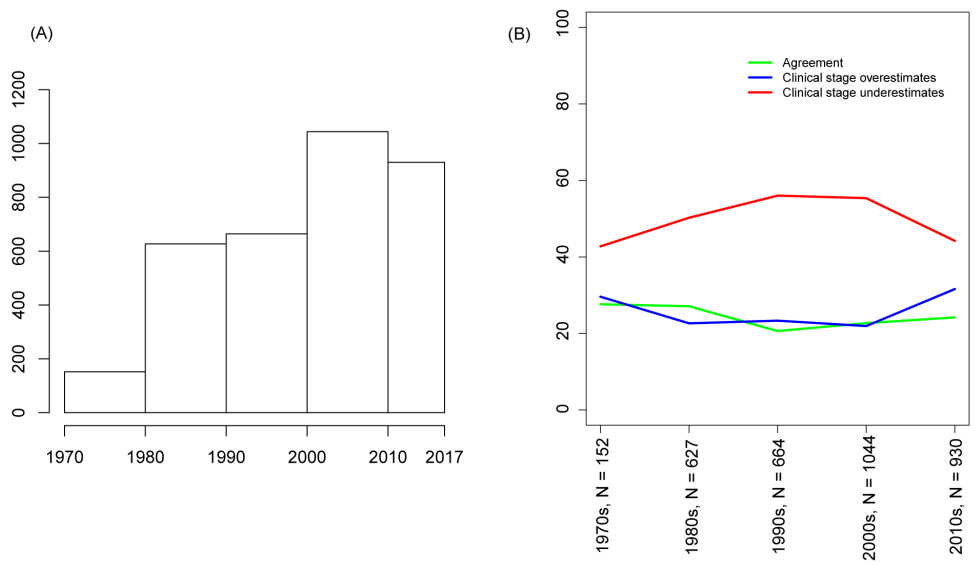
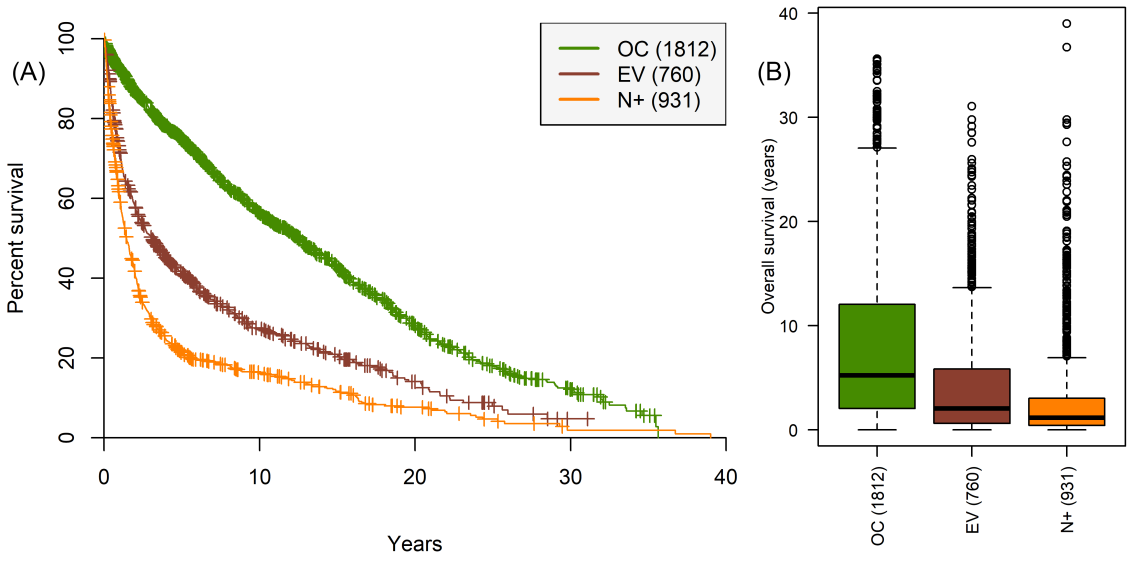


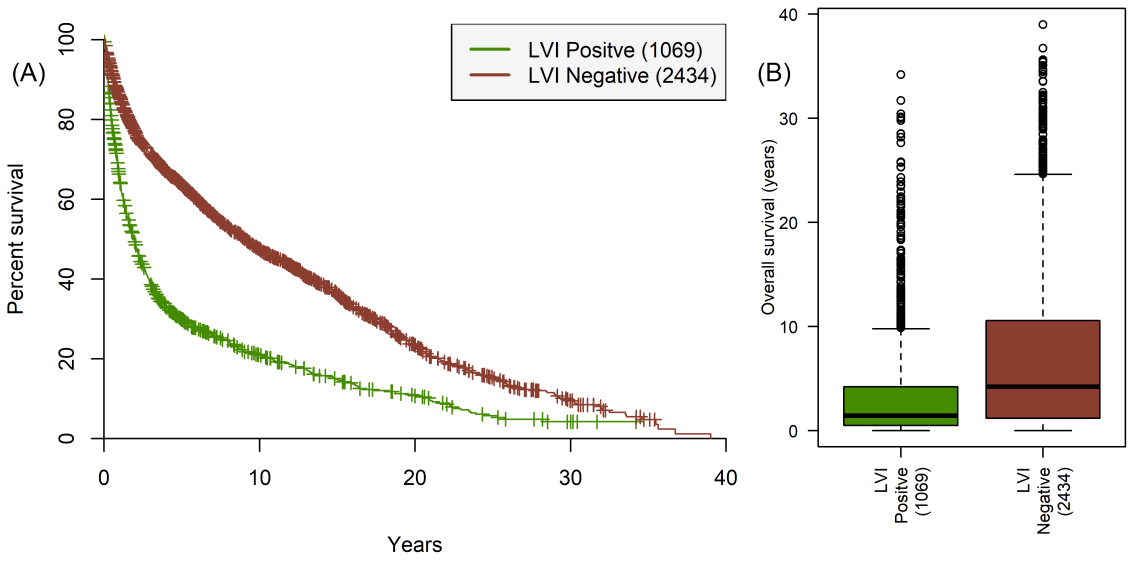
Figure 31: Discrepancy in staging over time. (A) Histogram of number of patients undergoing cystectomy for each decade of data collection. (B) The agreement between clinical and pathological staging over the period of data collection (1971-2016).



at risk (# censored)

OC	1812 (3)	564 (648)	160 (839)	30 (897)	0 (915)
EV	760 (0)	111 (180)	28 (223	1 (236)	0 (237)
N+	931 (1)	86 (152)	22 (179)	2 (188)	0 (188)

Figure 32: Bladder stage survival analysis. (A) Kaplan-Meier survival computed by bladder stage at time of cystectomy subgrouped into OC (T0, Ta, TIS, T1, T2a, T2b), EV (T3a, T3b, T4a, T4b), N+ irrespective of stage (metastasized) (p-value < 0.001). (B) Boxplot of OS by bladder stage at time of cystectomy.



at risk (# censored)

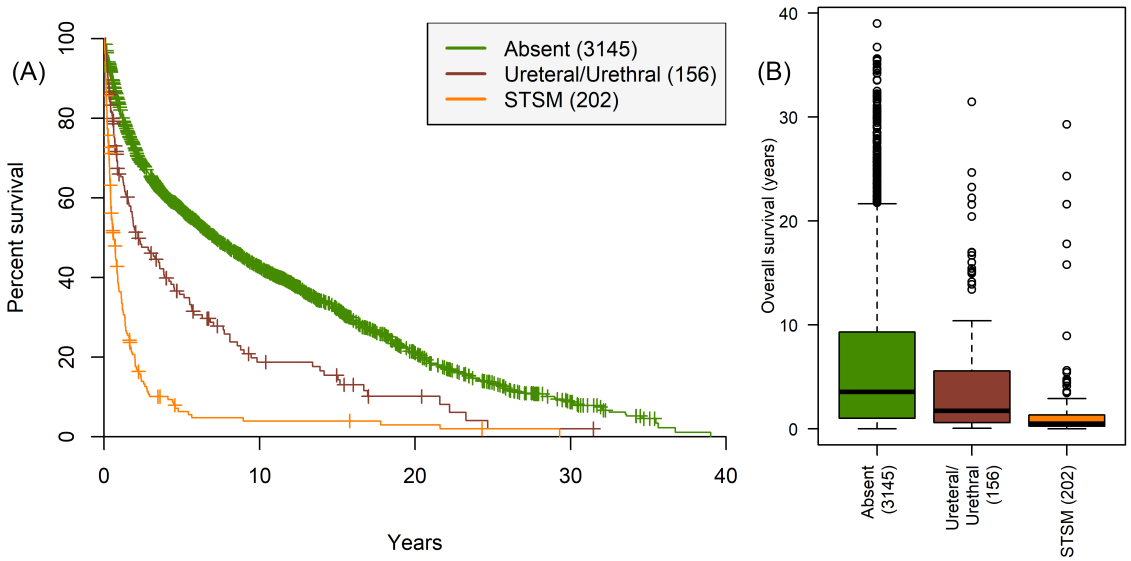
LVI Positive	1069 (1)	116 (218)	38 (249)	4 (265)	0 (269)
LVI Negative	2434 (3)	645 (762)	172 (992)	29 (1056)	0 (1071)

Figure 33: Lymphovascular invasion survival analysis. (A) Kaplan-Meier survival computed by lymphovascular invasion. (B) Boxplot of OS by lymphovascular invasion.

S3 Appendix. Chi-squared test of independence: predictors vs. recurrence and overall survival. Complete results of the Chi-squared test of independence between predictors and recurrence and OS are shown here in Table 15 and Table 16. The continuous variables age and OS are discretized into 10 bins of equal frequency. Predictors are ranked by (Eqs.56, 57, 58). The tests for association with OS are conducted on a dataset of 3499 patients, excluding the 4 patients who are missing survival data. The tests for recurrence exclude 163 patients whose recurrence status is unknown, resulting in a subset of the original data consisting of 3340 patients.

Included here is a ranking of the predictors based on Cramer's V shown in Fig.36 and listed in Table 17 and Table 18. Cramer's V normalizes the chi-squared values by the number of categories in the predictors and offers an alternate ranking of the predictor importance with recurrence and overall survival. A composite predictor ranking,

$$rank_i = \sqrt{(\bar{V}^2_{Rec,i})^2 + (\bar{V}^2_{OS,i})^2} \tag{59}$$



at risk (# censored)

Absent	3145 (4)	738 (927)	201 (1182)	32 (1260)	0 (1278)
Ureteral/ Urethral	156 (0)	18 (29)	6 (34)	1 (35)	0 (36)
STSM	202 (0)	5 (24)	3 (25)	0 (26)	0 (26)

Figure 34: Surgical margin survival analysis. (A) Kaplan-Meier survival computed by soft tissue surgical margin. (B) Boxplot of OS by soft tissue surgical margin.

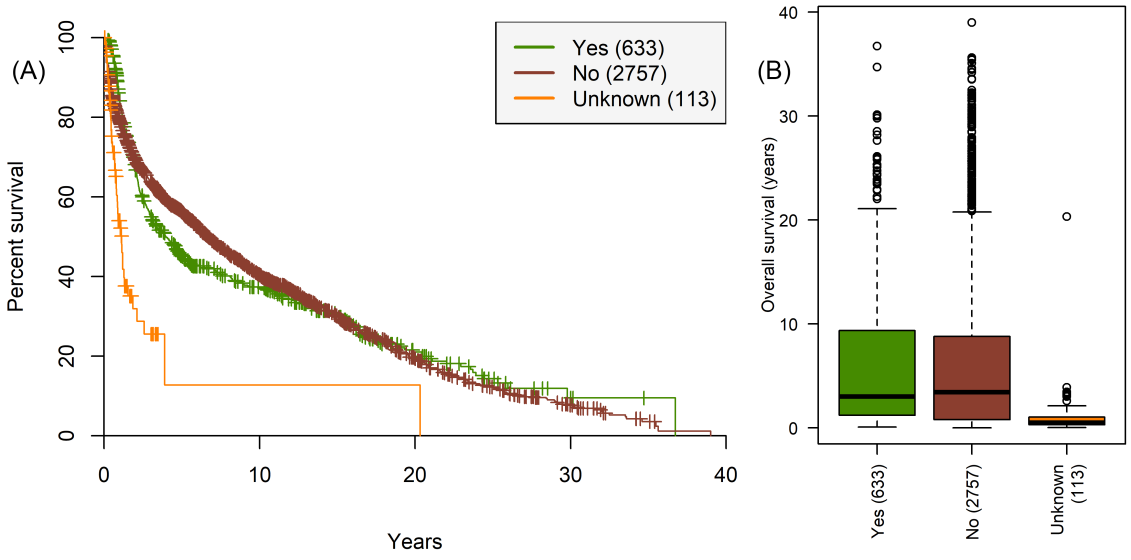
based on Cramer's V values for recurrence (V_{Rec}),

$$\bar{V}_{Rec,i} = \frac{V_{Rec,i}}{\sigma_{V_{\mathrm{Rec}}}} \tag{60}$$

and the Cramer's V values for overall survival (V_{OS}),

$$\bar{V}_{OS,i} = \frac{V_{OS,i}}{\sigma_{V_{\mathrm{OS}}}} \tag{61}$$

is used to identify predictor importance. The Cramer's V values for both outcomes are normalized by their respective standard deviations (Eqs.59, 60, 61) to weigh the effect of both outcomes equally.



at risk (# censored)

Yes	633 (0)	153 (123)	40 (187)	3 (211)	0 (213)
No	2757 (3)	607 (795)	169 (992)	30 (1048)	0 (1065)
Unknown	113 (1)	1 (62)	1 (62)	0 (62)	0 (62)

Figure 35: Adjuvant chemotherapy survival analysis. (A) Kaplan-Meier survival computed by receipt of adjuvant chemotherapy. (B) Boxplot of OS by receipt of adjuvant chemotherapy.

		Recurrence			**Overall survival**		
Rank	Predictor	χ^2	p-value	V	χ^2	p-value	V
1	pT stage TNM 7th Edition	344.3	0.000	0.321	713.0	0.000	0.150
2	pT stage TNM 5th Edition	340.3	0.000	0.319	717.0	0.000	0.151
3	pathologic stage subgroup (OC, EV, N+)	423.1	0.000	0.356	558.0	0.000	0.282
4	pM stage TNM 7th Edition	344.2	0.000	0.321	450.0	0.000	0.160
5	pN stage TNM 7th Edition	350.0	0.000	0.324	425.0	0.000	0.174
6	pN stage TNM 5th Edition	347.9	0.000	0.323	414.0	0.000	0.172
7	# of positive lymph nodes	350.6	0.000	0.324	392.0	0.000	0.137
8	pathologic positive lymph nodes	322.0	0.000	0.310	380.0	0.000	0.330
9	type of urinary diversion constructed	4.2	0.041	0.035	598.0	0.000	0.413
10	current urinary diversion performed	3.4	0.064	0.032	530.0	0.000	0.389
11	pathologic # of tumors	27.8	0.000	0.091	490.0	0.000	0.187
12	age at time of cystectomy (discretized)	11.7	0.230	0.059	467.0	0.000	0.122
13	type of surgical procedure	35.2	0.000	0.103	432.0	0.000	0.117
14	pathologic lymphovascular invasion	211.9	0.000	0.252	246.0	0.000	0.265
15	urinary diversion subgoups	17.9	0.000	0.073	380.0	0.000	0.190
16	# lymph nodes removed	7.2	0.407	0.046	374.0	0.000	0.124
17	clinical T stage (preoperative)	102.2	0.000	0.175	301.0	0.000	0.098
18	adjuvant chemotherapy subgroups	114.4	0.000	0.185	284.0	0.000	0.116
19	type of adjuvant chemotherapy	99.6	0.000	0.173	270.0	0.000	0.196
20	neoadjuvant chemotherapy subgroups	23.7	0.000	0.084	290.0	0.000	0.129
21	positive soft tissue surgical margin	26.8	0.000	0.090	251.0	0.000	0.190
22	pM stage TNM 5th Edition	6.8	0.034	0.045	234.0	0.000	0.183
23	intracorporeal diversion performed	15.6	0.000	0.068	225.0	0.000	0.180
24	# of adjuvant chemo cycles	70.8	0.000	0.146	187.0	0.000	0.094
25	# of neoadjuvant chemo cycles	21.7	0.003	0.081	213.0	0.000	0.093
26	clincial staging subgroup (preoperative)	57.3	0.000	0.131	191.0	0.000	0.165
27	robotic surgical procedure	14.9	0.000	0.067	207.0	0.000	0.243
28	pathologic predominant cell type	97.8	0.000	0.171	91.2	0.001	0.066
29	clinical N Stage (preoperative)	50.7	0.000	0.123	154.0	0.000	0.121
30	type of neoadjuvant chemotherapy	17.3	0.000	0.072	171.0	0.000	0.221
31	pathologic size of bladder tumor, dim-3 (cm)	19.9	0.006	0.077	154.0	0.000	0.079
32	pathologic size of bladder tumor, dim-2 (cm)	26.0	0.000	0.088	147.0	0.000	0.078
33	pathologic carcinoma in situ	88.6	0.000	0.163	46.1	0.000	0.081
34	# of neoadjuvant radiation to bladder	14.1	0.015	0.065	147.0	0.000	0.092
35	pathologic size of bladder tumor, dim-1 (cm)	6.2	0.290	0.043	144.0	0.000	0.091
36	clinical M stage (preoperative)	20.3	0.000	0.078	122.0	0.000	0.132

Table 15: Chi-squared test, ranked by Eq.1-3. Predictors 1-36.

		Recurrence			**Overall survival**		
Rank	Predictor	χ^2	p-value	V	χ^2	p-value	V
37	pathologic multifocal tumors	69.6	0.000	0.144	41.1	0.001	0.077
38	Charlson: renal disease, moderate or severe	1.9	0.397	0.024	110.0	0.000	0.125
39	pathologic other variant histology	19.4	0.007	0.076	97.9	0.003	0.063
40	Charlson: connective tissue disorder	2.2	0.337	0.026	102.0	0.000	0.121
41	pathologic tumor grade	57.8	0.000	0.132	36.8	0.006	0.072
42	Charlson: chronic obstructive pulmonary disease	12.1	0.002	0.060	96.2	0.000	0.117
43	pathologic squamous differentiation	47.4	0.000	0.119	55.8	0.000	0.073
44	Charlson: diabetes	7.4	0.025	0.047	93.5	0.000	0.116
45	pathologic glandular differentiation	51.0	0.000	0.124	34.4	0.156	0.057
46	predominant clinical cell type (preoperative)	20.5	0.005	0.078	81.6	0.057	0.058
47	indication for surgery	17.0	0.005	0.071	83.6	0.000	0.069
48	Charlson: any other malignancy	2.2	0.333	0.026	86.8	0.000	0.111
49	pathologic neuroendocrine differentiation	41.5	0.000	0.111	27.5	0.070	0.063
50	Charlson: congestive heart failure	6.6	0.037	0.044	70.6	0.000	0.100
51	clinical CIS	1.5	0.216	0.021	69.9	0.000	0.141
52	Charlson: myocardial infarction	4.7	0.095	0.038	64.7	0.000	0.096
53	Charlson: metastatic disease for any other primary	5.2	0.073	0.040	57.4	0.000	0.091
54	Charlson: peripheral vascular disease	2.3	0.314	0.026	57.2	0.000	0.090
55	Charlson: cerebrovascular disease	1.9	0.390	0.024	57.0	0.000	0.090
56	Charlson: diabetes with end-organ damage	1.9	0.393	0.024	54.4	0.000	0.088
57	Charlson: leukemia	1.6	0.453	0.022	53.4	0.000	0.087
58	Charlson: dementia	1.5	0.483	0.021	53.2	0.000	0.087
59	Charlson: hemiplegia	3.0	0.226	0.030	52.5	0.000	0.087
60	Charlson: liver disease, moderate or severe	5.9	0.052	0.042	51.4	0.000	0.086
61	Charlson: liver disease mild	4.0	0.134	0.035	51.5	0.000	0.086
62	Charlson: lymphoma	1.4	0.494	0.021	50.9	0.000	0.085
63	Charlson: ulcer disease	3.1	0.213	0.030	50.3	0.000	0.085
64	any prior pelvic radiation	7.7	0.006	0.048	40.9	0.000	0.108
65	history of smoking	12.3	0.006	0.061	37.4	0.088	0.060
66	intravesicle BCG treatment	0.2	0.655	0.008	39.1	0.000	0.106
67	neoadjuvant bladder radiation	9.3	0.002	0.053	33.2	0.000	0.097
68	clinical hydronephrosis	0.4	0.802	0.011	31.4	0.026	0.067
69	adjuvant radiation post cystectomy	2.5	0.115	0.027	29.0	0.000	0.091
70	pathologic micropapillary	4.0	0.046	0.034	21.9	0.009	0.079
71	any type of intravesicle treatment	0.1	0.746	0.006	20.6	0.015	0.077
72	gender	9.9	0.002	0.054	9.5	0.390	0.052

Table 16: Chi-squared test, ranked by Eq.1-3. Predictors 37-72.

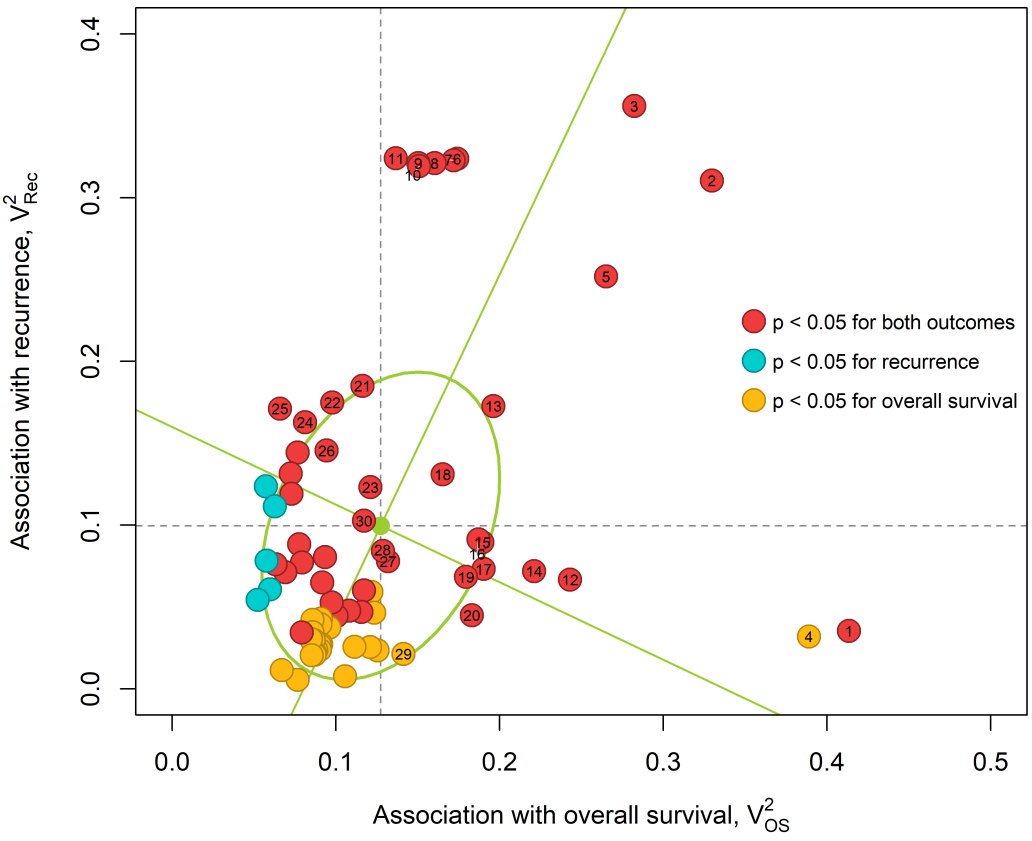


Figure 36: Predictors ranked by Cramer's V.

		Recurrence			**Overall survival**		
Rank	Predictor	χ^2	p-value	V	χ^2	p-value	V
1	type of urinary diversion constructed	4.2	0.041	0.035	598.0	0.000	0.413
2	pathologic positive lymph nodes	322.0	0.000	0.310	380.0	0.000	0.330
3	pathologic stage subgroup (OC, EV, N+)	423.1	0.000	0.356	558.0	0.000	0.282
4	current urinary diversion performed	3.4	0.064	0.032	530.0	0.000	0.389
5	pathologic lymphovascular invasion	211.9	0.000	0.252	246.0	0.000	0.265
6	pN stage TNM 7th Edition	350.0	0.000	0.324	425.0	0.000	0.174
7	pN stage TNM 5th Edition	347.9	0.000	0.323	414.0	0.000	0.172
8	pM stage TNM 7th Edition	344.2	0.000	0.321	450.0	0.000	0.160
9	pT stage TNM 7th Edition	344.3	0.000	0.321	713.0	0.000	0.150
10	pT stage TNM 5th Edition	340.3	0.000	0.319	717.0	0.000	0.151
11	# of positive lymph nodes	350.6	0.000	0.324	392.0	0.000	0.137
12	robotic surgical procedure	14.9	0.000	0.067	207.0	0.000	0.243
13	type of adjuvant chemotherapy	99.6	0.000	0.173	270.0	0.000	0.196
14	type of neoadjuvant chemotherapy	17.3	0.000	0.072	171.0	0.000	0.221
15	positive soft tissue surgical margin	26.8	0.000	0.090	251.0	0.000	0.190
16	pathologic # of tumors	27.8	0.000	0.091	490.0	0.000	0.187
17	urinary diversion subgoups	17.9	0.000	0.073	380.0	0.000	0.190
18	clincial staging subgroup (preoperative)	57.3	0.000	0.131	191.0	0.000	0.165
19	intracorporeal diversion performed	15.6	0.000	0.068	225.0	0.000	0.180
20	pM stage TNM 5th Edition	6.8	0.034	0.045	234.0	0.000	0.183
21	adjuvant chemotherapy subgroups	114.4	0.000	0.185	284.0	0.000	0.116
22	clinical T stage (preoperative)	102.2	0.000	0.175	301.0	0.000	0.098
23	clinical N Stage (preoperative)	50.7	0.000	0.123	154.0	0.000	0.121
24	pathologic carcinoma in situ	88.6	0.000	0.163	46.1	0.000	0.081
25	pathologic predominant cell type	97.8	0.000	0.171	91.2	0.001	0.066
26	# of adjuvant chemo cycles	70.8	0.000	0.146	187.0	0.000	0.094
27	clinical M stage (preoperative)	20.3	0.000	0.078	122.0	0.000	0.132
28	neoadjuvant chemotherapy subgroups	23.7	0.000	0.084	290.0	0.000	0.129
29	clinical CIS	1.5	0.216	0.021	69.9	0.000	0.141
30	type of surgical procedure	35.2	0.000	0.103	432.0	0.000	0.117
31	pathologic multifocal tumors	69.6	0.000	0.144	41.1	0.001	0.077
32	age at time of cystectomy (discretized)	11.7	0.230	0.059	467.0	0.000	0.122
33	# lymph nodes removed	7.2	0.407	0.046	374.0	0.000	0.124
34	Charlson: renal disease, moderate or severe	1.9	0.397	0.024	110.0	0.000	0.125
35	Charlson: chronic obstructive pulmonary disease	12.1	0.002	0.060	96.2	0.000	0.117
36	pathologic tumor grade	57.8	0.000	0.132	36.8	0.006	0.072

Table 17: Chi-squared test, ranked by Eq.4-6. Predictors 1-36.

Rank	Predictor	Recurrence			Overall survival		
		χ^2	p-value	V	χ^2	p-value	V
37	Charlson: connective tissue disorder	2.2	0.337	0.026	102.0	0.000	0.121
38	Charlson: diabetes	7.4	0.025	0.047	93.5	0.000	0.116
39	pathologic squamous differentiation	47.4	0.000	0.119	55.8	0.000	0.073
40	any prior pelvic radiation	7.7	0.006	0.048	40.9	0.000	0.108
41	Charlson: any other malignancy	2.2	0.333	0.026	86.8	0.000	0.111
42	# of neoadjuvant chemo cycles	21.7	0.003	0.081	213.0	0.000	0.093
43	pathologic glandular differentiation	51.0	0.000	0.124	34.4	0.156	0.057
44	pathologic neuroendocrine differentiation	41.5	0.000	0.111	27.5	0.070	0.063
45	Charlson: congestive heart failure	6.6	0.037	0.044	70.6	0.000	0.100
46	intravesicle BCG treatment	0.2	0.655	0.008	39.1	0.000	0.106
47	neoadjuvant bladder radiation	9.3	0.002	0.053	33.2	0.000	0.097
48	# of neoadjuvant radiation to bladder	14.1	0.015	0.065	147.0	0.000	0.092
49	pathologic size of bladder tumor, dim-2 (cm)	26.0	0.000	0.088	147.0	0.000	0.078
50	Charlson: myocardial infarction	4.7	0.095	0.038	64.7	0.000	0.096
51	pathologic size of bladder tumor, dim-3 (cm)	19.9	0.006	0.077	154.0	0.000	0.079
52	pathologic size of bladder tumor, dim-1 (cm)	6.2	0.290	0.043	144.0	0.000	0.091
53	Charlson: metastatic disease for any other primary	5.2	0.073	0.040	57.4	0.000	0.091
54	adjuvant radiation post cystectomy	2.5	0.115	0.027	29.0	0.000	0.091
55	Charlson: peripheral vascular disease	2.3	0.314	0.026	57.2	0.000	0.090
56	Charlson: cerebrovascular disease	1.9	0.390	0.024	57.0	0.000	0.090
57	Charlson: liver disease, moderate or severe	5.9	0.052	0.042	51.4	0.000	0.086
58	Charlson: diabetes with end-organ damage	1.9	0.393	0.024	54.4	0.000	0.088
59	Charlson: liver disease mild	4.0	0.134	0.035	51.5	0.000	0.086
60	Charlson: hemiplegia	3.0	0.226	0.030	52.5	0.000	0.087
61	Charlson: leukemia	1.6	0.453	0.022	53.4	0.000	0.087
62	Charlson: dementia	1.5	0.483	0.021	53.2	0.000	0.087
63	indication for surgery	17.0	0.005	0.071	83.6	0.000	0.069
64	Charlson: ulcer disease	3.1	0.213	0.030	50.3	0.000	0.085
65	Charlson: lymphoma	1.4	0.494	0.021	50.9	0.000	0.085
66	pathologic other variant histology	19.4	0.007	0.076	97.9	0.003	0.063
67	predominant clinical cell type (preoperative)	20.5	0.005	0.078	81.6	0.057	0.058
68	pathologic micropapillary	4.0	0.046	0.034	21.9	0.009	0.079
69	any type of intravesicle treatment	0.1	0.746	0.006	20.6	0.015	0.077
70	history of smoking	12.3	0.006	0.061	37.4	0.088	0.060
71	clinical hydronephrosis	0.4	0.802	0.011	31.4	0.026	0.067
72	gender	9.9	0.002	0.054	9.5	0.390	0.052

Table 18: Chi-squared test, ranked by Eq.4-6. Predictors 37-72.

	Ta	TIS	T1	T2a	T2b	T3a	T3b	T4a	T4b
T0	0.656	0.896	0.442	0.049	4.00E-05	0.086	3.48E-05	3.85E-05	1.06E-06
Ta	———	0.505	0.732	0.030	1.12E-07	0.104	1.03E-07	3.10E-07	9.80E-08
TIS	———	———	0.152	6.29E-04	4.27E-12	0.029	2.48E-13	1.58E-11	6.78E-10
T1	———	———	———	2.03E-04	1.44E-15	0.078	0	9.44E-15	4.98E-10
T2a	———	———	———	———	8.71E-09	0.536	3.97E-09	5.56E-08	3.98E-06
T2b	———	———	———	———	———	0.232	0.519	0.699	0.024
T3a	———	———	———	———	———	———	0.199	0.207	0.018
T3b	———	———	———	———	———	———	———	0.928	0.055
T4a	———	———	———	———	———	———	———	———	0.061

Table 19: p-value of paired log rank tests for clinical T stage subgroups.

	Pa	PIS	P1	P2a	P2b	P3a	P3b	P4a	P4b
P0	0·638	0·666	0·047	0·027	2·47E-07	0	0	0	0
Pa	———	0·791	0·357	0·243	0·005	3·45E-09	3·29E-10	0	0
PIS	———	———	0·083	0·062	1·07E-07	0	0	0	0
P1	———	———	———	0·724	1·48E-04	0	0	0	0
P2a	———	———	———	———	0·004	1·13E-14	0	0	0
P2b	———	———	———	———	———	6·59E-09	5·18E-12	0	0
P3a	———	———	———	———	———	———	0·239	3·13E-07	0
P3b	———	———	———	———	———	———	———	8·12E-06	0
P4a	———	———	———	———	———	———	———	———	2·89E-13

Table 20: p-value of paired log rank tests for pathologic T stage subgroups.

5 Feature and model based characterization of spatial and temporal organoid dynamics

5.1 Abstract

Organoid model systems recapitulate key features of mammalian tissues and enable high throughout experiments. However, the impact of these experiments is limited by manual, qualitative phenotypic analysis. We developed an automated image analysis pipeline to quantify shape dynamics in mammary organoids. Our pipeline consists of i) segmentation of image series ii) preprocessing, iii) geometrical and signal processing feature extraction, iv) dimensionality reduction to differentiate dynamical paths, v) time series clustering, and vi) dynamical modeling using point distribution models to explain shape variation. The pipeline is applied to time series from 10 untreated and 30 FGF2-treated organoids, revealing three FGF2-treated subgroups characterized by i) high-area high-circularity, ii) low-area high-circularity, and iii) high-area low-circularity. Our pipeline can characterize, cluster, and model differences among unique dynamical paths that define diverse final shapes, thus enabling quantitative analysis of the molecular basis of tissue development and disease.

5.2 Introduction

Our modern understanding of cell, developmental, and cancer biology relies heavily upon the analysis of isolated cells cultured at low density on flat, rigid substrates. These experiments are typically termed two-dimensional (2D) or conventional cell culture. These model systems have defined our understanding of subcellular compartments, single cell behavior, and the molecular regulation of cell structure and dynamics. However, 2D cell culture is inherently limited in its ability to model complex 3D tissue architecture. This problem is aggravated in mammalian systems due to the large size of the organs, the diversity of their cell types, their relatively slow development, and their relative inaccessibility deep inside the body. To overcome these challenges, a broad range of 3D culture environments and organoid techniques have been developed to provide more physiologically accurate models of tissues and organs ([172, 52]).

The term organoid, defined variously by different groups, here connotes the use of primary tissue collected directly from a mammalian organ and cultured acutely in a manner that preserves important features of the *in vivo* cell and tissue state. The large size of mammalian organs becomes an advantage in these assays, as a single normal organ or tumor can generate thousands to hundreds of thousands of organoids, each representing a sample of the heterogeneity of cell states present *in vivo*. However, conventional approaches to quantify the phenotypes of these diverse cell samples have often been limited to qualitative scoring systems,

such as binned percentages or categorical scales. This has limited the utility of organoid assays relying on manual inspection, subjective judgment, and qualitative criteria. Conversely, microscopic automation enables sequential imaging of dozens to hundreds or thousands of organoids across multiwell or microfluidic plates containing diverse experimental conditions. With abundant tissue and automated imaging, analysis quickly becomes the rate limiting step for determining the molecular regulation of these biological systems.

We sought to overcome these limitations by developing an automated imaging pipeline that can extract quantitative shape descriptor features from time series image data of organoids, analyze the dynamics of organoid shape descriptors over time, compare average differences in features between experimental conditions, and classify the phenotypic heterogeneity across experimental conditions. At the core of the pipeline are geometric and signal processing based descriptive features of organoid shape, as these are often the focus of shape assessment studies ([168, 110, 19, 152, 125, 219, 124]). We also introduce local curvature shape descriptors, and leverage all these features concurrently in order to increase robustness of downstream phenotypic analyses. As a proof of principle example, we focused on a 3D culture model of normal mouse mammary development ([56, 137]). Briefly, mouse mammary glands are mechanically and enzymatically processed to generate thousands of primary tissue organoids, each consisting of 100-300 epithelial cells ([137]). In basal culture medium, the organoids undergo growth arrest with minimal cell division. Addition of nanomolar concentrations of a growth factor, such as fibroblast growth factor 2 (FGF2), induces branching morphogenesis in a majority of organoids. We selected this model as we had qualitatively observed large differences in growth and shape between the basal and FGF2-treated organoid groups, but we were previously limited in our ability to robustly quantify and describe the diversity of the phenotypes we observed. In this study, we apply our novel imaging pipeline to characterize the developmental trajectory of organoids in both basal and FGF2-treated groups. We quantified the anticipated growth induction by FGF2 and quantitatively characterized the heterogeneity in growth and branching responses to FGF2. We anticipate that our pipeline will enable quantitative analysis of the molecular and mechanical regulators of tissue growth and invasion in a broad range of 3D culture and organoid assays.

5.3 Methods

Mammary organoids were prepared and embedded in a 3D matrix composed of equal parts Collagen-I and Matrigel according to published protocols ([56, 137]). Organoids were cultured either in basal media (DMEM-F12 with 1% insulin-transferrin-selenium and 1% penicillin-streptomycin) or in media supplemented with 2.5nM FGF2. The organoids were maintained at 37oC and 5% CO2 for 130 hours and imaged every 30 minutes using Differential Interference Contrast (DIC) microcopy for a total of 261 time points, creating a

time series of contours of each organoid's boundary. The following sections describe each step of the analysis pipeline shown in Fig.37.

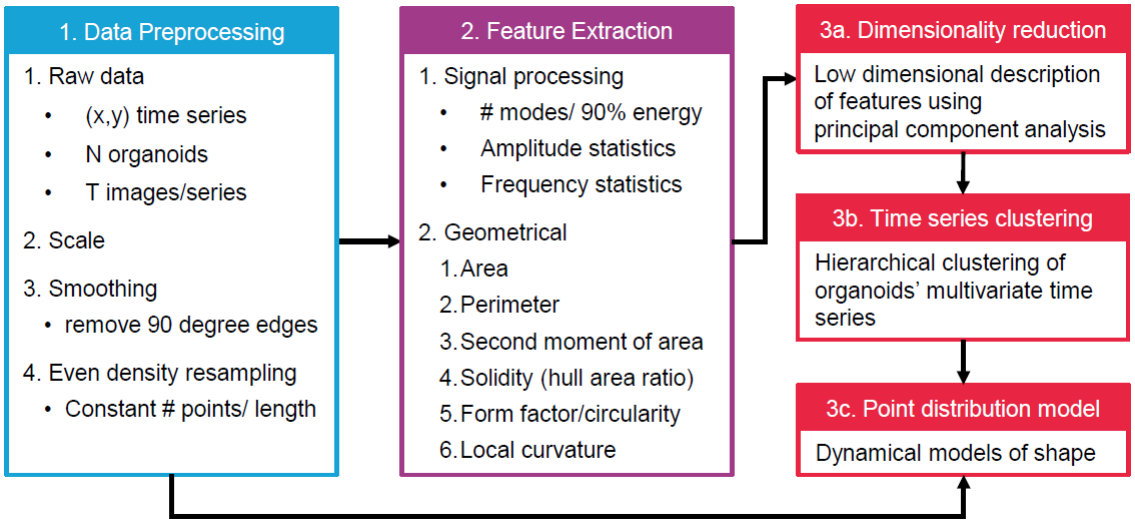


Figure 37: **Schematic highlighting the main steps in the analytical tool pipeline.** The three major steps in the pipeline are: (i) data preprocessing, (ii) feature extraction, and (iii) time series analyses.

5.3.1 Preprocessing

Each organoid image is converted to a two dimensional contour using the Automated Contour Macro (ACM) to get contours $(\vec{x}, \vec{y}) \in \mathbb{R}^{N\times 2}$ consisting of N points. The automated contour detection method produces contours composed of horizontal and vertical line segments. In the third preprocessing step, these contours are smoothed by replacing the original set of vertices by midpoints of each original line segment (S1 Appendix). In addition to smoothing the boundary, this step also avoids inflating the area and perimeter calculations. Finally, each smoothed contour of perimeter P is redrawn using a constant point density $\rho = N/P = 6\text{px}/\mu\text{m}$ by increasing the number of points in the contour N while maintaining an arc length of $1/\rho$ between points. Fig. 38 shows examples of preprocessed untreated and FGF2-treated organoid contours.

5.3.2 Feature extraction

Ten features are extracted from each contour (Table 21), where area A and perimeter P are the two basic descriptors of organoid size (S2 Appendix).

Form factor a_f and solidity a_h are two commonly used and useful area ratios. The form factor a_f is defined as the ratio of the organoid area and the area of a circle whose perimeter is equivalent to the

organoid's perimeter,

$$a_f = \frac{A}{P^2/4\pi} = \frac{4\pi A}{P^2}. \tag{62}$$

The solidity a_h is defined as the ratio of the area of the organoid and the area of the convex hull for the contour,

$$a_h = \frac{A}{A_{\text{Hull}}} \tag{63}$$

Note that a_f reaches a maximum of 1 for circles, and both a_f and a_h measure the area-packing efficiency, or complexity, of a curve (for which $a_f < 1$ and $a_h < 1$).

Another measure of contour irregularity is the polar moment of area J_{zz} (S2 Appendix). While area and perimeter directly describe the size of an organoid, the area ratios capture the increasing complexity of contours due to localized protrusions and can identify organoids that do not maintain shape over time. The polar moment of area J_{zz} furthers this concept by calculating explicitly the squared distances of each infinitesimal unit of area from the normal axis at the contour's centroid and describes the distribution of an organoid's area.

A greater granularity in contour description is achieved by studying the local curvature along the contour, with the fraction of convex f_{vex} and collinear f_{col} points in a contour serving as summary statistic features (S2 Appendix). The fraction of concave points f_{cav} is omitted to maintain linear independence in the feature set.

A final set of features is derived from the discrete Fourier transform (DFT) of a contour (S2 Appendix). We find the set of modes with the highest amplitude that comprise 90% of the total energy (sum of all mode amplitudes) of a signal, E_T, to calculate i) the number of modes N_{90}, ii) the mean mode amplitude $\bar{A}_{90}$, and iii) the standard deviation σ_{90} of mode amplitudes in the 90% E_T signal. The 90% E_T signal attenuation serves to omit higher frequency modes that are more likely due to contour segmentation artifacts in the first preprocessing step.

The p features (Tab.21) are extracted from an organoid's time series of contours, and together they represent a multivariate time series $X^o \in \mathbb{R}^{T\times p}$ of length $T = 261$ and dimensionality p for an organoid o. These multivariate time series are compared directly for the clustering task (Fig.373b). The time series are concatenated for the low dimensional representation (Fig.373a), where the entire set of p features for $n_O = 40$ organoids are stacked in a data matrix $\mathbf{X} \in \mathbb{R}^{n_O T\times p}$. Each column represents a single feature for all $n_O T$ organoid contours in the dataset, and each row is a single time point or contour of an organoid.

Extracted features		
1	A	area
2	P	perimeter
3	a_f	form factor
4	a_h	solidity
5	J_{zz}	polar moment of area
6	f_{vex}	fraction of convex points
7	f_{col}	fraction of collinear points
8	N_{90}	number of modes in $90\%E_T$
9	$\bar{A}_{90}$	mean mode amplitude in $90\%E_T$
10	σ_{90}	standard deviation of mode amplitude in $90\%E_T$

Table 21: List of quantitative features extracted from organoid contours.

5.3.3 Low dimensional representation

Even though the feature space of the data matrix $\mathbf{X}$ is small compared to the number of sample points, principal component analysis (PCA) is performed in order to chart the composite evolution of all the geometrical and DFT properties of the entire set of organoids. A reduced data matrix

$$\mathbf{Y}_r = \mathbf{XV} \in \mathbb{R}^{{n_O}^T \times p_r} \tag{64}$$

$$= \mathbf{USV}^T\mathbf{V} \tag{65}$$

$$= \mathbf{US} \tag{66}$$

of dimension p_r is calculated using singular value decomposition of the original matrix $\mathbf{X} = \mathbf{USV}^T$, where $\mathbf{V}$ is the matrix of principal components ([103]). To obtain a reduced data matrix with $p_r < p$, the number of columns of $\mathbf{V}$ are reduced to p_r. The features A, P, J_{zz}, and $\bar{A}_{90}$ are log-transformed before the PCA in order to linearize relationships involving these features.

5.3.4 Time series clustering

The total distance between two organoids i and j,

$$\mathbf{D}(Y^i, Y^j) = \frac{1}{p_r}\sum_{k=1}^{p_r} g\left(Y_k^i, Y_k^j\right) \tag{67}$$

is the mean distance over all p_r principal components as measured by the distance metric g. The distance matrix is symmetric, $\mathbf{D} \in \mathbb{R}^{n_O \times n_O}$, where $n_O = 40$ is the total number of organoids in the dataset. The principal components are not standardized prior to applying the distance function g in order to maintain the variance they describe for the clustering algorithm. This distance matrix is used to cluster organoids, and the choice of p_r is based on the variance explained in the original data matrix as well as the quality of the

resulting clusters. Here, organoids are grouped using a hierarchical agglomerative average-linkage clustering algorithm based on the cosine similarity measure (S3 Appendix). Although the cosine similarity measure is not a true metric as it does not satisfy the triangle inequality, it is among the commonly used notions of distance and similarity like the L_p Minkowski and entropy based measures ([31]).

Point distribution model

Statistical models of variation of an individual organoid's spatio-temporal changes are created using the point distribution model (PDM) [36, 37] whose outputs are principal modes of variation about the mean organoid shape from a sequence of contours. Contour points are stacked in a column vector for M contours of length N,

$$(x_i, y_i),\ i \in [1, N] \tag{68}$$

$$\vec{q}_k = \begin{bmatrix} x_1 \\ \vdots \\ x_N \\ y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^{2N},\ k \in [1, M]. \tag{69}$$

The mean shape organoid contour over all M samples,

$$\underline{\vec{q}} = \frac{1}{M} \sum_{k}^{M} \vec{q}_k, \tag{70}$$

is then used to calculate the covariance matrix,

$$\mathbf{S} = \frac{1}{M-1} \sum_{k}^{M} (\vec{q}_k - \underline{\vec{q}})(\vec{q}_k - \underline{\vec{q}})^T. \tag{71}$$

The eigendecomposition of $\mathbf{S}$,

$$(\lambda_i, \vec{u}_i),\ i \in [1, 2N] \tag{72}$$

reveals the principal modes of variation and the fraction of temporal variance described by a single mode,

$$\text{Fraction of variance } \vec{u}_i = \frac{\lambda_i}{\sum_i \lambda_i}. \tag{73}$$

More complex evolution dynamics require a larger number of modes to capture the temporal shape variance therefore we use the fraction of variance explained by first eigenmode, and the area under the variance versus number of eigenmodes curve as metrics of dynamical complexity. A statistically viable shape,

$$\vec{q}_{sim} = \underline{\vec{q}} + \sum_{i=1}^{K} b_i \vec{u}_i, \;\; K << 2N \tag{74}$$

is simulated using the K eigenmodes which modulate the mean shape $\vec{q}_{sim}$ by scalar factors b. The K principal modes of variation u_i are ranked in decreasing order of eigenvalues λ_i. Simulated shapes generated by combinations of K eigenmodes $\vec{u}_i$ and scalar factors b_i describe the dynamical evolution of an organoid's shape. The choice of b_i is guided by the range observed in the dataset. It is typically bounded within a range of three standard deviations [36],

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i} \tag{75}$$

because the eigenvalues λ_i describe the variance of b_i in the dataset. Alternatively, the range of mode coefficients can be calculated directly from the dataset,

$$\vec{b} = \mathbf{U}^T(\vec{q}_k - \underline{\vec{q}}) \tag{76}$$

for $k = 1, .., M$ contours, where the matrix $\mathbf{U}$ has K columns, each being one of the K eigenmodes.

5.4 Results

5.4.1 Feature-based description

The time series of extracted features (Table 21) for the untreated (green) and FGF2-treated (purple) organoids are depicted in Fig. 39. Fitted exponential curves $A_{untreated} = 1843e^{-0.00305t}$ and $A_{FGF2} = 1507e^{0.0106t}$ in Fig. 39A, and $P_{untreated} = 181e^{-0.00109t}$ and $P_{FGF2} = 153e^{0.00636t}$ in Fig. 39B show a substantially greater increase in area and perimeter for FGF2 organoids compared to basal types. The area ratio time series are noisy and do not show a significant separation between the two phenotypes. The fitted

curves have similar parameters in Fig. 39C,

$$a_{f,\text{untreated}} = 0.709e^{-0.000869t} \tag{77}$$

$$a_{f,\text{FGF2}} = 0.803e^{-0.00212t}, \tag{78}$$

and in Fig. 39D

$$a_{h,untreated} = 0.903e^{-0.000460t} \tag{79}$$

$$a_{h,FGF2} = 0.936e^{-0.000569t}. \tag{80}$$

Nevertheless, the decreases in a_f and a_h reflect the transformation from generally elliptical contours to irregular shapes at later times in the organoid evolution.

In contrast to the area ratios, second moment of area in Fig. 39E shows the greatest temporal separation between untreated and treated organoids, with corresponding curve fits $J_{zz,untreated} = 584500e^{-0.00598t}$ and $J_{zz,FGF2} = 373000e^{0.0223t}$. The observed separation occurs because the area of the FGF2-treated organoids is distributed further away from the perpendicular centroid axis over time, while the untreated organoids' areas do not change with time as they maintain their original elliptical contours.

The local curvature features in Figs. 39F and 39G also reveal an interesting divergence in the shape changes of the two treatment groups. The curve fits for convexity in Fig. 39F, $f_{\text{vex,untreated}} = 0.257e^{-0.000742t}$ and $f_{vex,FGF2} = 0.265e^{0.000500t}$ show a slight increase in convexity from FGF2 treatment and a greater decrease in untreated organoids over time. The opposite trend is seen in the collinearity curve fits $f_{col,untreated} = 0.430e^{0.000431t}$ and $f_{col,FGF2} = 0.427e^{-0.000528t}$ in Fig. 39G. The fits for concavity are $f_{cav,untreated} = 0.311e^{-0.0000518t}$ and $f_{cav,FGF2} = 0.305e^{0.000308t}$. Therefore, for untreated organoids, increase in local collinearity over time is matched by a slight decrease in concavity and a larger decrease in convexity. In contrast, FGF2-treated organoids become less collinear and more convex and concave over time as they grow complex features.

The DFT features are shown in Fig. 39H-J. The mean N_{90} is 41.1 and 46.2 for untreated and treated organoids respectively, and the trend for N_{90} remains linear in time. Mean amplitude (Fig. 39I) shows an exponential increase for FGF2 types, where curve fits are $\bar{A}_{90,untreated} = 463e^{-0.00310t}$ and $\bar{A}_{90,FGF2} = 217e^{0.01486t}$. The standard deviation of the mode amplitudes (Fig.39J) has a similar trend, $\sigma_{90,untreated} = 699e^{-0.00347t}$ and $\sigma_{90,FGF2} = 369e^{0.01464t}$. The exponential increase in $\bar{A}_{90}$ for FGF2-treated organoids is consistent with their size increases seen in Figs. 39A and 39B.

The changing relationship between area and perimeter over time (Fig. 40) reveals the departure from

circularity that some organoids undergo. The untreated organoids remain in a low-area and low-perimeter region in Fig. 40, except for one organoid that behaves like the FGF2-treated organoids and is seen as an outlier in Fig. 39 as well. In contrast, the FGF2 organoids deviate further from the circular limit of $A = P^2/4\pi$ (solid black line, Fig. 40).

Similarly, the temporal changes of the organoid shape projected into the composite principal component space (Fig. 41) of the whole feature set shows a clear distinction in the shape changes of untreated and treated organoids. As seen in Fig. 41A, the first three principal components capture 92.4% of the variance of the entire feature set; the inset plot shows the feature weights of the first two principal components. The feature σ_{90} was removed from the dataset for the PCA due to strong linear correlation with $\bar{A}_{90}$. While other dependencies are present in the feature set, this choice of features results in a desirable clustering which separates FGF2 and untreated organoids and identifies FGF2 subtypes in the next section.

Plots of the first two principal components and convex hulls of the entire data sets, Fig. 41B, show that the FGF2 types differ from the range of basal contours. The untreated and FGF2-treated organoids projected onto principal components are shown in Figs. 41C and D, colored to specify early (t_1), mid (t_2), and late (t_3) time periods. The 30 treated organoids occupy a larger region than the 9 basal types in the projected space because the mid and late period FGF2-treated organoids move away from the initial region, while the untreated organoids remain close to their initial region. Although the first two principal components explain 80.9% of the variance in the organoid shapes, Figs. 41B-D illustrate that the untreated organoids do not change projected location substantially over time. In contrast, the FGF2 treatment may lead to distinct patterns of shape changes because the t_3 data points do not coalesce to a single region of the principal component space. We turn to clustering to explore general shape characteristics that distinguish the separate groups.

5.4.2 Clustering

Hierarchical clustering of the organoids based on $p_r = 5$ principal components is shown in Fig. 42. The first bifurcation in the clustering (Fig. 42A) is between untreated and FGF2-treated organoids: cluster #1 (green) includes all but one of the untreated organoids, and clusters #2-4 (yellow, blue, and red respectively) are entirely FGF2-treated, with the exception of organoid #8 which is an outlier in the set of untreated. The overlapped first and last contour drawings in Fig. 42A show that untreated organoids' areas decreases with time, except for #7 whose area remains nearly constant. Although the FGF2-treated organoids start in the same region of the principal component space as the untreated organoids (Fig. 41B-D), they evolve into three clusters which occupy distinct regions of the principal component space (Fig. 42B) and are characterized by unique feature combinations shown by the distributions of features during t_3 in Fig. 42C-F. All three

FGF2 clusters have higher perimeter and area than the untreated cluster. However, there is a gradation in this effect of treatment: clusters #3 and #4 have the highest area perimeter (Fig. 42C,D). The primary distinction between the large organoid clusters #3 and #4 is that cluster #4 organoids are relatively more circular as seen by the low a_f numbers in Fig. 42F. Cluster #2 includes the smallest and most circular of the FGF2-treated organoids. Cluster #3 organoids are large and most irregularly shaped organoids. Cluster #4 organoids are large yet remain circular. Differences in local curvature also stratify organoid clusters. The untreated organoids tend to have the lowest f_{vex}, Fig. 42E. Consequently, the differences in clusters #2-4 reveal the distinct phenotypic outcome of the FGF2 treatment.

5.4.3 Dynamical model

A first-mode simulation of a single organoid from each of the four clusters in Fig. 42 is shown over four time intervals of the evolution in Fig. 43. The first mode's coefficient b_1 is taken as the median of the value from the organoid's contours during the corresponding time period.

In Fig. 43A, organoid 7 from cluster #1 undergoes small fluctuations but a negligible change in overall size. On the other hand, organoid 34 in Fig. 43B maintains its general shape and expands in time. A larger increase in size is observed for both organoids 29 and 40 in Fig. 43C and D respectively. However, organoid 29 elongates primarily along one axis, whereas organoid 40 expands at each point of convexity leading to a slightly greater circularity than organoid 29.

Next, we generated and compared PDMs from each organoid's time series of contours, then examined the temporal variance described by the PDM modes. Generally, contours that undergo a more complex evolution require a greater number of modes K to fully describe the correspondingly large temporal variance. The fraction of variance explained by the first mode in each organoid's PDM is compared in Fig. 42G for the four detected organoid clusters. The first modes of cluster #3 organoids, which have the least regular boundaries, describe the least variance in their respective organoids' evolution. Aside from a few outliers, Fig. 42G provides evidence that cluster #2 organoids have simpler dynamics because this cluster has the highest fraction of the total variance explained by the first mode. This observation is consistent with the previous visual inspection of cluster boundaries.

The cumulative fraction of variance explained by PDM modes $\vec{u_i}$ reaches saturation near five modes for all organoids. The area under the cumulative variance curve (AUC) also describes the complexity of the contour dynamics because a higher variance in a series of contours will have a smaller AUC as more modes are needed to capture all sources of temporal variance. The AUC for $K = 5$ modes is shown in Fig. 42H for the four detected clusters. Clusters #2 and #4 have higher AUCs as these clusters' organoids remain consistently circular throughout the evolution. Cluster #1 has a smaller AUC due to smaller fluctuations in

shape; however, cluster #3 has the lowest AUC as these organoids transform significantly over time.

5.5 Discussion

We developed a computational pipeline that provides an automated platform for studying the spatial and temporal dynamics of shape in complex 3D organoids. For simplicity in this first implementation, we restricted our analyses to 2D projections of 3D volumes and found that this was sufficient to detect average differences in organoid shape between conditions and to characterize the heterogeneity in response of individual organoids within the same experimental condition. Since our organoid assays generate 1000s of genotypically equivalent replicates, this analytical framework was also useful for identifying precisely when and how the experimental organoids diverged in their shape trajectories from control organoids.

Furthermore, the set of geometric and discrete Fourier transform features together provide phenotypic signatures of organoid morphological changes. The low-dimension representation of the extracted feature set summarizes the divergent paths taken by FGF2-treated organoids in comparison to untreated organoids. Particularly, the treated organoids span a larger region of principal component space due to their complex shaped contours at later times. Point distribution based on sequences of contours succinctly describe the shape dynamics with three or fewer modes for most organoids and are useful for ranking clusters of organoids by their dynamical complexity. Hierarchical clustering of organoids in the reduced feature set not only separates untreated and FGF2-treated organoids but reveals three phenotypic subgroups in the FGF2-treated set: i) low-area high-circularity, ii) high-area high-circularity, and iii) high-area low-circularity.

Biologically, these sub-clusters suggest that some organoids do not respond morphologically to FGF2, others grow in response to FGF2 but do not branch, and others both grow and branch. Since a cell must express an appropriate FGF receptor in order to recognize and respond to FGF2, an possible explanation is that FGF receptor expression is low or absent in the non-responding organoids. Similarly, although the genetic state is identical across organoids, differences in cell state defined by transcription or by epigenetics may be responsible for differences in FGF2-response.

Thus, the ability to characterize the growth and branching response of organoids to FGF2 enables future experiments to characterize the molecular basis of their differential response. The underlying basic biology is relevant to branching morphogenesis in normal development and potentially to aberrant cell behavior in cancer. Though a simple proof-of-principle example, the discovery of patterns of organoid behavior highlights the power of automated image analysis to identify biologically cohesive phenotypes in large populations in an unbiased fashion, or at least in a fashion in which the phenotypic boundaries are clearly defined. The distinct phenotypic groups can then be recovered for molecular analysis. We anticipate that this approach

will find broad utility in determining the molecular regulators of cell structure and function within epithelial tissues and tumors.

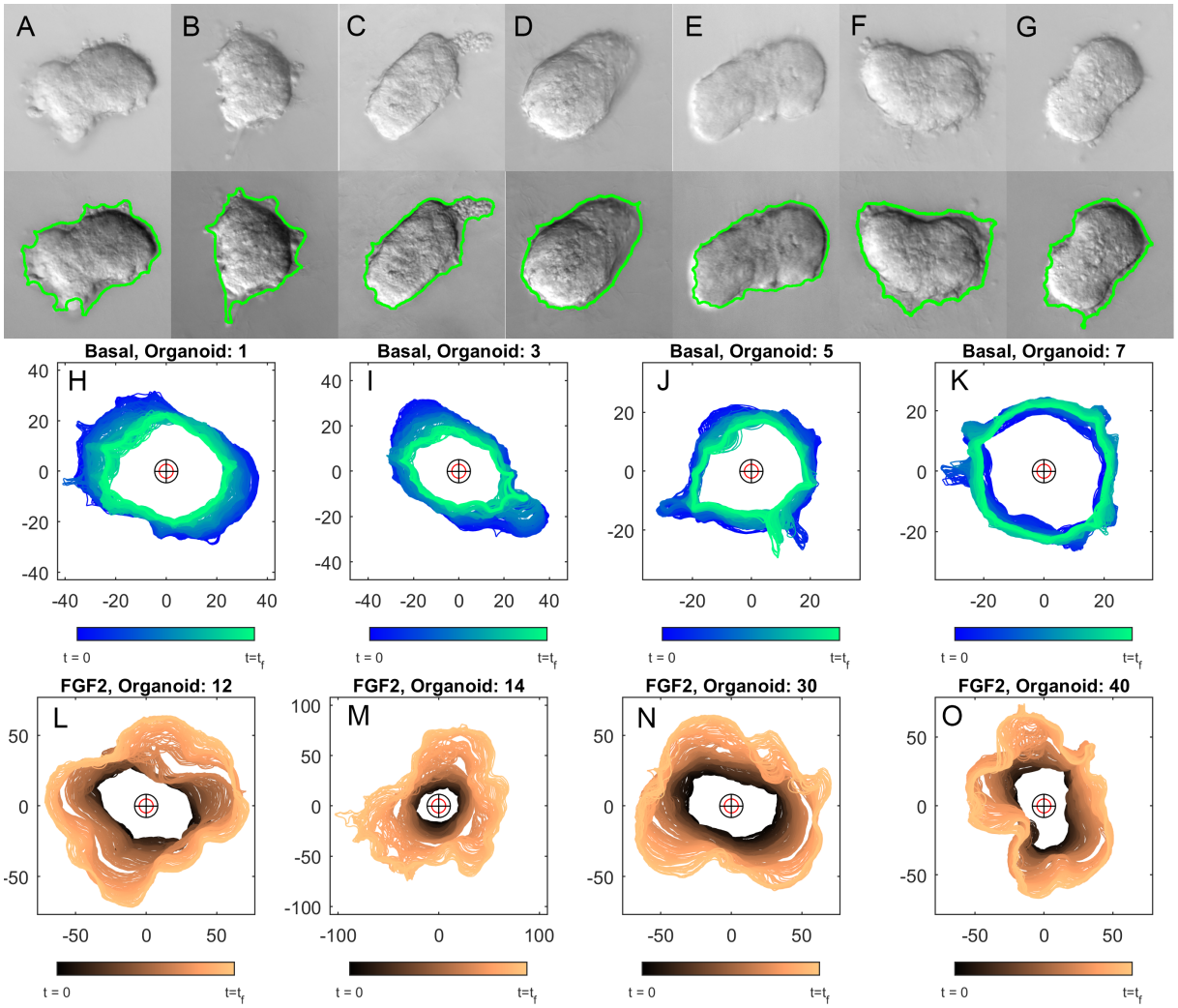


Figure 38: **Examples of organoid images, segmentation, and observed spatio-temporal changes. A-G** Original organoid image (top) and segmented boundary (bottom, green). **A-C** are untreated basal organoids, and **D-G** are FGF2 treated organoids at initial time points. **A-D** untreated and **E-H** FGF2-treated organoids' series of observed contours color coded by time point and overlapped to show the spatio-temporal changes from $t = 0$ to $t_f = 130$ hours.

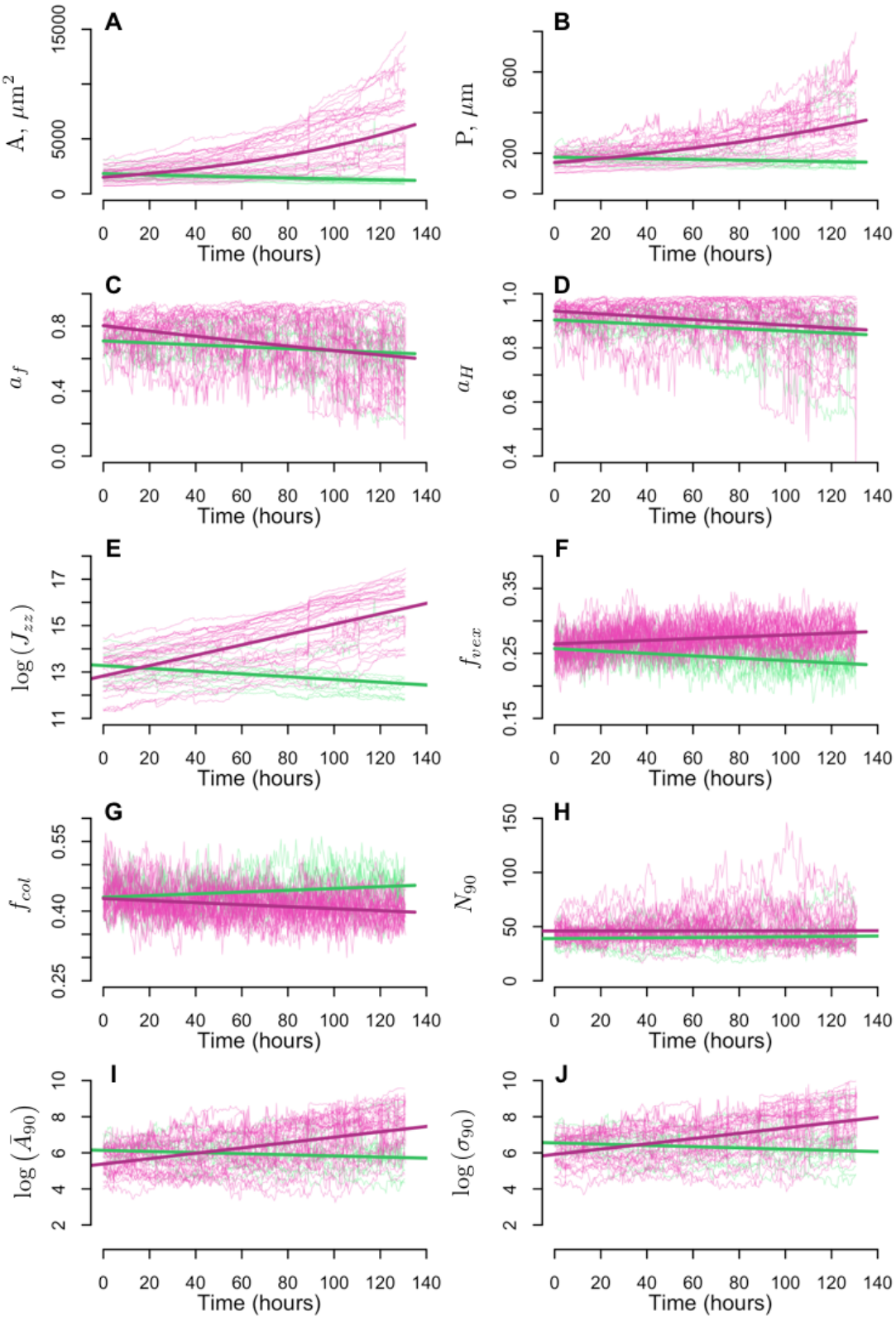


Figure 39: **Extracted feature time series for all organoids.** Basal: green; FGF2: purple. **A**, area; **B**, perimeter; **C**, form factor; **D**, solidity; **E**, log of second moment of area; **F**, fraction of convex points; **G**, fraction of collinear points; **H**, number of modes in the 90% E_T; **I**, log of mean mode amplitude in the 90% E_T; **J**, log of mode amplitude standard deviation in the 90% E_T.

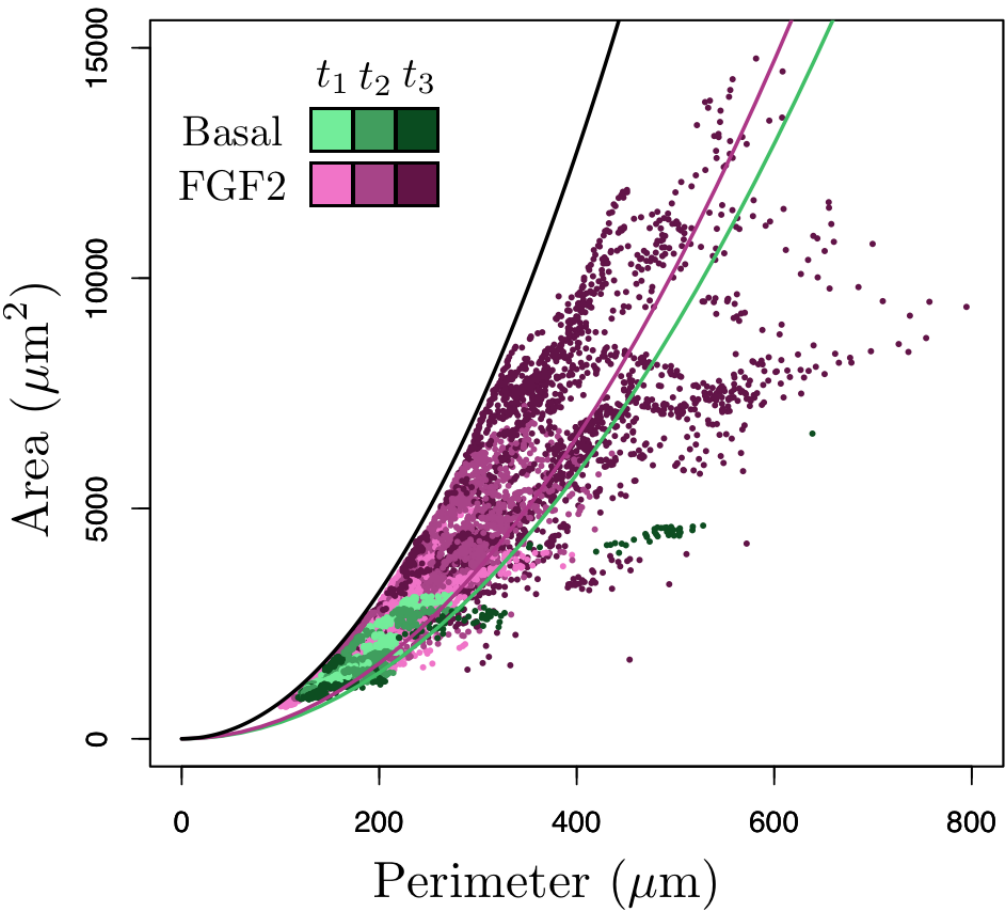


Figure 40: **Relationship between area and perimeter over time.** Temporal data is shown by a three tier encoding: $0 \leq t_1 < 43.5$ hr $\leq t_2 < 87.5$ hr $\leq t_3 \leq 130.5$ hr. The area-perimeter relationship for a circle, $A_{\text{circle}} = P^2/4\pi$ is shown by the black curve. Quadratic curve fits for untreated organoids, $A_{\text{untreated}} \approx 0.0360P^2$, and FGF2-treated organoids, $A_{\text{FGF2}} \approx 0.0410P^2$ are shown in green and purple respectively.

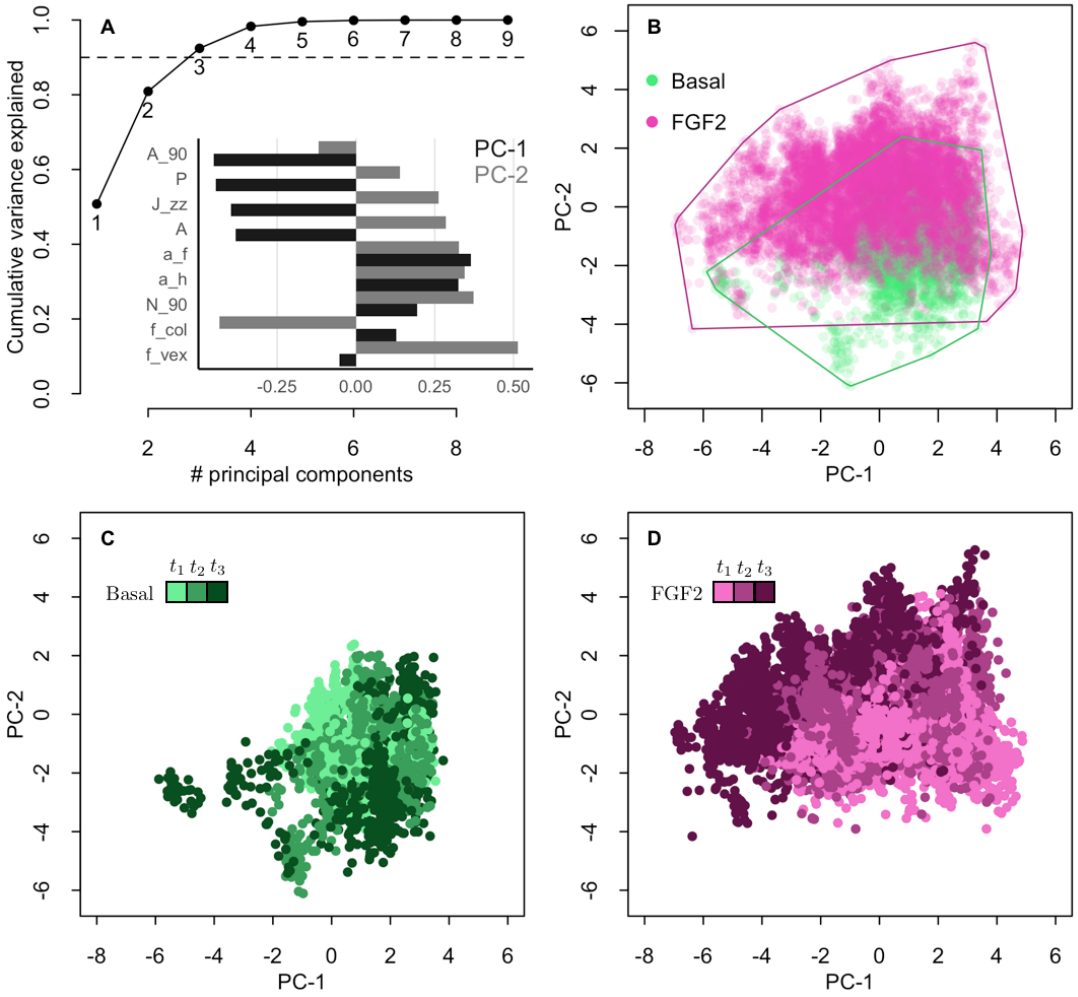


Figure 41: **Organoid shape dynamics in reduced principal component space. A**, fraction of variance explained by principal components in the original dataset and the feature weights for the first two principal components. **B**, all FGF2-treated and untreated organoid data points plotted in subspace of first two principal components, the solid purple and green lines are convex hulls of all the FGF2-treated and untreated organoids respectively. **B**, **C**, **D**, organoid data points plotted in subspace of first two principal components for both FGF2-treated and untreated, only untreated, and only treated respectively. Temporal data in **C** and **D** is shown by a three tier encoding: $0 \leq t_1 < 43.5$ hr $\leq t_2 < 87.5$ hr $\leq t_3 \leq 130.5$ hr.

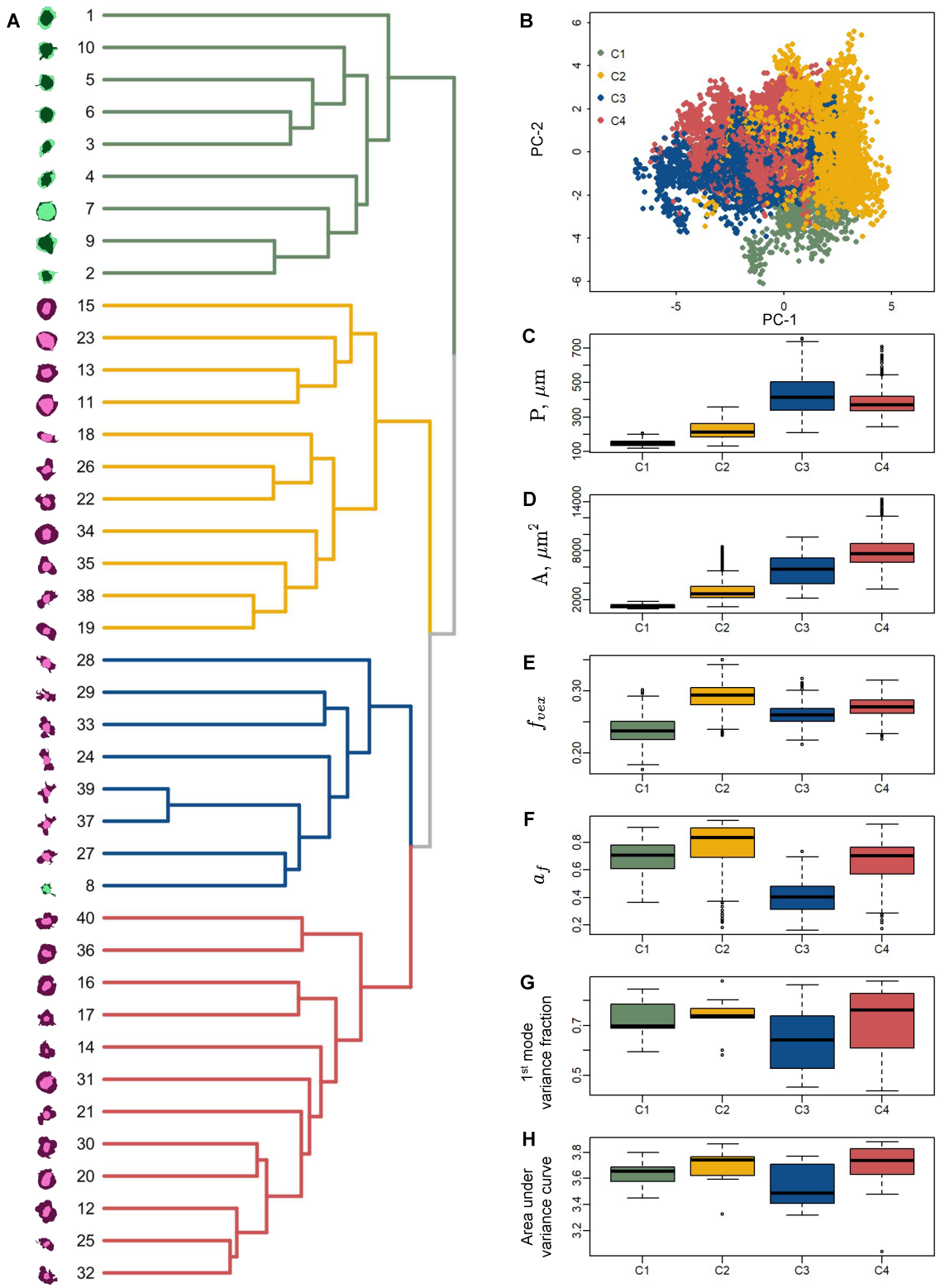


Figure 42: **Hierarchical clustering of organoids. A** Organoids grouped into four clusters (1: green, 2: yellow, 3: blue, 4: red) with each organoid's first (light shade) and last (dark shade) contour. Untreated organoids are shown in green and FGF2-treated organoids are shown in purple. **B** All data points are colored by clusters in the space of the first two principal components. **C,D,E,F** Boxplots of perimeter, area, fraction of convex points, and form factor for the four clusters during the late regime $t_3 > 87.5$ hr. **G** Fraction of temporal variance in time series of contours explained by first mode in the point distribution models. **H**

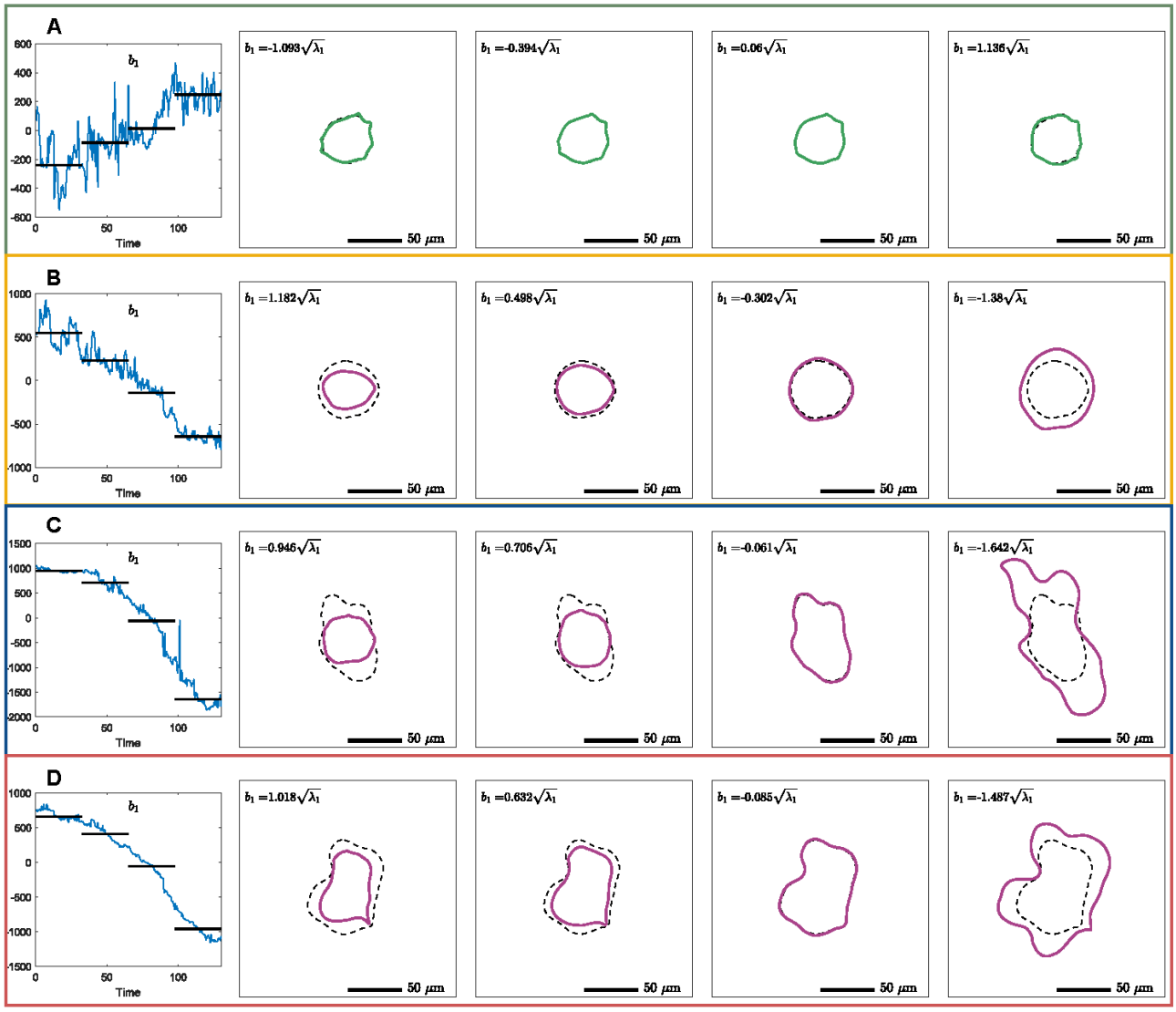


Figure 43: **Point distribution models of shape variation.** Far left subplots show first mode coefficient b_1 over time (blue), and median values (black) over four time quarters ($0 \leq t_1 < 32.6$ hr $\leq t_2 < 65.3$ hr $\leq t_3 < 97.9$ hr $\leq t_4 \leq 130.5$ hr). **A**,**B**,**C**, and **D** show contour simulations $\vec{q}_{sim}$ for each time quarter using only the first mode in the PDM for organoids 7, 34, 29, 40 respectively. Mean contours are shown by a black dashed line.

5.6 Appendix

5.6.1 S1 Appendix

Preprocessing algorithms.

Algorithm 1: Contour scaling and mid-point smoothing

Data: Contour $(\vec{x}, \vec{y}) \in \mathbf{R}^{N\times 2}$
Result: Smoothed and rescaled contour $(\vec{x}_{smooth}, \vec{y}_{smooth})$
scaling factor c;
$\vec{x} = c\vec{x}$;
$\vec{y} = c\vec{y}$;
for $i = 1 : N-1$ **do**
  $\vec{x}_{smooth}(i) = (\vec{x}(i) + \vec{x}(i+1))/2$;
  $\vec{y}_{smooth}(i) = (\vec{y}(i) + \vec{y}(i+1))/2$;
end
$\vec{x}_{smooth}(N) = (\vec{x}(N) + \vec{x}(1))/2$;
$\vec{y}_{smooth}(N) = (\vec{y}(N) + \vec{y}(1))/2$;

5.6.2 S2 Appendix

Derivation and description of extracted organoid contour features are given here. Area is calculated by applying Green's theorem

$$A = \iint dxdy = \oint_C xdy = -\oint_C ydx \tag{81}$$

$$\Rightarrow \oint_C xdy = \sum_i \int_{C_i} xdy = \sum_i A_{C_i} \tag{82}$$

$$A_{C_i} = \begin{cases} \int_{x_i}^{x_{i+1}} mxdx = \frac{m}{2}\left(x_{i+1}^2 - x_i^2\right) & \text{if } x_i \neq x_{i+1}, y_i \neq y_{i+1} \\ \int_{y_i}^{y_{i+1}} x_c dy = x_c\left(y_{i+1} - y_i\right) & \text{if } x_i = x_{i+1} = x_c \\ \int_{y_i}^{y_{i+1}} xdy = 0 & \text{if } y_i = y_{i+1} = y_c \end{cases} \tag{83}$$

Algorithm 2: Resample contour with constant density

Data: Contour $(\vec{x}, \vec{y}) \in \mathbf{R}^{N\times 2}$
Result: Contour $(\vec{x}', \vec{y}') \in \mathbf{R}^{N'\times 2}$ with point density ρ
Residual length $l_r = 0$;
New contour length $n^* = 1$;
for $i = 1 : N - 1$ **do**
  Direction from point i to $i + 1$;
  $u_x = \vec{x}(i + 1) - \vec{x}(i)$;
  $u_y = \vec{x}(i + 1) - \vec{x}(i)$;
  $u_x = u_x/\sqrt{u_x^2 + u_y^2}$;
  $u_y = u_y/\sqrt{u_x^2 + u_y^2}$;
  Effective length;
  $l_e = \sqrt{U_x^2 + u_y^2} + l_r$;
  Number of points to add;
  $n_{add} = \lfloor l_e\rho \rfloor$;
  Add first point (may be closer than $1/\rho$;
  $\vec{x}'(n^* + 1) = \vec{x}(i) + ((1/\rho) - l_r)u_x$;
  $\vec{y}'(n^* + 1) = \vec{y}(i) + ((1/\rho) - l_r)u_y$;
  $n_{add} = n_{add} - 1$;
  $n^* = n^* + 1$;
  Add remaining points;
  for $k = 1 : n_{add}$ **do**
    $\vec{x}'(n^* + 1) = \vec{x}'(n^*) + (1/\rho)u_x$;
    $\vec{y}'(n^* + 1) = \vec{y}'(n^*) + (1/\rho)u_y$;
    $n^* = n^* + 1$;
  end
  Left over distance on current segment;
  $l_r = l_e$ modulo $1/\rho$;
end

where each summand A_{C_i} of the line integral is calculated for a single line segment component $y = mx + b$. The centroid of the contour is calculated similarly,

$$\bar{x} = \frac{\iint x dx dy}{\iint dx dy} = \frac{1}{2A} \oint_C x^2 dy \tag{84}$$

$$\Rightarrow \frac{1}{A} \oint_C \frac{x^2}{2} dy = \frac{1}{A} \sum_i \int_{C_i} \frac{x^2}{2} dy = \frac{1}{A} \sum_i \bar{x}_i \tag{85}$$

$$\bar{x}_i = \begin{cases} \displaystyle\int_{x_i}^{x_{i+1}} \frac{1}{2} m x^2 dx = \frac{m}{6} \left(x_{i+1}^3 - x_i^3 \right) & \text{if } x_i \neq x_{i+1}, y_i \neq y_{i+1} \\ \displaystyle\int_{y_i}^{y_{i+1}} \frac{1}{2} x_c^2 dy = \frac{x_c^2}{2} \left(y_{i+1} - y_i \right) & \text{if } x_i = x_{i+1} = x_c \\ \displaystyle\int_{y_i}^{y_{i+1}} \frac{1}{2} x^2 dy = 0 & \text{if } y_i = y_{i+1} = y_c \end{cases} \tag{86}$$

$$\bar{y} = \frac{\iint y dx dy}{\iint dx dy} = -\frac{1}{2A} \oint_C y^2 dx \tag{87}$$

$$\Rightarrow \frac{1}{A} \oint_C -\frac{y^2}{2} dx = \frac{1}{A} \sum_i \int_{C_i} -\frac{y^2}{2} dx = \frac{1}{A} \sum_i \bar{y}_i \tag{88}$$

$$\bar{y}_i = \begin{cases} \displaystyle\int_{y_i}^{y_{i+1}} -\frac{y^2}{2m} dy = -\frac{1}{6m} \left(y_{i+1}^3 - y_i^3 \right) & \text{if } x_i \neq x_{i+1}, y_i \neq y_{i+1} \\ \displaystyle\int_{x_i}^{x_{i+1}} -\frac{y^2}{2} dx = 0 & \text{if } x_i = x_{i+1} = x_c \\ \displaystyle\int_{x_i}^{x_{i+1}} -\frac{y_c^2}{2} dx = -\frac{y_c^2}{2} \left(x_{i+1} - x_i \right) & \text{if } y_i = y_{i+1} = y_c \end{cases} \tag{89}$$

Perimeter is calculated by summing along the line segments:

$$P = \sum_i^N = \left((x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right)^{1/2}. \tag{90}$$

Second moment of area is calculated about both the axes,

$$I_{xx} = \iint y^2 dA, \tag{91}$$

$$I_{yy} = \iint x^2 dA, \tag{92}$$

and only the polar moment of area,

$$J_{zz} = I_{xx} + I_{yy} \tag{93}$$

is maintained in the extracted feature set.

Locally convex, collinear, and concave points are defined using the orientation matrix,

$$\mathbf{O} = \begin{bmatrix} 1 & x_{i-1} & y_{i-1} \\ 1 & x_i & y_i \\ 1 & x_{i+1} & y_{i+1} \end{bmatrix} \tag{94}$$

and its determinant

$$\det(\mathbf{O}) = (x_i - x_{i-1})(y_{i+1} - y_{i-1}) - (x_{i+1} - x_{i-1})(y_i - y_{i-1}) \tag{95}$$

for pairs of line segments connected by point i. For a clockwise oriented contour, a point i is convex if $\det(\mathbf{O}) < 0$, collinear if $\det(\mathbf{O}) = 0$, or concave if $\det(\mathbf{O}) > 0$. The conditions for convexity and concavity are reversed for a counter-clockwise oriented contour. The set of N points of a contour satisfy the relationship

$$f_{vex} + f_{col} + f_{cav} = 1 \tag{96}$$

from which the fraction of convex f_{vex} and collinear f_{col} points in a contour are used as summary statistic features, and fraction of concave points f_{cav} is ignored to maintain linear independence in the feature set.

The discrete Fourier transform (DFT) of a contour $(\vec{x}, \vec{y})$ is calculated by first casting it into the complex plane,

$$z_n = x_n + iy_n \tag{97}$$

$$n = 1, ..., N \tag{98}$$

and then applying the transform,

$$Z_k = \sum_{n=0}^{N-1} z_n e^{-i2\pi kn/N} \tag{99}$$

$$k = 0, 1, .., N-1, \tag{100}$$

which has a sampling period $T_p = 1/\rho$, frequency $f_s = \rho$, and DFT frequency domain resolution f_s/N. The

total energy of a signal, E_T, is the sum of each mode's amplitude,

$$E_T = \sum_{k=1}^{N-1} \|Z_k\|, \tag{101}$$

$$\|Z_k\| = \sqrt{\mathrm{Re}(Z_k)^2 + \mathrm{Im}(Z_k)^2}. \tag{102}$$

We sort the modes in order of descending amplitude and find the set of modes with the highest amplitude that comprise $90\%E_T$ to calculate i) the number of modes N_{90}, ii) the mean mode amplitude $\bar{A}_{90}$, and iii) the standard deviation σ_{90} of mode amplitudes in the $90\%E_T$ signal. The $90\%E_T$ signal attenuation serves to omit higher frequency modes, which are more likely due to contour segmentation in the first preprocessing step.

5.6.3 S3 Appendix

The cosine similarity measure,

$$cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{103}$$

$$g_{cosine}(A, B) = 1 - cos(\theta) \tag{104}$$

acts on two time series A and B of length n and used in the time series clustering analysis as the distance metric.

6 Temporal differentiation of phenotypes using a Hidden Markov model of organoid shape dynamics

6.1 Abstract

Phenotype differentiation is a central focus of analytical tools designed to study outcomes in organoid experiments where the goal is to understand and reveal biomechanisms by associating observable features with various organoid treatments. Feature based clustering techniques offer static differentiation and may be adapted to study temporal differences, however such tools are not suitable for describing phenotype dynamics. Presented here, is a hidden Markov model of organoid shape geometry, which not only differentiates phenotypes but also links them to distinct dynamics by considering hidden states as unique states of genetic expression which produce similar phenotypes. Furthermore, the task of model selection for Hidden Markov model parameters provides a natural process of determining the number of distinct genetic states in a dataset. Models are trained using perimeter and area time series from two dimensional contours segmented from a sequence of microscopy images of basal and fibroblast growth factor (FGF2) treated organoids. Particularly, dynamical models of area and perimeter are created which find i) the number of distinct phenotypes ii) the evolution of hidden states responsible for observed contour dynamics, and iii) the characteristic dynamics associated with each hidden state. The hidden Markov model of shape geometry is an informative dynamical model capable of detecting phenotype groups and corresponding dynamical behavior regimes.

6.2 Introduction

Organoids are clusters of organ-specific cells, derived from either animal or human tissue [23], which mimic functionality of the target organ [113] and are grown in controlled environments to study the effects of various treatments and disruptions in bioprocesses [119]. The resulting behavior of the organoid is captured via imaging techniques which allow micro-scale assessment of organoid anatomy [162]. This combination of highly tailored experiments and fine resolution spatio-temporal observations provide an opportunity for quantitative analyses and dynamical modeling of organoid morphology. Such models can help bridge the gap towards understanding and controlling multiscale morphogensis of organoids, which is a current limitation in the field of tissue engineering where biomimicry is of high precedence [119]. Particularly, questions regarding phenotype count, differentiation, and behavior are appropriate given time series data from differentially treated organoid experiments. We propose hidden Markov models to leverage spatio-temporal data from organoid experiments in order to shed light on organoid behavior and dynamics.

Hidden Markov models are diversely applicable and have been used successfully in the fields of speech

recognition,[199], genomics [181], motion recognition [214], and finance [82]. Although the application areas are varied, hidden Markov models are generally used in one or a combination of three ways: i) classification ii) forecasting iii) general inference [154]. Classification is made possible when separate models can be trained for each target category in a supervised learning fashion and subsequently be used to calculate the likelihood of an observed sequence. However, when labeled data is not available, or it is not possible to know the generating source of a sequence, hidden Markov models can be used to understand how the sequence is formed, analogous to the unsupervised learning paradigm. Forecasting applications emphasize accuracy and are less concerned with understanding the underlying generating process. Here, we use hidden Markov models for general inference to discover and understand temporal differences in geometery of basal and FGF2 treated organoids. We hypothesize that distinct genomic expressions are responsible for the observed features of an organoid, and these unique states can be modeled as the hidden states in a hidden Markov model, thereby yielding a quantitative model which explains spatio-temporal heterogeneity of organoid phenotypes.

Basal and FGF2 hidden Markov models are created using two steps, preprocessing and model selection (Fig.44), using a dataset of 39 organoids. Subsequently, the models are used to i) associate observed time series to sequences of hidden states and ii) generate synthetic time series samples which further elucidate the role and behavior of the hidden states learned during model selection.

6.3 Methods

6.3.1 Dataset

The dataset is comprised of 9 basal organoids and 30 FGF2 treated organoids which are photographed at 30 min intervals for a timespan of 130 hours for a total of 261 time points. During preprocessing (Fig.44), the organoid contours are segmented at each time point into a set of two dimensional coordinates, which are then used to extract the area and perimeter of the organoid contour. Consequently, each of the $O = 39$ organoids is described by two time series, area and perimeter, of length $L = 261$. We express this as a multivariate time series $\mathbf{X}^o \in \mathbb{R}^{L\times 2}$ for $o = [1, O]$.

6.3.2 Model of organoid shape

To create dynamical models of organoid shape, we train basal and FGF2 hidden Markov models following the schematic in Fig.44 using the multivariate time series $\mathbf{X}^o \in \mathbb{R}^{L\times 2}$. Individual hidden states represent a single genomic expression state and observations generated from the same hidden state represent phenotypes of the same kind. Therefore, transitions between hidden states represent dynamical evolution from one genomic expression to another, and the hidden state Markov chain is the corresponding governing mechanism.

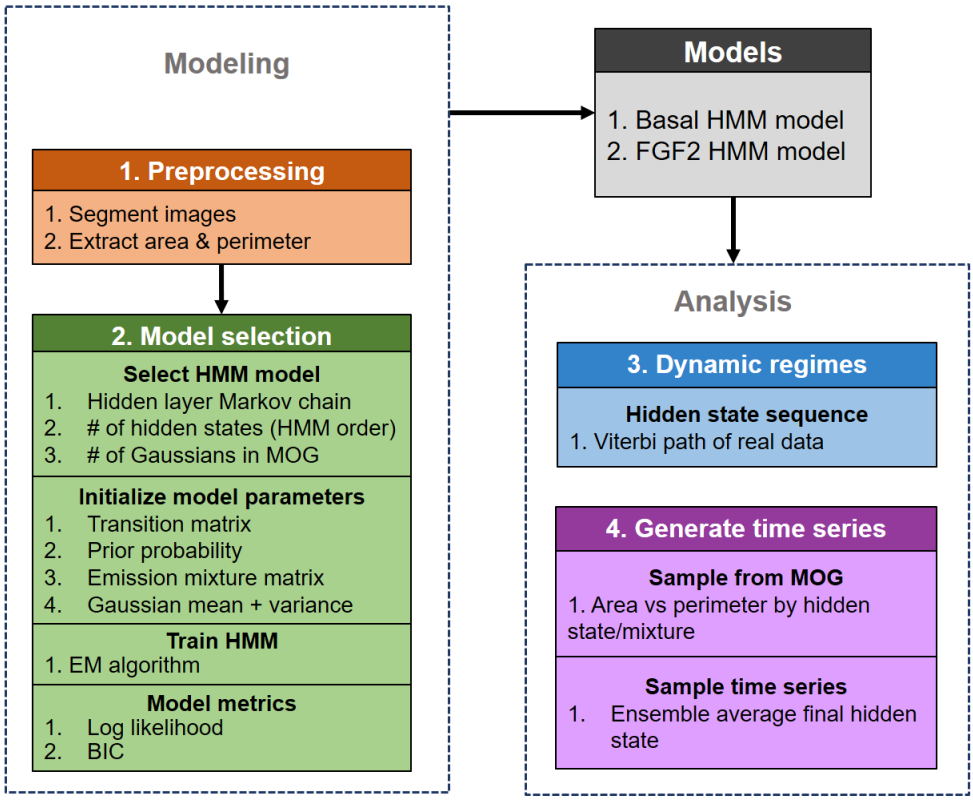


Figure 44: Analysis pipeline. Modeling procedure consists of preprocesing (step 1) of organoid images and model selection (step 2), which yield two distinct models for Basal and FGF2 types. Viterbi and forward algorithms are used to perform the analyses in steps 3 and 4.

Organoids whose geometery varies less throughout the examined time period remain in the same hidden states of genomic expression, whereas organoids which differentiate substantially from the original state are hypothesized to transition through a series of hidden states in the hidden Markov model. The Markov chain order selection process yields an optimal number of hidden states, and in our application to organoid shape modeling, this provides the optimal number of genomic states responsible for the observed range of phenotypes. However, we seek the fewest number of hidden states which can capture the full range of geometric variation in a set of organoid shape time series, as well as a Markov chain whose absorption states model the observed dynamical endpoints in the data. We characterize organoid shape by perimeter and area, as these two descriptors together can be used to define organoid circularity,

$$a_f = \frac{4\pi A}{P^2} \tag{105}$$

which differentiates more regular shaped organoids from more complex shaped organoids that are more invasive. Therefore, circularity, and in turn the combination of perimeter and area forms a computationally

simple and intuitive first order of approximation of contour complexity. Nevertheless, the adaptation of hidden Markov models to model organoid dynamics described here is generalizable to any combination of organoid descriptors.

For the purpose of model selection, the basal model $n_o = 9L$ and the FGF2 $n_o = 30L$ in the BIC expression (Eq.34), and $d = 2$ is the dimensionality of the multivariate observation time series in the number of parameters n_p (Eq.35)

6.4 Results

6.4.1 Perimeter-area space

Basal and FGF2 organoids' perimeter and area time series are shown in Fig.45. Median perimeter decreases for basal organoids Fig.45A and doubles for FGF2 organoids Fig.45C. Median area decreases by 30% for basal organoids Fig.45B, and increases 380% for FGF2 organoids Fig.45D. While the the basal types stagnate or decrease in size, the FGF2 treated organoids grow exponentially.

A more expressive view of the discrepancy in morphogenesis is shown in Fig.46 in the perimeter-area space, where the circular limit of area per perimeter forms an upper bound for all observations. Although basal organoids' area and perimeter decreases, they remain confined to a smaller area of the perimeter-area subspace, and the best fit quadratic has a coefficient of 0.0529. In contrast, FGF2 organoids span a much larger portion of the perimeter-area space and the best fit quadratic has a lower coefficient of 0.0410 because more FGF2 organoids diverge from circularity. Furthermore, unlike the basal organoids which have less varied dynamics, the FGF2 organoids evolve towards a few distinct regions of the perimeter-area space. This heterogeneity in behaviour is captured by the optimal hidden Markov model in the following sections.

6.4.2 Model selection

Selection of parameters for GMM-HMM via BIC, shown in Fig.47, is performed to find the best combination of Markov chain structure $\mathbf{A}$, number of hidden states Q, and number of Gaussians in the mixture of Gaussians M, for both the basal (Fig.47A-C) and FGF2 models (Fig.47D-F). In addition to left-to-right and upper-triangle, bifurcated Markov chain structures which are two-branch tree graphs, are also trained in Fig.47. Each combination of $\mathbf{A}$, Q, and M is simulated 10 times and mean BIC values along with standard deviation bars are shown in Fig.47. BIC decreases with higher values of Q except for basal bifurcated models with $M = 3, 4, 5$ Gaussians. Despite this decrease in BIC, the left-to-right, $Q = 2$, and $M = 2$ model is selected for the basal organoids because of its simplicity and the small basal dataset size of 9 organoids. Given a fixed value of Q, all the FGF2 model BIC values are approximately within one standard deviation of

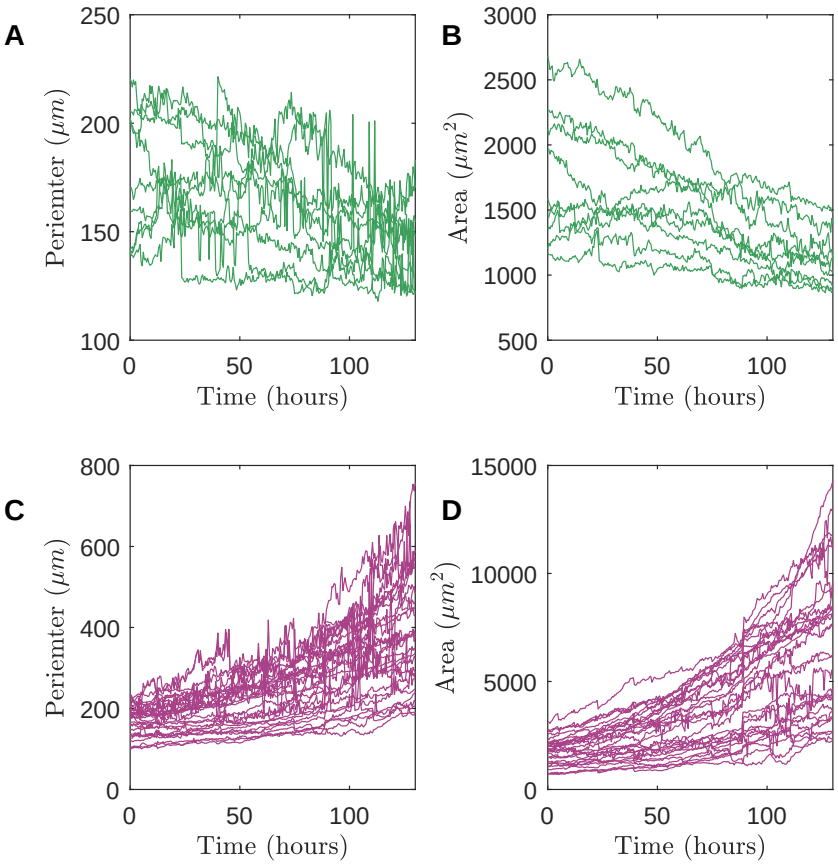


Figure 45: Time series of organoid geometry. Basal organoid (A) perimeter, and (B) area. FGF2 treated organoids' (C) perimeter , and (D) area . Initial perimeter range for basal and FGF2 organoids are 140-220 μm (median = 170 μm) and 102-236 μm (median = 181 μm) respectively. Initial area range for basal and FGF2 organoids are 1170-2680 μm^2 (median = 1590 μm^2) and 718-3150 μm^2 (median = 1840 μm^2) respectively. Final perimeter range for basal and FGF2 organoids are 121-183 μm (median = 148 μm) and 189-757 μm (median = 379 μm) respectively. Final area range for basal and FGF2 organoids are 868-1480 μm (median = 1110 μm) and 2210-14300 μm (median = 6920 μm) respectively.

each other. Although their BIC values are comparable to models with other Markov structures, the left-to-right FGF2 models are not suitable because FGF2 organoids do not converge towards a single region of the perimeter-area space as seen in Fig.46. The bifurcated model holds an advantage over the left-to-right model in that it has two absorption states thus it has the potential to model two distinct dynamical endpoints. Furthermore, compared to upper-triangle structures, bifurcated Markov chain transition matrices **A** have fewer non-zero terms which results in a simpler model. Therefore, the bifurcated, $Q = 5$, and $M = 2$ model is selected for the FGF2 organoids as the other $Q = 5$ models' BIC values are similar, and the bifurcated

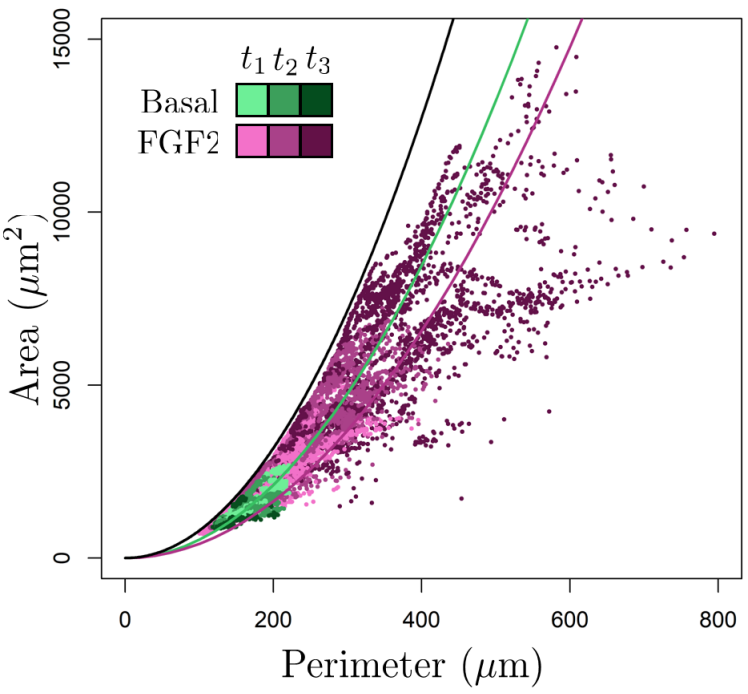


Figure 46: Perimeter and area over three time periods. Three time periods of organoid dynamics shown by individually plotting each contour in a time series: $0 \geq t_1 < 43.5$ hr $\geq t_2 < 87.5$ hr $\geq t_3 \leq 130.5$ hr. The area-perimeter relationship for a circle, $A_{\text{circle}} = \frac{1}{4\pi}P^2 = 0.08P^2$ is shown by the black curve. Quadratic curve fit for 9 basal organoids $A_{\text{Basal}} \approx 0.0529P^2$, and 30 FGF2 organoids, $A_{\text{FGF2}} \approx 0.0410P^2$ are shown in green and purple respectively.

$Q = 7$ BIC is not considerably lower.

6.4.3 Basal and FGF2 hidden Markov models

Fig.48 and Fig.49 summarize the results of i) training, ii) finding the most likely sequence of hidden states associated with each observed organoid time series, and iii) time series sampling of the basal and FGF2 hidden Markov models respectively.

Basal model prior distribution of hidden states,

$$\pi = \begin{pmatrix} 1 & 0 \end{pmatrix}^T \tag{106}$$

transition matrix,

$$\mathbf{A} = \begin{pmatrix} 0.995 & 0.005 \\ 0 & 1 \end{pmatrix} \tag{107}$$

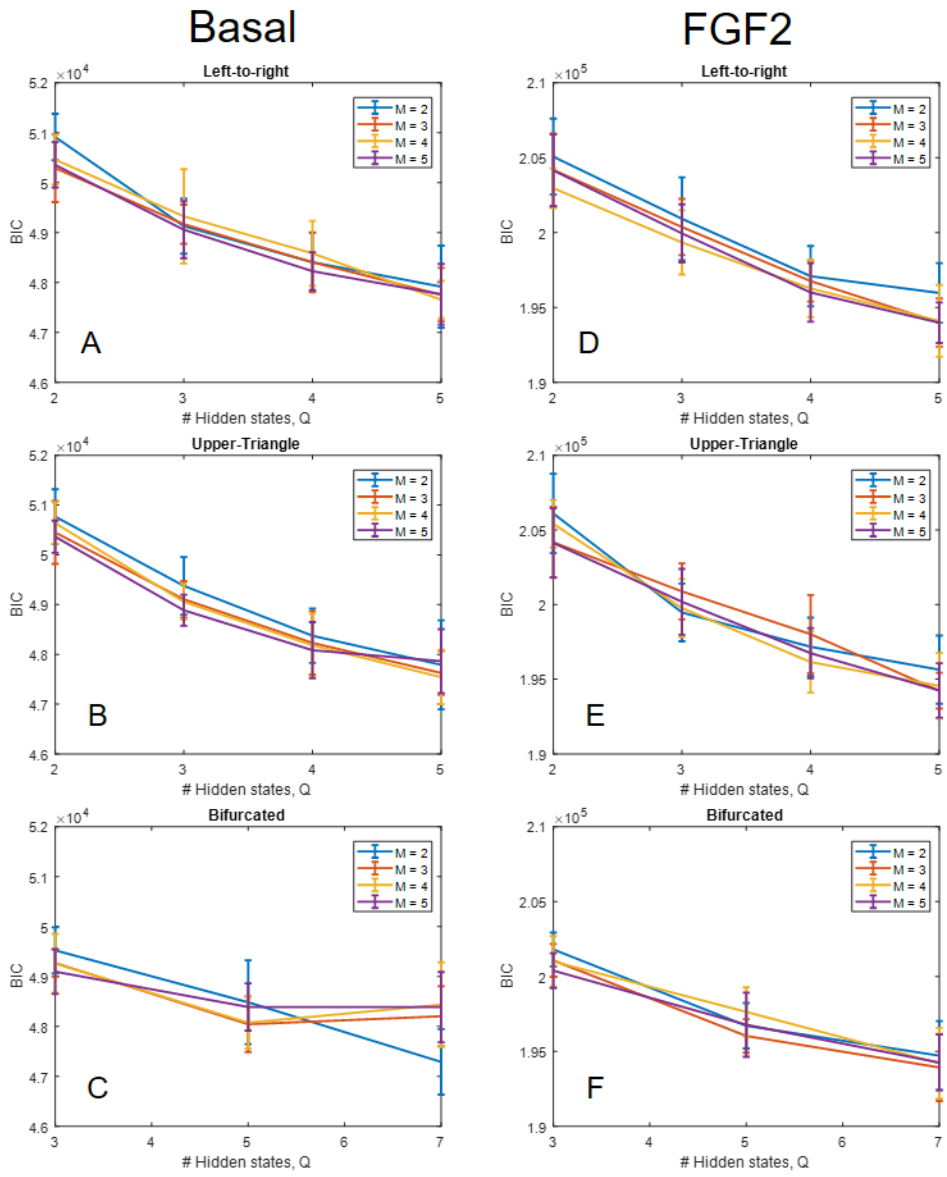


Figure 47: Model selection for basal and FGF2 Hidden Markov models. (A), (B), (C) show BIC values for varying number of hidden states Q, and number of Gaussians M in the emission mixture for left-to-right, upper-triangle and bifurcated hidden state Markov chain structures. (D), (E), (F) show the corresponding BIC values for FGF2 Hidden Markov models.

and emission matrix,

$$\mathbf{B} = \begin{pmatrix} 0.598 & 0.402 \\ 0 & 1 \end{pmatrix} \tag{108}$$

are learned from the 9 basal organoid observed time series. Similarly the FGF2 model is trained using the

30 organoid observed time series, resulting in the following prior,

$$\pi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T \tag{109}$$

transition matrix,

$$\mathbf{A} = \begin{pmatrix} 0.988 & 0.005 & 0 & 0.007 & 0 \\ 0 & 0.99 & 0.01 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.992 & 0.008 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{110}$$

and emission matrix

$$\mathbf{B} = \begin{pmatrix} 0.378 & 0.622 \\ 0.363 & 0.637 \\ 0.434 & 0.566 \\ 0.556 & 0.444 \\ 0.795 & 0.205 \end{pmatrix} \tag{111}$$

500 points are sampled from the trained models to illustrate the region of the perimeter-area space occupied by each hidden state in Fig.48B and Fig.49B. Basal state X_1 generates larger area and perimeter organoids compared to state X_2 (Fig.48B), however, neither states' samples diverge away from the circular upper limit more than the other. The FGF2 initial state X_1 represents the smallest organoids in terms of perimeter and area, and the subsequent states' organoids are progressively larger (Fig.49B). Unlike the basal states, the FGF2 branch formed by states X_2 and X_3 diverges further away from the circular limit line than the X_4 and X_5 branch (Fig.49B). Furthermore, the variance of the mixture of Gaussians in the FGF2 model increases for the later states, unlike the basal model.

The Viterbi algorithm uses the learned models to label the observed perimeter and area time series with most likely hidden states in Fig.48C-D and Fig.49C-D for basal and FGF2 models respectively. All but two basal organoids transition to X_2, as these organoids' final area and perimeter is relatively smaller (Fig.48C-D). All FGF2 organoids transition out of X_1: seven of the smaller organoids end up in intermediate states X_2 and X_4 and 23 larger organoids eventually reach states X_3 and X_5 (Fig.49C-D). Although both FGF2 absorbing states' (X_3 and X_4) organoids are larger than the preceding states' organoids, the organoids which have lower area and high perimeter transition to state X_3 and the organoids which are closer to the circular

limit transition to X_5. In this manner, organoids can be grouped into subcategories of similar phenotypes by labeling their times series with sequences of most likely hidden states using the Viterbi algorithm and the trained models.

To further understand the behavior of the learned basal and FGF2 models, 1000 time series are sampled and averaged from the respective models using the forward algorithm. The generated time series are grouped and colored according to the final hidden state achieved in Fig.48E-I and Fig.49E-I for basal and FGF2 models respectively. The resulting time-averaged trajectories in the perimeter-area space are plotted in Fig.48E and Fig.49E, where the basal state X_1 remains confined to the initial point while X_2 moves along the circularity limit towards lower area and lower perimeter. On average, the basal X_1 states' shapes stagnate (Fig.48F-G) while organoids in state X_2 decrease in size. In contrast, the FGF2 branch states move towards larger perimeter and larger area, and X_5 differs from X_3 in that it produces time series with larger areas than X_3 and stays closer to the circular limit (Fig.49E). Therefore FGF2 state X_3 represents more complex and irregular shapes than the equally large X_5 organoids. This behavior is made clear in Fig.49F-G where state X_1 stagnates, intermediate states X_2 and X_4 grow to a fraction of the size achieved by organoids which transition to states X_3 and X_5 which themselves have the similar mean perimeter but state X_5 does so with more regular larger area organoids.

The average timing of transitions between the states is calculated by computing the mode of hidden states for each hidden state group of generated time series (Fig.48H and Fig.49H). The basal state samples which transition to X_2 do so mostly at the 40 hour mark (Fig.48H). Similarly, in the FGF2 model, on average, the organoids which reach absorption states X_3 and X_5 transition earliest, before the 40 hour mark (Fig.48H) to the intermediate states, and then transition again at approximately 60 hours to reach X_3 and X_5. Therefore, organoids must transition early to achieve large area and large perimeter.

Mean circularity of the groups of generated time series summarizes the corresponding hidden state characteristics in Fig.48I and Fig.49I for basal and FGF2 models respectively. Although the two basal states differ in size, their generated time series have the same mean circularity and never become irregular (Fig.48I). Mean circularity of FGF2 generated time series groups vary considerably (Fig.49I): X_1 and X_4 time series groups maintain high circularity ($a_f \approx 0.8$), X_2 and X_5 time series groups' circularity converges to $a_f \approx 0.6$ and $a_f \approx 0.7$ respectively, with X_2 organoids becoming irregularly earlier than X_5 organoids. Finally, the most irregular shaped X_3 state time series' mean circularity decreases most rapidly towards $a_f < 0.5$.

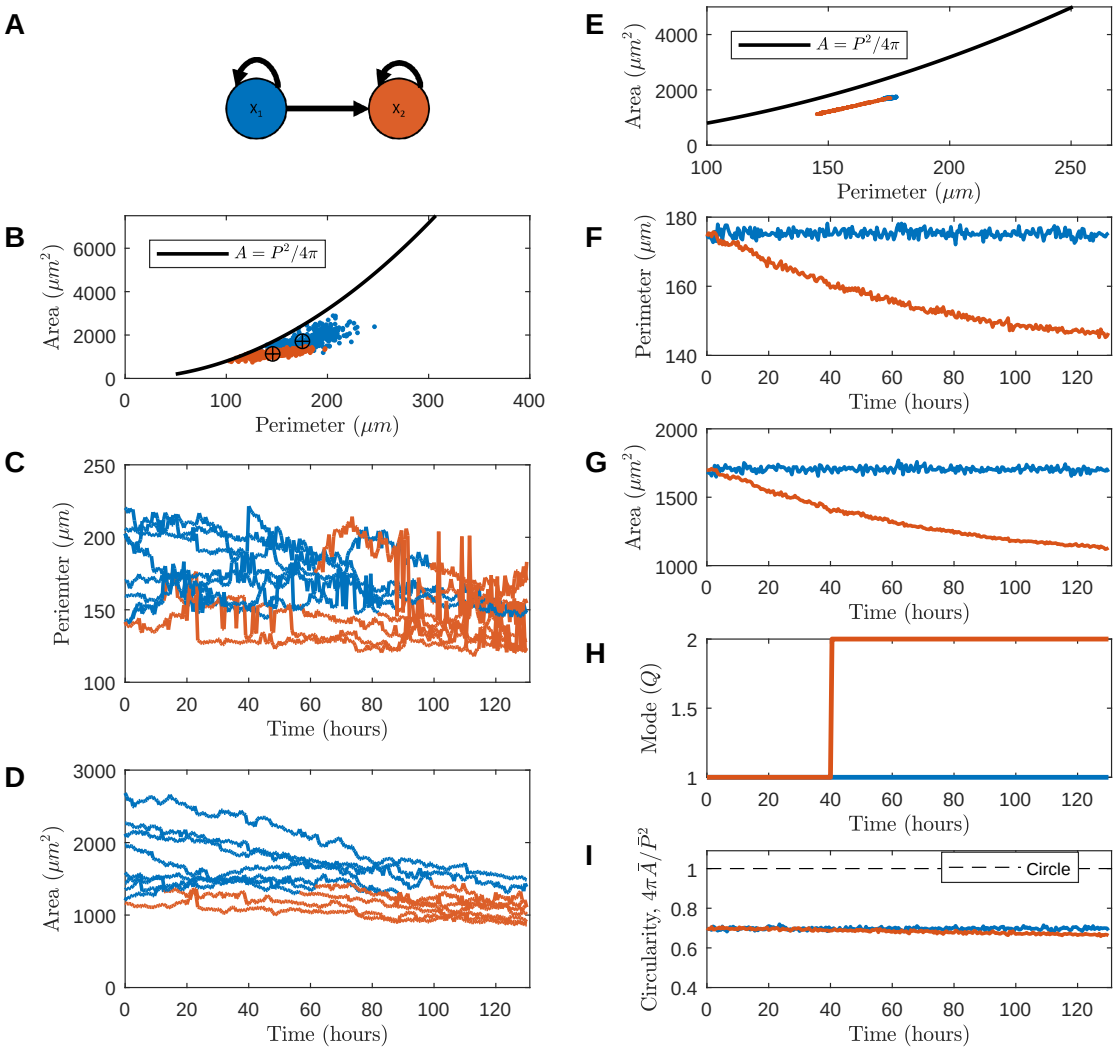


Figure 48: Basal Hidden Markov Model results. (A) left-to-right hidden layer Markov chain with $Q = 2$ hidden states (hidden state color key for the entire figure). (B) 500 individual perimeter and area paired samples from $Q = 2$ hidden states in perimeter-area space. 9 observed organoid (C) perimeter and (D) area time series color coded by most likely hidden state as per Viterbi algorithm. 1000 time series are generated from the learned model and grouped by last hidden state: (E) mean trajectory in perimeter-area space, (F) mean perimeter time series, (G) mean area time series, (H) mode hidden state of all generated series, and (I) mean circularity time series. In (E-I) time series are grouped and colored by the final hidden state at $t = 130$ hours.

6.5 Conclusion

The basal and FGF2 hidden Markov models discover the number of unique genomic expression states responsible for the set of observed contour geometry time series. The basal hidden Markov model is a relatively

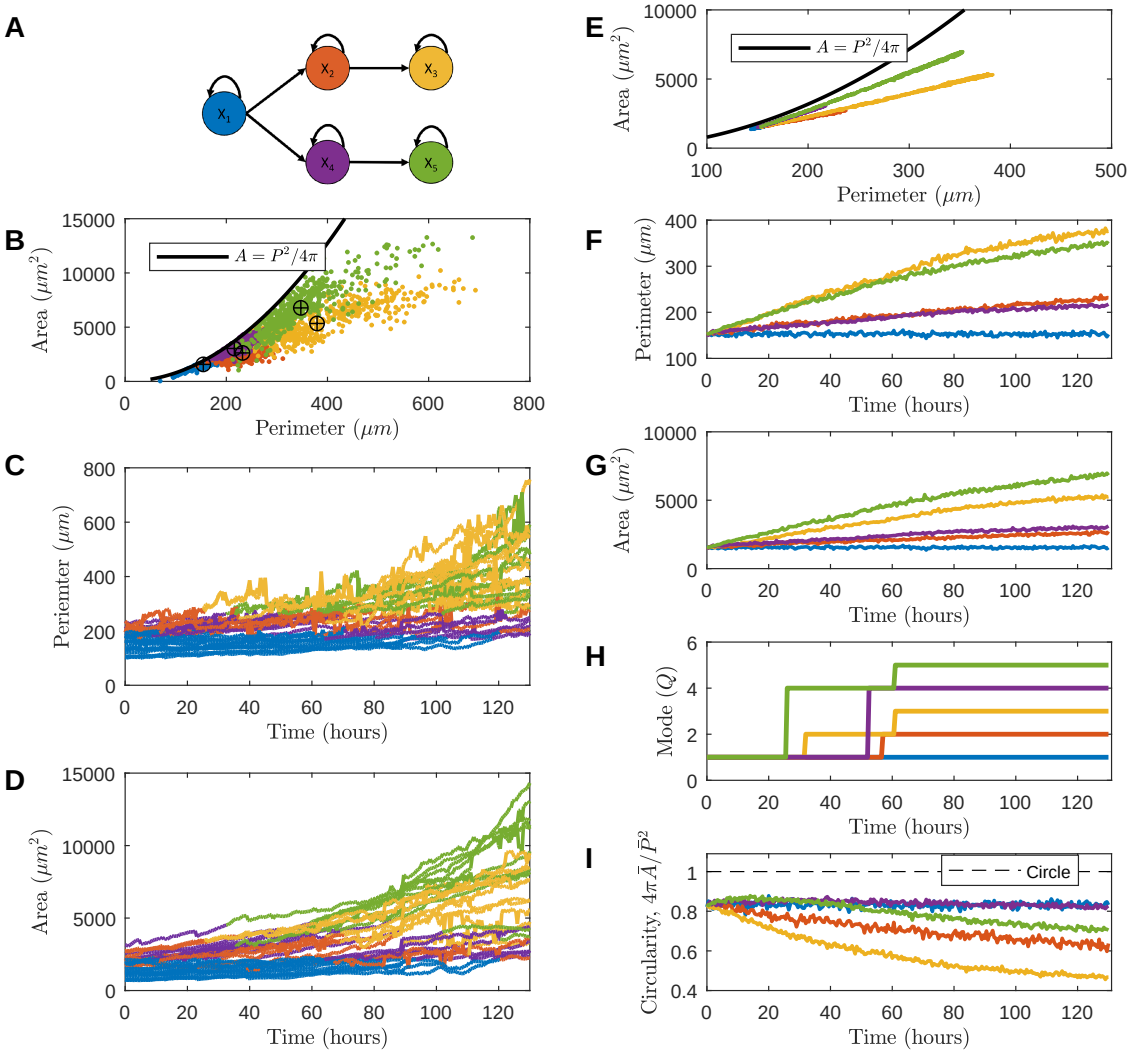


Figure 49: FGF2 Hidden Markov Model results. (A) bifurcated hidden layer Markov chain with $Q = 5$ hidden states (hidden state color key for the entire figure). (B) 500 individual perimeter and area paired samples from $Q = 5$ hidden states in perimeter-area space. 30 observed organoid (C) perimeter and (D) area time series color coded by most likely hidden state as per Viterbi algorithm. 1000 time series are generated from the learned model and grouped by last hidden state: (E) mean trajectory in perimeter-area space, (F) mean perimeter time series, (G) mean area time series, (H) mode hidden state of all generated series, and (I) mean circularity time series. In (E-I) time series are grouped and colored by the final hidden state at $t = 130$ hours.

simple two state model where one state produces organoids whose perimeter and area stay constant and the organoids in the second state shrink to a smaller size than the initial perimeter and area. Both basal hidden states produce organoids of the same circularity, therefore the geometric variation among basal organoids is

low.

The FGF2 hidden Markov model is a more complex bifurcated model with one initial state which connects to two branches for a total of five hidden states. The bifurcated Markov chain model is suitable for FGF2 treated organoids as there are at least two distinct dynamical end points in the perimeter-area space which a majority of the organoids progress towards. Sampling time series from this model reveals the major differences between the two FGF2 phenotype branches. Particularly, large FGF2 organoids can be more circular with matching perimeter and area or irregularly shaped with high perimeter per area which is indicative of more invasive behavior. Furthermore the bifurcated model structure also captures behavior of the FGF2 organoids which do not complete morphogenesis towards the two dynamical extremes. Therefore, the dataset of FGF2 treated organoids contains organoids with five distinct genetic expression states, and the Viterbi algorithm is used to find that all 30 of the FGF2 organoids are in one of four genetic states at the end of the experiment.

We have demonstrated here a novel modeling approach for understanding morphogenesis of organoids. Hidden Markov models for organoid shape time series can learn the number of distinct underlying genetic states in a set of organoids, and then temporally differentiate the organoids into this set of states. Furthermore, the models can be used to generate time series of organoid geometry based on the learned stochastic dynamical progression between the hidden genetic states. These three features provide an advantage over other methods which i) are not directly suited for spatio-temporal differentiation, ii) do not have rigorous methods for detecting number of generating states, and iii) cannot be used to generate samples. The proof-of-concept model shown here with a dataset of 39 organoids can easily be adapted to larger, more heterogeneous datasets which may require additional geometric features to detect more hidden genetic states. The modeling technique developed here can readily be used to study datasets generated from organoid growth experiments, and provide a common quantitative language to compare complex behavior across studies and organoid types.

7 Future work

Future work of each project presented in the previous chapters is discussed below with emphasis on extracting more relevant and insightful results from each data source and experiment type.

The feature and model based techniques developed here to analyze organoid growth need to be employed on larger more heterogeneous datasets to further understand the utility of the feature set, and to identify need for additional features. The binary treatment dataset comprised of growth factor and untreated organoids studied here only shows the potential of these tools, and experiments where organoids are treated with a combination of agents can especially take advantage of the present work. Specifically, the strength of the geometric and signal processing feature based clustering is to identify similar organoids, and the hidden Markov models of organoid shape reveal groups of organoids in similar genomic states and their respective dynamics. Standardizing the application of this analytically pipeline across organoid studies can provide a common quantitative basis with which organoid morpghogenesis can be studied.

The ability to reliably record the motion of patients in the clinic is a crucial step in achieving predictive feature and model based tools which are clinically relevant. At present, there is no clear successor technology to the discontinued Microsoft Kinect camera used in this thesis, and this limits the opportunities for deriving motion data non-invasively. Nevertheless, motion capture data from patient exercises performed in the clinic can be used for various ailments besides cancer. Secondly, such data can be correlated with and consequently used to learn other risk factors besides subjective physician assigned fitness scores. Therefore applying the toolset developed here to study patients suffering non-cancer conditions, and learning the corresponding risk factors is a fruitful direction. However, to better understand the utility of a non-invasive motion capture based fitness predictor, longer longitudinal studies with greater patient enrollment is required. Deploying in-home or mobile versions of this combination of motion imaging technology and performance predicting tools is perhaps the most forward-looking application, and needs to be a focus of future study.

Long-term outcome predictors of bladder cancer cystectomy patients developed here need to be tested on patients at the facilities which provided the data used to train the models as well as external centers in order to truly test the accuracy of predicting survival and recurrence of disease. In order to do this, the models created here need to be deployed as clinician-friendly applications. Concurrently, similar datasets from external centers ought to be used to create similar models to verify the accuracy achieved with the USC dataset, and to fully verify the models' transformational potential of changing identifying at-risk patients.

8 Appendix I: The prisoner's dilemma as a cancer model

8.1 Abstract

Tumor development is an evolutionary process in which a heterogeneous population of cells with different growth capabilities compete for resources in order to gain a proliferative advantage. What are the minimal ingredients needed to recreate some of the emergent features of such a developing complex ecosystem? What is a tumor doing before we can detect it? We outline a mathematical model, driven by a stochastic Moran process, in which cancer cells and healthy cells compete for dominance in the population. Each are assigned payoffs according to a prisoner's dilemma evolutionary game where the healthy cells are the cooperators and the cancer cells are the defectors. With point mutational dynamics, heredity, and a fitness landscape controlling birth and death rates, natural selection acts on the cell population and simulated 'cancer-like' features emerge, such as Gompertzian tumor growth driven by heterogeneity, the log-kill law which (linearly) relates therapeutic dose density to the (log) probability of cancer cell survival, and the Norton–Simon hypothesis which (linearly) relates tumor regression rates to tumor growth rates. We highlight the utility, clarity, and power that such models provide, despite (and because of) their simplicity and built-in assumptions.

8.2 Introduction

Cancer is an evolutionary process taking place within a genetically and functionally heterogeneous population of cells that traffic from one anatomical site to another via hematogenous and lymphatic routes [8, 41, 61, 186, 207]. The population of cells associated with the primary and metastatic tumors evolve, adapt, proliferate, and disseminate in an environment in which a fitness landscape controls survival and replication [126]. Tumorigenesis occurs as the result of inherited and acquired genetic, epigenetic and other abnormalities accumulated over a long period of time in otherwise normal cells [122, 215]. Before we can typically detect the presence of a tumor, the cells are already competing for resources in a Darwinian struggle for existence in tissues that progressively age and evolve. It is well established that the regenerative capacity of individual cells within a tumor, and their ability to traffic multi-directionally from the primary tumor to metastatic tumors all represent significant challenges associated with the efficacy of different cancer treatments and our resulting ability to control systemic spread of many soft-tissue cancers [59, 208]. Details of the metastatic and evolutionary process are poorly understood, particularly in the subclinical stages when tumors are actively developing but not yet clinically visible [180]. It could be argued that in order to truly understand cancer progression at the level in which quantitative predictions become feasible, it is necessary to understand

how genetically and epigenetically heterogeneous populations of cells compete and evolve within the tumor environment well before the tumor is clinically detectable. Additionally, a better understanding of how these populations develop resistance to specific therapies [73, 93] might help in developing optimal strategies to attack the tumor, slow disease progression, or maintain it at a stable level.

Evolutionary game theory is perhaps the best quantitative framework for modeling evolution and natural selection. It is a dynamic version of classical game theory in which a game between two (or more) competitors is played repeatedly, giving each participant the ability to adjust their strategy based on the outcome of the previous string of games. While this may seem like a minor variant of classical (static) game theory, as developed by the mathematicians von Neumann and Morgenstern in the 1940's [202], it is not. Developed mostly by the mathematical biologists John Maynard Smith and George Price in the 1970s [178, 179] and Martin Nowak and Karl Sigmund [141, 145] more recently, this dynamic generalization of classical game theory has proven to be one of the main quantitative tools available to evolutionary biologists (if coupled with a fitness landscape) whose goal is to understand natural selection in evolving populations. In this biological context, a strategy is not necessarily a deliberate course of action, but an inheritable trait [171]. Instead of identifying Nash equilibria, as in the static setting [131, 130], one looks for evolutionary stable strategies (ESS) and fixation probabilities [88, 141] of a subpopulation. This subpopulation might be traced to a specific cell with enhanced replicative capacity (high fitness), for example, that has undergone a sequence of mutations and is in the process of clonally expanding [146]. A relevant question in that case is what is the probability of fixation of that subpopulation? More explicitly, how does one subpopulation invade another in a developing colony of cells?

One game in particular, the Prisoner's Dilemma game, has played a central role in cancer modeling (as well as other contexts such as political science and economics) [9, 10, 12, 48, 49, 69, 70, 83, 84, 88, 91, 22, 95, 142, 144, 145, 193, 197, 201, 206]. It was originally developed by Flood, Dresher and Tucker in the 1950s as an example of a game which shows how rational players might not cooperate, even if it seems to be in their best interest to do so. The evolutionary version of the Prisoner's Dilemma game has thus become a paradigm for the evolution of cooperation among a group of selfish individuals and thus plays a key role in understanding and modeling the evolution of altruistic behavior [9, 10]. Perhaps the best introductory discussion of these ideas is found in Dawkins' celebrated book, The Selfish Gene [44]. The framework of evolutionary game theory allows the modeler to track the relative frequencies of competing subpopulations with different traits within a bigger population by defining mutual payoffs among pairs within the group. From this, one can then define a fitness landscape over which the subpopulations evolve. The fitness of different phenotypes is frequency dependent and is associated with reproductive prowess, while the 'players' in the evolutionary game compete selfishly for the largest share of descendants [88, 206]. Our goal in this article is provide a

brief introduction to how the Prisoner's Dilemma game can be used to model the interaction of competing subpopulations of cells, say healthy, and cancerous, in a developing tumor and beyond.

8.3 The prisoner's dilemma evolutionary game

An evolutionary game between two players is defined by a 2 x 2 payoff matrix which assigns a reward to each player (monetary reward, vacation time, reduced time in jail, etc.) on a given interaction. Let us call the two players A and B. In the case of a prisoner's dilemma game between cell types in an evolving population of cells, let there be two subpopulations of cell types which we will call 'healthy', and 'cancerous'. We can think of the healthy cells as the subpopulation that is cooperating, and the cancer cells as formerly cooperating cells that have defected via a sequence of somatic driver mutations. Imagine a sequence of 'games' played between two cells (A and B) selected at random from the population, but chosen in proportion to their prevalence in the population pool. Think of a cancer-free organ or tissue as one in which a population of healthy cells are all cooperating, and the normal organ functions are able to proceed, with birth and death rates that statistically balance, so an equilibrium healthy population is maintained (on average). Now imagine a mutated cell introduced into the population with enhanced proliferative capability as encoded by its genome as represented as a binary sequence of 0's and 1's carrying forward its genetic information (which is passed on to daughter cells). A schematic diagram associated with this process is shown in Figure 1. We can think of this cancer cell as a formerly cooperating cell that has defected and begins to compete against the surrounding population of healthy cells for resources and reproductive prowess. From that point forward, one can imagine tumor development to be a competition between two distinct competing subpopulations of cells, healthy (cooperators) and cancerous (defectors). We are interested in the growth rates of a 'tumor' made up of a collection of cancer cells within the entire population, or equivalently, we are interested in tracking the proportion of cancer cells, $i(t)$, vs. the proportion of healthy cells, $N - i(t)$, in a population of N cells comprising the simulated tissue region.

To quantify how the interactions proceed, and how birth/death rates are ultimately assigned, we introduce the 2 x 2 prisoner's dilemma payoff matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}. \tag{112}$$

What defines a prisoner's dilemma matrix are the inequalities $c > a > d > b$. The chosen values in (112) are relatively standard, but not unique[1]. The essence of the prisoner's dilemma game is the two players

[1] A general investigation of how the values in the PD payoff matrix affects evolutionary dynamics of the subpopulations is addressed in [205].

compete against each other, and each has to decide what best strategy to adopt in order to maximize their payoff. This 2 x 2 matrix assigns the payoff (e.g. reward) to each player on each interaction. My options, as a strategy or, equivalently, as a cell type, are listed along the rows, with row 1 associated with my possible choice to cooperate, or equivalently my cell type being healthy, and row 2 associated with my possible choice to defect, or equivalently my cell type being cancerous. Your options are listed down the columns, with column 1 associated with your choice to cooperate (or you being a healthy cell), and column 2 associated with your choice to defect (or you being a cancer cell). The analysis of a rational player in a prisoner's dilemma game runs as follows. I do not know what strategy you will choose, but suppose you choose to cooperate (column 1). In that case, I am better off defecting (row 2) since I receive a payoff of 5 instead of 3 (if I also cooperate). Suppose instead you choose to defect (column 2). In that case, I am also better off defecting (row 2) since I receive a payoff of 1 instead of 0 (if I were to have cooperated). Therefore, *no matter what you choose, I am better off (from a pure payoff point of view) if I defect.* What makes this game such a useful paradigm for strategic interactions ranging from economics, political science, biology, and even psychology [9, 178, 206] is the following additional observation. *You will analyze the game in exactly the same way I did (just switch the roles of me and you in the previous rational analysis), so you will also decide to defect no matter what I do.* The upshot if we both defect is that we will each receive a payoff of 1, instead of each receiving a payoff of 3 if we had both chosen to cooperate. The defect-defect combination is a Nash equilibrium [131, 130], and yet it is sub-optimal for both players and for the system as a whole. Rational thought rules out the cooperate-cooperate combination which would be better for each player (3 points each) and for both players combined (6 points). In fact, the Nash equilibrium strategy of defect-defect is the worst possible system wide choice, yielding a total payoff of 2 points, compared to the cooperate-defect or defect-cooperate combination, which yields a total payoff of 5 points, or the best system-wide strategy of cooperate-cooperate yielding a total payoff of 6 points.

The game becomes even more interesting if it is played repeatedly [206], stochastically [197], and with spatial structure [116] with each player allowed to decide what strategy to use on each interaction so as to accumulate a higher payoff than the competition over a sequence of N games. In order to analyze this kind of an evolving set-up, a fitness function must be introduced based on the payoff matrix A. Let us now switch our terminology so that the relevance to tumor cell kinetics becomes clear. When modeling cell competition, one has to be careful about the meaning of the term 'choosing a strategy'. Cells do not choose a strategy, but they do behave in different ways depending on whether they are normal healthy cells cooperating as a cohesive group, with birth and death rates that statistically balance, or whether they are cancer cells with an overactive cell division mechanism (as triggered by the presence of oncogenes) and an underactive 'break' mechanism (as triggered by the absence of tumor suppressor genes) [207]. In our context, it is not

the strategies that evolve, as cells cannot change type based on strategy (only based on mutations), but the prevalence of each cell type in the population is evolving, with the winner identified as the sub-type that first saturates in the population.

8.4 A tumor growth model

Consider a population of N cells driven by a stochastic birth-death process as depicted in Figure 50, with red cells depicting cancer cells (higher fitness) and blue cells depicting healthy cells (lower fitness, but cooperative). We model the cell population as a stochastic Moran process [210] of N cells, 'i' of which are cancerous, '$N-i$' of which are healthy. If each cell had equal fitness, the birth-death rates would all be equal and a statistical balance would ensue. At each step, a cell is chosen (randomly but based on the prevalence in the population pool) and eliminated (death), while another is chosen to divide (birth). If all cells had equal fitness, the birth/death rates of the cancer cells would be i/N, while those of the healthy cells would be $(N-i)/N$. With no mechanism for introducing a cancer cells in the population, the birth/death rates of the healthy cells would be equal, and no tumor would form.

Now, introduce one cancer cell into the population of healthy cells, as shown in Figure 50A. At each step, there would be a certain probability of this cell dividing ($P_{i,i+1}$), being eliminated ($P_{i,i-1}$), or simply not being chosen for either division or death ($P_{i,i}$). Based on this random process, it might be possible for the cancer cells to saturate the population, as shown by one simulation in Figure 51 depicting $N = 1000$ cells, with initially $i = 1$ cancer cell, and $N - i = 999$ healthy cells. However, the growth curve would not show any distinct shape (Figure 51 (black)), and might well become extinct after any number of cell divisions, as opposed to reaching saturation. But we emphasize that without mutational dynamics, heritability, and natural selection operating on the cell population, the shape of the growth curve would look random, and we know this is not how tumors tend to grow [111, 112]. By contrast, Figure 51 (red) shows a Gompertzian growth curve starting with exponential growth of the cancer cell subpopulation, followed by linear growth, ending with saturation. The growth rate is not constant throughout the full history of tumor development, but after an initial period of exponential growth, the rate decelerates until the region saturates with cancer cells. The basic ingredients necessary to sustain Gompertzian growth seem to be: an underlying stochastic engine of developing cells, mutational dynamics, heritability, and a fitness landscape that governs birth and death rates giving rise to some sort of natural selection.

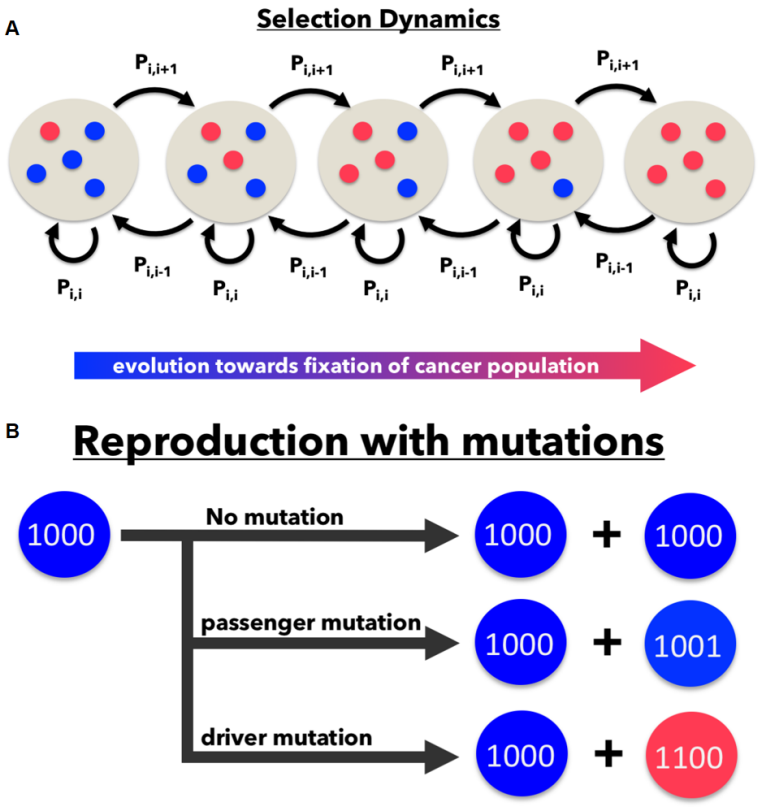


Figure 50: **Schematic of the Moran Process** — (a) The number of cancer cells, i, is defined on the state space $i = 0, 1, \ldots, N$ where N is the total number of cells. The cancer population can change at most by one each time step, so a transition exists only between state i and $i-1$, i, and $i+1$. (b) During each time step, a single cell is chosen for reproduction, where an exact replica is produced. With probability m $(0 \leq m \leq 1)$, a mutation may occur.

8.4.1 Mutations and heritability

Each of the N cells in our simulated population carries with it a discrete packet of information that represents some form of molecular differences among the cells. In our model, we code this information in the form of a 4-digit binary string from 0000 up to 1111, giving rise to a population made up of 16 distinct cell types. At each discrete step in the birth-death process, one of the digits in the binary string is able to undergo a point mutation [62, 122], where a digit spontaneously flips from 0 to 1, or 1 to 0, with probability p_m. The mutation process is shown in Figure 50, while a mutation diagram is shown in Figure 52 in the form of a directed graph. This figure shows the possible mutational transitions that can occur in each cell, from step to step in a simulation. A typical simulation begins with a population of N healthy cells, all with identical binary strings 0000. The edges on the directed graph represent possible mutations that could occur on a given step. The first 11 binary string values (0-10) represent healthy cells in our model that are at different stages in their evolutionary progression towards becoming a cancer cell (the exact details of this genotype

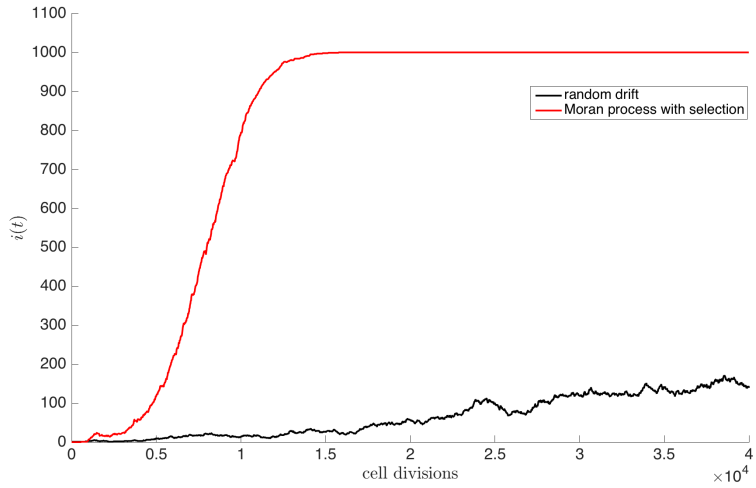


Figure 51: **Emergence of Gompertzian growth via selection** — Random drift (black) plotted for a single simulation of 10^3 cells for $4 \cdot 10^4$ generations shows no particular shape. A single simulation of the Moran process (red) with selection ($w = 0.5$) and mutations ($m = 0.1$) gives rise to the characteristic S-shaped curve associated with Gompertzian growth.

to phenotype map do not matter much). Mutations strictly within this subpopulation are called passenger mutations as the cells all have the same fitness characteristics. The first driver mutation occurs when a binary string reaches value 11-15. The first cell that transitions from the healthy state to the cancerous state is the renegade cell in the population that then has the potential to clonally expand and take over the population. How does this process occur?

8.4.2 The fitness landscape

At the heart of how the Prisoner's Dilemma evolutionary game dictates birth and death rates which in turn control tumor growth, is the definition of cell fitness. Let us start by laying out the various probabilities of pairs of cells interacting and clearly defining payoffs when there are i cancer cells, and $N - i$ healthy cells in the population. The probability that a healthy cell interacts with another healthy cell is given by $(N - i - 1)/(N - 1)$, whereas the probability that a healthy cell interacts with a cancer cell is $i/(N - 1)$. The probability that a cancer cell interacts with a healthy cell is $(N - i)/(N - 1)$, whereas the probability that a cancer cell interacts with another cancer cell is $(i - 1)/(N - 1)$.

In a fixed population of N cells, with i cancer cells, the number of healthy cells is given by $N - i$. The average payoff of a single cell (π^H, π^C), is dependent on the payoff matrix value weighted by the relative frequency of types in the current population:

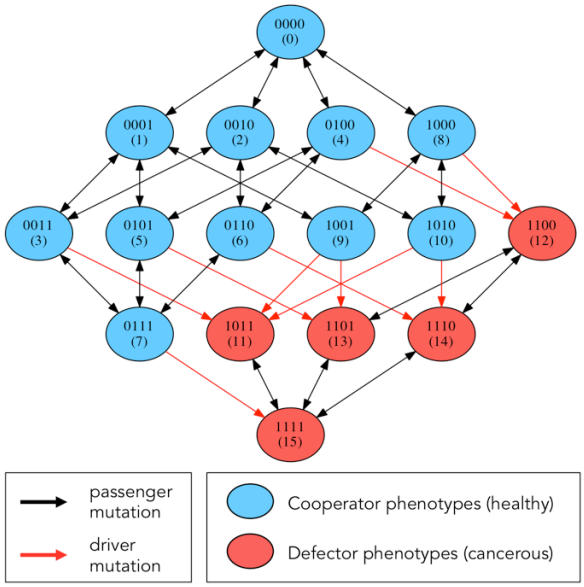


Figure 52: **Markov Point Mutation Diagram** — Diagram shows 16 genetic cell types based on 4-digit binary string and the effect of a point mutation on each cell type. Blue indicates healthy cell type (0000 — 1010), red indicates cancerous cell type (1011 — 1111). Black arrows indicate passenger mutations (healthy to healthy or cancer to cancer), red arrows indicate driver mutations (healthy to cancer).

$$\pi_i^H = \frac{a(N-i-1)+bi}{N-1} \tag{113}$$

$$\pi_i^C = \frac{c(N-i)+d(i-1)}{N-1} \tag{114}$$

Here, $a = 3$, $b = 0$, $c = 5$, $d = 1$ are the entries in the Prisoner's Dilemma payoff matrix (112). For the Prisoner's dilemma game, the average payoff of a single cancer cell is always greater than the average payoff for a healthy cell (Figure 53C). With the invasion of the first cancer cell, the higher payoff gives a higher probability of survival when in competition with a single healthy cell.

Selection acts on the entire population of cells as it depends not on the payoff, but on the effective fitness of the subtype population. The effective fitness of each cell type (f^H, f^C) is given by the relative contribution of the payoff of that cell type, weighted by the selection pressure:

$$f_i^H = 1 - w + w\pi_i^H \tag{115}$$

and the fitness of the cancer cells as:

$$f_i^C = 1 - w + w\pi_i^C \tag{116}$$

The probability of birthing a new cancer cell depends on the relative frequency (random drift) weighted by the effective fitness, and the death rate is proportional to the relative frequency. The transition probabilities can be written:

$$P_{i,i+1} = \frac{if_i^C}{if_i^C + (N-i)f_i^H}\frac{N-i}{N} \tag{117}$$

$$P_{i,i-1} = \frac{(N-i)f_i^H}{if_i^C + (N-i)f_i^H}\frac{i}{N} \tag{118}$$

$$P_{i,i} = 1 - P_{i,i+1} - P_{i,i-1}; \quad P_{0,0} = 1; \quad P_{N,N} = 1. \tag{119}$$

In the event of the first driver mutation, the first cancer cell is birthed. At the beginning of the simulation, the effective fitness of the healthy population is much greater than the fitness of the cancer population (Figure 53B). This is because although the single cancer has a higher *payoff* than any of the healthy cells, the number of healthy cells far outnumber the single cancer cells. That single cancer cell initiates a regime of explosive high growth and the fitness of the cancer population steadily increases. Cancer cells are continually competing with healthy cells and receiving a higher payoff in this regime (compare the payoff entries of a cancer cell receiving $c = 5$ vs a healthy cell receiving $b = 0$). At later times, growth slows because cancer cells are competing in a population consisting mostly of other cancer cells. The payoff for a cancer cell is dramatically lower when interacting with a cancer cell (observe the payoff entry of both cancer cells receiving $d = 1$ when interacting). As the cancer population grows, the payoff attainable decreases and growth slows. In addition, the average fitness of the total population steadily declines because each interaction derives less total payoff, from $c+b = 5$ to $d+d = 1$. It is precisely the payoff structure of the Prisoner's Dilemma matrix that produces this declining average population fitness as the cancer cells saturate the population. Although they receive higher payoffs than healthy cells on pairwise interactions, these cancer-healthy interactions mostly take place early on in the evolution of the tumor. As the cancer cells take over the population, most of the interactions take place between pairs of cancer cells (i.e. they eventually start competing with each other) causing the population fitness to decline.

This complex process of competition among cell types and survival of subpopulations, where defection is selected over cooperation, produces a Gompertzian growth curve shown in Figure 54, and compared with a compilation based on a wide range of data first shown in [111, 112]. It is now well established that tumor cell

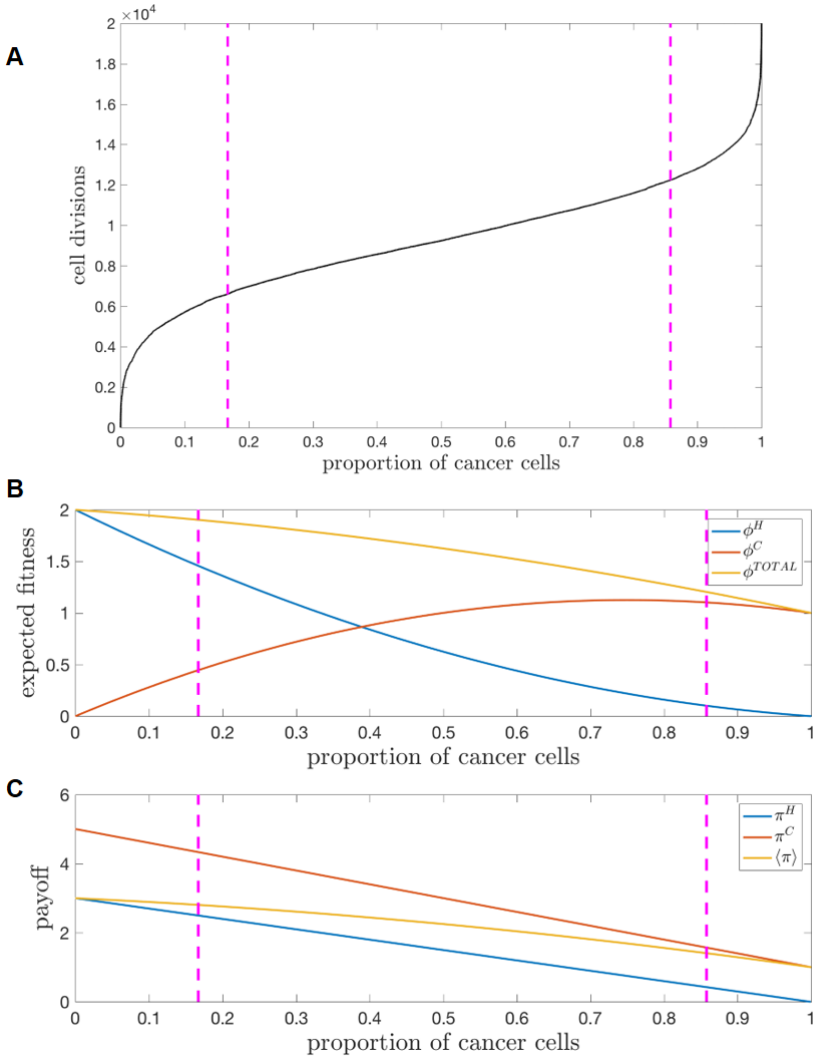


Figure 53: **Tumor fitness drives tumor growth** — (A) The average of 25 stochastic simulations ($N = 1000$ cells, $w = 0.5$, $m = 0.1$) is plotted for 20,000 cell divisions to show the cancer cell population (defectors) saturating. The pink lines delineate the regions of tumor growth (defined by the maximum and minimum points of the second-derivative of $i(t)$). (B) Fitness of the healthy population, cancer population, and total population plotted for the range cancer cell proportion. (C) Average payoff of a single healthy cell, cancer cell, and all cells plotted for the range cancer cell proportion.

populations (and other competing populations, such as bacteria and viral populations) generally follow this growth pattern, although the literature is complicated by the fact that different parts of the growth curve have vastly different growth rates [111, 112], and it is nearly impossible to follow the growth of a population of cancer cells *in vivo* from the first cancer cell through to an entire tumor made up of $O(10^9 - 10^{12})$ cells. Growth rates are typically measured for a short clinical time period [111, 112], and then extrapolated back to the first renegade cell, and forward to the fully developed tumor population.

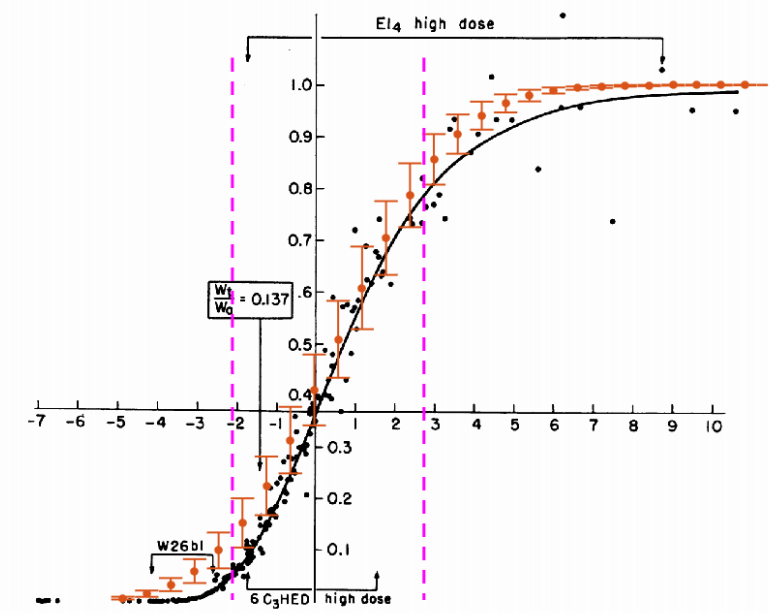


Figure 54: **Moran Process fit to Gompertzian Growth Data** — The mean and deviation of 25 stochastic simulations ($N = 10^3$ cells, $w = 0.7$, $m = 0.3$) is overlaid on data from a "normalized" Gompertzian [111, 112]. Values for m and w were chosen by implementing a least-squares fit to the data over a range of m ($0 \leq m \leq 1$), and w ($0 \leq w \leq 1$). Pink lines delineate regions of growth (defined by the maximum and minimum points of the second-derivative of $i(t)$).

8.4.3 Heterogeneity drives growth

Insights into the process by which growth rates vary and conspire to produce a Gompertzian shape can be achieved by positing that growth is related to molecular and cellular heterogeneity of the developing population [27, 100, 186]. Indeed, an outcome of the model is that molecular heterogeneity (i.e. the dynamical distribution of the 4-digit binary string 0000—1111 making up the population of cells) drives growth. Consider entropy [40, 134] of the cell population as a measure of heterogeneity:

$$E(t) = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{120}$$

(here, log is defined as base 2). The probability p_i measures the proportion of cells of type i, with $i = 1, \ldots, 16$ representing the distribution of binary strings ranging from 0000 to 1111. We typically course-grain this distribution further so that cells having strings ranging from 0000 up to 1010 are called 'healthy', while those ranging from 1011 to 1111 are 'cancerous'. Then growth is determined by:

$$\frac{dn_E}{dt} = \alpha E(t) \tag{121}$$

It follows from (133) that the cancer cell proportion $n_E(t)$ can be written in terms of entropy as:

$$n_E(t) = \alpha \int_0^t E(t)dt \tag{122}$$

This relationship between growth of the cancer cell population and entropy is pinned down and detailed in [210]. We consider it to be one of the key emergent features of our simple model.

A typical example of the emergence of genetic heterogeneity in our model system is shown in the form of a phylogenetic tree in Figure 55. This particular tree is obtained via a simulation of only 30 healthy phenotypic cells (0000), which during the course of a simulation expand out (radially in time) to form a much more heterogeneous population of cells at the end of the simulation. In our model, the genetic time-history of each cell is tracked and the population can be statistically analyzed after the simulation finishes.

8.5 Simulated drug dosing strategies and therapeutic response

Figure 56 shows the clear advantage of early stage therapy in our model system. We compare the effect of therapy given at an early stage, mid-stage, and late stages of the Gompertzian growth of the tumor. The dashed black Gompertzian curve is the freely growing cancer cell population. In each of the figures, we depict the effect of a range of drug dose densities, D, where

$$D = c \cdot t. \tag{123}$$

The dose density is the product of the drug concentration, c, and the time over which the therapy is administered, t, (123). Here, the drug concentration value is a weighting ($0 \leq c \leq 1$) which determines the intensity of the drug treatment. A higher value of c will alter the selection pressure in favor of healthy cells (and to the disadvantage of cancer cells) more dramatically.

Figure 56 varies the drug dose density by varying the drug concentrations ($c = 0.2, 0.4, 0.6, 0.8, 1.0$) administered for a constant time (t = 5000 cell divisions, black arrow). The colored curves show the subsequent decline of the cancer cell population under therapeutic pressure. Clearly, to obtain the desired effect of driving the cancer cell population down to manageable levels, one needs to (i) use a sufficient dose density, and (ii) initiate therapy early enough in the growth history. These figures are meant to paint a broad brush with respect to the simulated advantages of early therapy and to show the capability of the model with respect to addressing questions of this type in a quantitative way. A detailed investigation is left

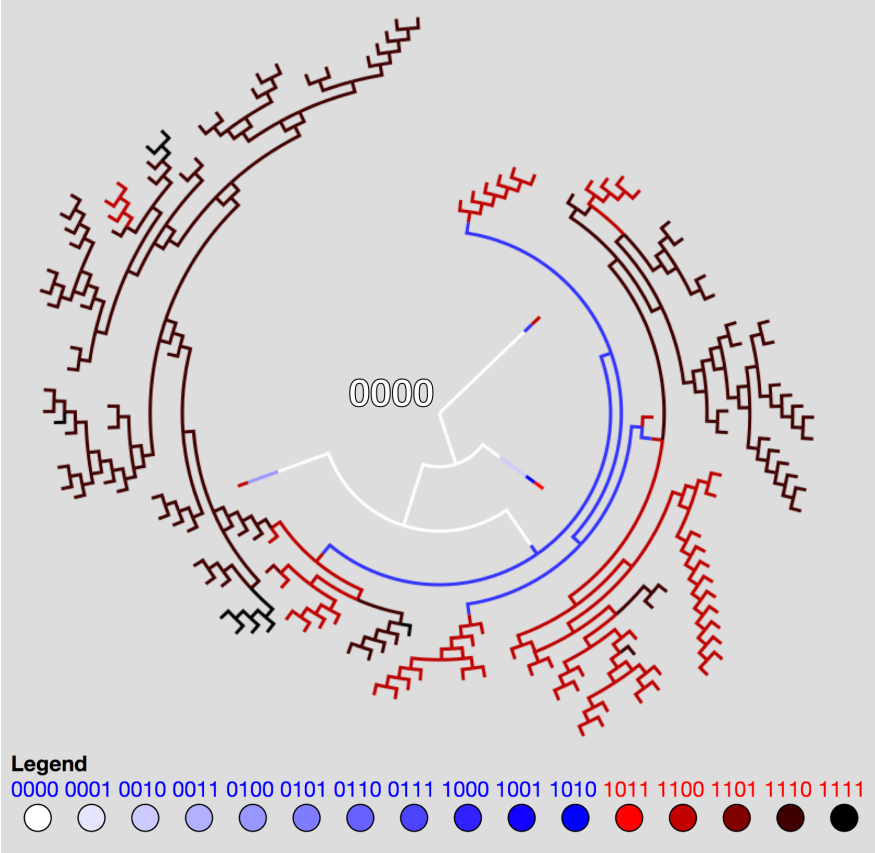


Figure 55: **Sample Dendritic Phylogenetic Tree** — Sample dendritic phylogenetic tree tracking point mutations as time extends radially, depicting the emergence of molecular heterogeneity. The tree shows a simulation of 30 cells (all with genetic string 0000 at the beginning of the simulation) with strong selection ($w = 1$, $m = 0.2$). Pathways are color coded to indicate genetic cell type.

for a separate publication.

An established empirical law which relates drug dose density to its effectiveness in killing off cancer cells is known as the 'log-kill' law [177]. The log kill law states that a given dose of chemotherapy kills the same fraction of tumor cells (as opposed to the same number of tumor cells), regardless of the size of the tumor at the time the therapy is administered [177], a consequence of exponential growth with a constant growth rate. This effect is best illustrated on a dose-response curve, plotting the dose density, D, with respect to the probability of tumor cell survival, P_S. Thus, the log-kill law states the following:

$$\log(P_S) = -\beta D. \tag{124}$$

As an example, if there are 1000 cancer cells in a tumor population, and the first therapy dose kills off 90% of them (i.e. $\beta = 0.9$), then after the first round of therapy there will be 100 cancer cells remaining. If

a second round of therapy is administered, exactly as the first round, starting soon enough so that no new cancer cells have formed, then this next round will also kill off 90% of the cells, leaving 10 cells, and so on for each future round of therapy. In a sense, since the first round killed 900 cells, while the second identical round killed only 90 cells, the population gets increasingly more difficult to kill off using the same treatment on each cycle. The log-kill law, a fundamentally static law (it does not say anything about the relationship of the fraction of cells killed vs. the growth rate of the tumor), is verified in our model system, as shown in the dose response curve in Figure 57D. On the x-axis, we increase the dose density D, and we plot the number of surviving cancer cells. The slope of this straight line (verifying the log-kill law) is the rate of regression of the tumor, β. Alternatively, β can be estimated using an exponential fit of $i(t)$ during therapy (i.e. $i(t) = i_0\texttt{exp}(-\beta(t - t_0))$, where i_0 is the initial tumor size and t_0 is the time therapy is initiated).

So how is the rate of regression, β, related to the growth rate of the tumor, γ? This is relevant, since we know from the shape of the Gompertzian curve, the growth rate is highest (exponential) at the beginning stage of tumor development and lowest at the late saturation stage. Figure 57A shows therapy is more effective (i.e. a higher rate of regression, β) for earlier stage therapy. These early stage therapies correspond to a higher growth rate, shown in Figure57B. The Norton-Simon hypothesis [139, 140, 140] states that the rate of regression is proportional to the instantaneous growth rate for an untreated tumor of that size at the time therapy is first administered. A faster growing tumor (early stage) should show a higher rate of regression than a more slowly growing tumor (late stage). This hypothesis is also verified in our model system, and shown clearly in Figure 57C. The reality of this growth-dependent tumor regression rate effect (where early stage faster growing tumors are more vulnerable to therapy than later stage, slowly growing tumors) dramatically reinforces the need to administer drug treatment early in tumor progression when growth rates are high and there are fewer cancer cells to kill off.

8.6 Mathematical modeling and tumor analytics

It is important to keep in mind that no mathematical model captures all aspects of reality, so choices must be made which involve prioritizing the features that are most essential in capturing the essence of a complex process and which are not. Most experts now agree that the evolutionary processes in a tumor played out among subpopulations of competing cells are key to understanding aspects of growth and resistance to chemotherapy, which will ultimately lead the way toward a quantitative understanding of tumor growth and cancer progression [126, 208, 207]. The paradigm of the cancer cell subpopulation and the healthy cell subpopulation competing as the defectors and cooperators in a Prisoner's Dilemma evolutionary game has been useful in obtaining a quantitative handle on many of these processes and frames the problem in an

intuitive yet predictive way.

Nonetheless, the mathematical 'taste' of the modeler plays a role in what techniques are selected and ultimately where the spotlight shines. This fact makes clinicians uncomfortable and can lead to deep suspicion of the mathematical modeling enterprise as a whole. Aren't the outcomes and predictions of mathematical models a straightforward consequence of the model assumptions? Once those choices are made, isn't the cake already baked? So why should we be surprised if you tell us it tastes good? Why not simply use tried and true statistical tools like regression methods to curve-fit the data directly, with no built in assumptions, and be satisfied with uncovering correlations and trends? Clinicians (and experimentalists, in general) feel that they are dealing directly with reality, so why mess around with 'toy' systems based on possibly 'ad hoc' or incorrect assumptions that create artificial realities that may or may not be relevant? To a theoretician, calling their assumptions ad hoc, as opposed to natural, is as insulting as calling a clinician sloppy and uncaring (try this for yourself at the next conference you go to! But please use the term 'somewhat ad hoc' to lessen the blow.) And if you want to deliver an even harsher insult, you could comment that the model seems like an exercise in curve fitting.

But the usefulness of mathematical models built on simplified assumptions is well established in the history of the physical sciences, as detailed beautifully in Peter Dear's book, *The Intelligibility of Nature: How Science Makes Sense of the World* [45]. Bohr's simple model of the structure of the atom was crucial in moving the community forward towards a deeper understanding of cause and effect, and ultimately pushing others to develop more realistic atomic models. The same could be said for many other important, but ultimately discarded models of reality (e.g. the notion of aether used as a vehicle to understand the mysterious notion of action-at-a-distance [45]) now relegated to footnotes in the history of the physical sciences.

Lessons from this history highlight the importance of using the principle of Occam's razor (law of parsimony) as a heuristic guide in developing models: (1) keep things simple, but not too simple; (2) see what can be explained by using a given set of assumptions, and try to identify what is either wrong or cannot be explained; (3) add complexity to the model, but do this carefully. Since ultimately, the model will always be wrong (with respect to some well chosen and specific new question being posed about a system), it is important that it be *useful as a vehicle of intelligibility* [45] associated with the set of questions surrounding the phenomena it was built to explain. Answers to some new questions will be found using the model as a temporary crutch, and new questions will emerge in the process that had not yet been asked, as their relevance had never previously been realized. A new quantitative language will emerge in which aspects of the model will be associated with the underlying reality it is attempting to describe, predictions will be easier to frame and test, and shortcomings will be exposed. In his famous article [212], Eugene Wigner writes

compellingly that 'the miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps also to our bafflement, to wide branches of learning.'

In general, the more complex the model (as measured, for example, by the number of independent parameters associated with it), the less useful it will be, and the less likely it is to be adopted by the community at large. After all, if the model is as complex as the phenomena it was built to understand, why not stick with reality? Effective models can be thought of as *low-dimensional approximations of reality*, surrogates that help us bootstrap our way forward. They arise as the outcome of a complex balancing act between simplicity of the ingredients, and complexity of the reality the model is meant to describe. They generally do not arise in a vacuum, but are built in the context of informed and sustained discussions among people with different expertise. In the context of medical oncology, this means physical scientists developing ongoing interactions with clinical oncologists, radiologists, pathologists, molecular and cell biologists and other relevant medical specialists.

Appropriate data is a necessary ingredient in developing and testing any successful model, and treasure troves of medical data sit unexamined in patient files and government databases across the country waiting to be put to good use. There is no doubt that they are telling an interesting and important story that we have yet to fully understand. It is not currently possible for the computer to simulate all of the complex, relevant, and systemic ingredients at play to faithfully recreate all aspects of cancer progression and treatment response in patients. It is hard to imagine that a deep and actionable understanding can ever be obtained without the combined use of data, models, and computer simulations to help guide us and highlight some of the underlying causal mechanisms of this complex and deadly disease.

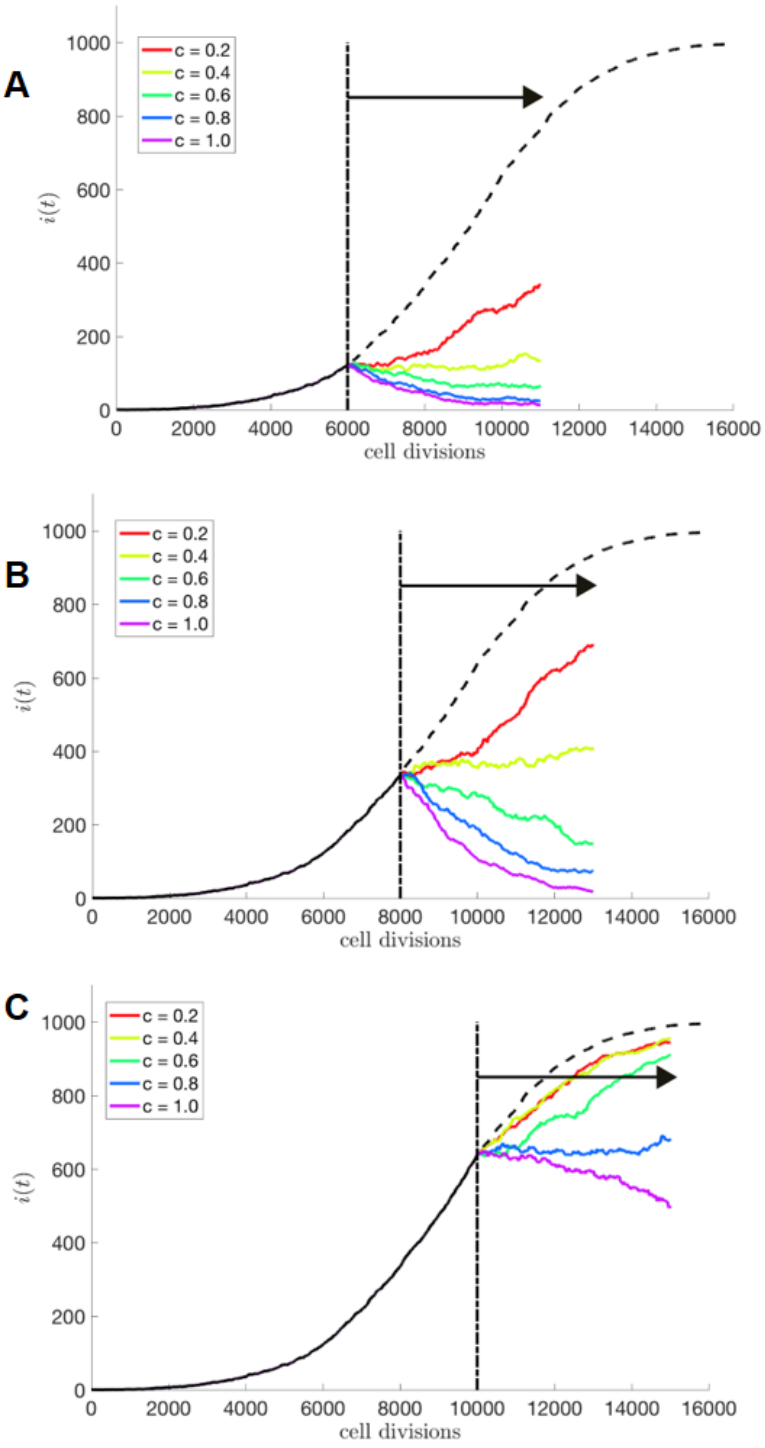


Figure 56: **Effects of varied dose density for early-stage, mid-stage, and late-stage therapies** — An average of 25 stochastic simulations of unperturbed tumor growth ($N = 10^3$ cells, $w = 0.5$, $m = 0.1$, no therapy) is plotted (black dashed line). The effect of varied drug dose density (eqn. 123), is shown by administering a range of drug concentration values ($c = 0.2, 0.4, 0.6, 0.8, 1.0$) for constant length of time ($t = 5000$ cell divisions, black solid arrows). This process is repeated for (A) high growth, early-stage, (B) linear growth, mid-stage, and (C) slowed growth, late-stage. The kill effect is highest for high drug concentration values (i.e. high dose density) and early therapy.

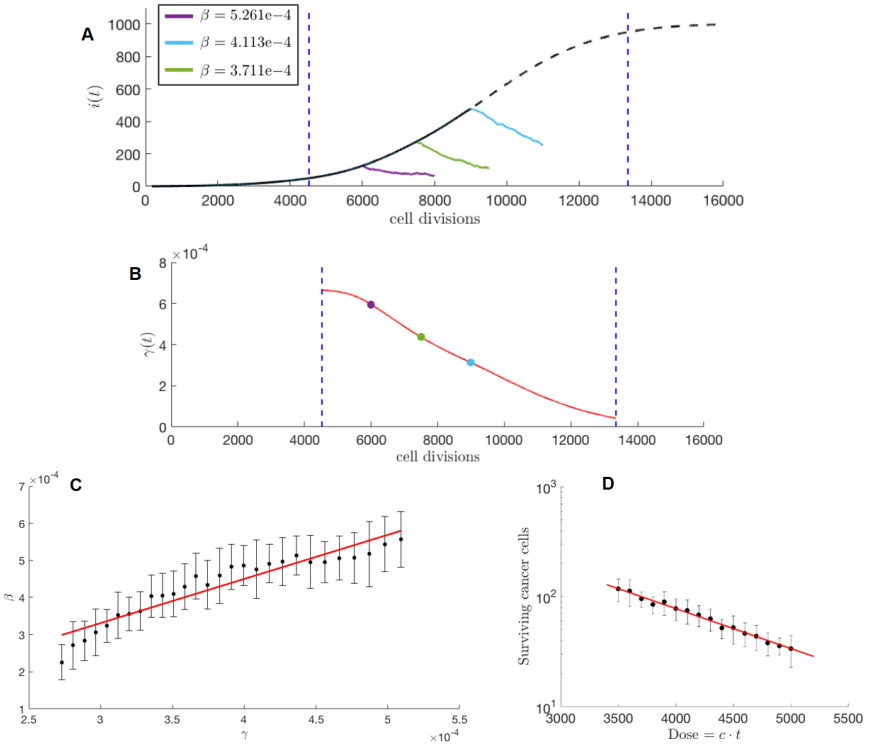


Figure 57: **Growth-dependent tumor regression** — (A) An average of 25 stochastic simulations of unperturbed tumor growth ($N = 10^3$ cells, $w = 0.5$, $m = 0.1$, no therapy) is plotted (black dashed line) with (B) corresponding instantaneous growth rate, $\gamma(t)$, of the unperturbed tumor (red). Tumor regression, β, (estimated using an exponential fit of $i(t)$ during therapy, shown in legend) during therapy (constant dose density: $c = 1.0$, $t = 2000$) is calculated for a high growth, early-stage therapy (purple), linear-growth, mid-stage therapy (green), and late-stage, slowed growth (light blue); (C) This process is repeated for a full range of growth rates (between vertical blue dashed lines). Average values of β are plotted with standard deviations. Regression is proportional to growth rate (linear fit in red), with higher regression rates associated with high growth rates of early stage tumors. (D) Tumor regression, β, can also be calculated as the slope of a dose response curve (red), where therapy is administered for a range of dose densities ($0.7 \leq c \leq 1.0$) for a single timepoint, 8000 cell divisions (i.e. single growth rate).

9 Appendix II: An evolutionary model of tumor cell kinetics and the emergence of molecular heterogeneity driving Gompertzian growth

9.1 Abstract

We describe a cell-molecular based evolutionary mathematical model of tumor development driven by a stochastic Moran birth-death process. The cells in the tumor carry molecular information in the form of a numerical genome which we represent as a four-digit binary string used to differentiate cells into 16 molecular types. The binary string is able to undergo stochastic point mutations that are passed to a daughter cell after each birth event. The value of the binary string determines the cell fitness, with lower fit cells (e.g. 0000) defined as healthy phenotypes, and higher fit cells (e.g. 1111) defined as malignant phenotypes. At each step of the birth-death process, the two phenotypic sub-populations compete in a prisoner's dilemma evolutionary game with the healthy cells playing the role of cooperators, and the cancer cells playing the role of defectors. Fitness, birth-death rates of the cell populations, and overall tumor fitness are defined via the prisoner's dilemma payoff matrix. Mutation parameters include passenger mutations (mutations conferring no fitness advantage) and driver mutations (mutations which increase cell fitness). The model is used to explore key emergent features associated with tumor development, including tumor growth rates as it relates to intratumor molecular heterogeneity. The tumor growth equation states that the growth rate is proportional to the logarithm of cellular diversity/heterogeneity. The Shannon entropy from information theory is used as a quantitative measure of heterogeneity and tumor complexity based on the distribution of the 4-digit binary sequences produced by the cell population. To track the development of heterogeneity from an initial population of healthy cells (0000), we use dynamic phylogenetic trees which show clonal and sub-clonal expansions of cancer cell sub-populations from an initial malignant cell. We show tumor growth rates are not constant throughout tumor development, and are generally much higher in the subclinical range than in later stages of development, which leads to a Gompertzian growth curve. We explain the early exponential growth of the tumor and the later saturation associated with the Gompertzian curve which results from our evolutionary simulations using simple statistical mechanics principles related to the degree of functional coupling of the cell states. We then compare dosing strategies at early stage development, mid-stage (clinical stage), and late

stage development of the tumor. If used early during tumor development in the subclinical stage, well before the cancer cell population is selected for growth, therapy is most effective at disrupting key emergent features of tumor development.

9.2 Introduction

At the molecular and cellular levels, cancer is an evolutionary process [126, 138, 138, 138] driven by random mutational events [138, 104, 192, 166] responsible for genetic diversification which typically arises via waves of clonal and sub-clonal expansions [75, 146], operating over an adaptive fitness landscape in which Darwinian selection favors highly proliferative cell phenotypes which in turn drive rapid tumor growth [111, 65, 17]. The tumor environment should be viewed as a complex Darwinian adaptive eco-system consisting of cell types which have evolved over many years [126]. As a result, all but the most well designed and tailored therapeutic strategies often deliver disappointing outcomes and potentially introduce a potent new source of selective pressure for the proliferation of variant cells which develop an enhanced ability to resist future therapeutic assaults [93, 54, 106, 105, 74]. The prospects for influencing and controlling such a system are likeliest at the emerging early stages of tumor development when the cell population has not yet been selected for growth and survival, and the tumor size is small. But by the time a typical tumor becomes clinically detectable (often after several years of growth), it already contains $O(10^8)$ or more malignant cells (and potentially occupies a volume of $1-2$ $\mathtt{mm}^3$), some of which may have entered the blood circulation [65]. Since there is very little human data available in this early subclinical stage of tumor development, computational models can serve as a useful surrogate in this critical developmental stage which clearly influences and determines many important emergent features of the tumor at later stages.

Our goals in this paper are to describe a mathematical model for stochastic cell kinetics in the beginning stages of tumor development (from a single malignant cell) that includes cell reproduction and death, mutations, evolution, and the subsequent emergence of genetic heterogeneity well documented in many soft-tissue tumors [133, 123, 55, 2, 85, 173, 71, 186, 86]. The model is a computational one, driven by a stochastic Moran (birth-death) process with a finite cell population, in which birth-death rates are functions of cell fitness. The fitness is determined by the cell's numerical genome in the form of a four-digit binary string capable of undergoing point mutational dynamics with one digit in the string flipping values stochastically. The corresponding numerical value of the binary string determines whether the cell is healthy (low-fitness) or cancerous (high fitness). These two classes of cells compete against each other at each birth-death event, with fitness calculated according to the payoff matrix associated with the prisoner's dilemma evolutionary zero-sum game [179, 143, 201, 10]. The healthy cells play the role of cooperators, while the cancer cells play

the role of defectors [143, 10]. Our goal is to understand how the model parameters: passenger (m_p) and driver mutation rates (m_d), selection strength (w), birth and death rates, affect tumor growth characteristics, such as tumor growth rates, fixation probabilities of malignant and healthy cell types, saturation rates of cancer cells, and the emergence of genetic heterogeneity in a tumor at later stages of development when the tumor is clinically detectable.

An important outcome of the model is that growth of the cancer cell population is directly influenced by the intratumor heterogeneity (represented as the distribution of the 4-digit binary strings throughout the cell population), with high heterogeneity driving more rapid growth. The connection between heterogeneity and growth has been discussed in the literature [100, 112, 3, 209, 7, 86]. We quantify heterogeneity in a tumor using tools from information theory [39, 134], as well as quantitative analysis of phylogenetic trees associated with clonal and sub-clonal expansions [170, 186] in the developing tumor. Because our numerical simulations are carried out from initial conditions corresponding to a homogeneous population of healthy cells (0000) all the way to a saturated population of cancer cells, we can use the model to test basic dose and scheduling strategies [195, 1] at the very early stages of tumor development in the subclinical range, well before a tumor would be clinically detectable by current technology. Our point of view is that this emerging subclinical tumor should be more amenable (and potentially vulnerable) to a well planned therapeutic assault than a more mature tumor comprised (on average) of larger numbers of cells with more aggressive proliferative capabilities (having undergone generations of selection), that are potentially in the early stages of migration to other organs. More complex features that might influence early stage dynamics, like human-immune response [3] and the tumor microenvironment [117] are not included in this model in order to keep things as simple and clear as possible.

9.3 Description of the model

The ingredients in our model includes a stochastic birth-death process that is the engine which drives tumor growth, with heritable mutations operating over a fitness landscape so that natural selection can play out over many cell division timescales. Genetic mutations (point mutations) are modeled using a four-digit binary string of information that each cell carries with it.[2] This simple sequence divides the cells into 16 different "genotypes", ranging from 0000 up to 1111, and this information is passed on to the daughter cell during a birth event. The birth-death replacement process is based on a fitness function defined in terms of interactions quantified by the prisoner's dilemma payoff matrix which operates on two general classes of cells: healthy (the cooperators), and cancerous (the defectors). Natural selection acts on each generation of the

[2]To be clear, the four digit sequence is not meant as a bare-bones representation of the full human genome, but as a simple representation of the *relevant differences* in genetic information contained in different cells, allowing us to course-grain the cells into 16 different categories based on their genetic/epigenetic profiles.

cell population as the computational simulation proceeds on a cell division timescale. In this version of the model we typically simulate up to $O(10^{11})$ cell divisions, so our mutation rates are chosen to be relatively high to accommodate these somewhat modest timescales. See [192] for discussions on mutation rates in cancer.

9.3.1 The Moran birth-death process

The stochastic engine [165] that drives tumor growth in our model is a finite cell-based Moran process consisting of a population of N cells, divided into two sub-populations consisting of i cancer cells, and $N - i$ healthy cells. In all of our simulations, N is large enough so that there is not a significant difference between the results from our finite-cell model and the (deterministic) replicator equation approach for infinite populations, a connection that is discussed in detail in [196]. At each time-step in the simulation, one cell is chosen for reproduction and one cell is chosen for elimination. The cells are chosen randomly, based on their prevalence in the population pool which, in turn, is weighted by the fitness function based on a chosen payoff matrix. The probability of choosing a cancer cell at any given step is i/N, while the probability of choosing a healthy cell is $(N - i)/N$. As it unfolds, the process is a stochastic birth-death process where the total population size, N, stays constant and the number of cancer cells in the population, i, is the stochastic state variable. At any given step, the probability of transitioning from i cancer cells to j cancer cells is denoted P_{ij} , with $j = i + 1$ or $j = i - 1$. These probabilities are determined by the birth/death rates associated with the cancer cell population, which in turn are determined by a cell population fitness function. Each cell carries with it a binary string in the form of a four digit binary sequence from 0000 up to 1111. This defines 16 different cell types, which are course-grained into two groups: healthy cells (0000 - 1010), and cancer cells (1011-1111). These two sub-populations interact at each birth-death time-step with fitness defined in terms of the prisoner's dilemma payoff matrix. The algorithmic details are shown in the appendix Figure 67. To set the stage for more complex simulations, Figure 58 shows the result of a stochastic simulation (depicting i) driven by the Moran process alone, with no mutations, and no selection. Figure 58 shows three different simulations, one leading to the elimination of all cancer cells via random drift (red), another fluctuates between a mixed cell population after 10,000 cell divisions (yellow), and a third leading to fixation of the cancer cell population (blue) after around 5000 cell divisions. The average of 25 stochastic simulations is also plotted (note that the average will converge to half cancer cells and half healthy cells by the law of large numbers). The mean time to fixation of the cancer cell population which starts with 'i' cells in this simple setting (no mutations, no selection) is given by

$$k = N\left[\sum_{j=1}^{i}\frac{N-i}{N-j} + \sum_{j=i+1}^{N-1}\frac{i}{j}\right]. \tag{125}$$

With no mechanism for natural selection, there is no shape to the growth curves.

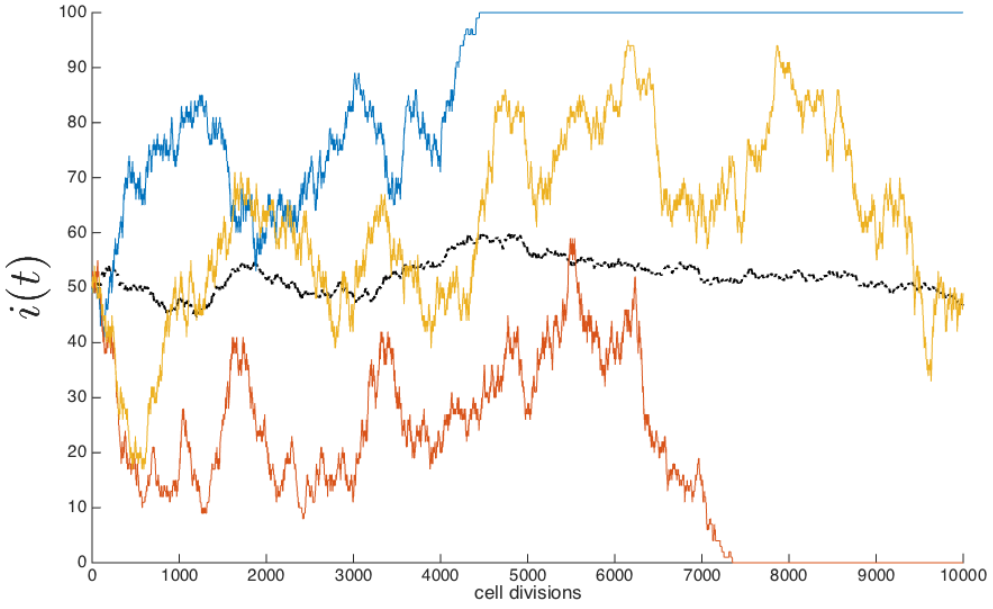


Figure 58: **Stochastic Moran birth-death process** — Cancer cell population, $i(t)$, during three stochastic simulations of the Moran birth-death process in a population of 100 cells and an initial condition of $i = 50$ cells. The blue curve leads to fixation of the cancer cell population, the red curve leads to elimination of the cancer cell population, and the yellow curve remains fluctuating in a mixed population of cells after 10,000 cell divisions. An average of 25 stochastic simulations (black dashed line) is also plotted.

9.3.2 The prisoner's dilemma payoff matrix

To introduce the effect of selection which will regulate cell birth and death rates, we use the prisoner's dilemma evolutionary game in which two players compete against each other for the best payoff. Each has to decide whether to cooperate (healthy cell) or defect (cancer cell) and each receives a payoff determined from the prisoner's dilemma payoff matrix[3], A:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}. \tag{126}$$

[3]What defines a prisoner's dilemma matrix are the inequalities $c > a > d > b$. The chosen values in (126) are relatively standard, but not unique. More discussion of why the prisoner's dilemma matrix, which models the evolution of defection, is a useful paradigm for cancer can be found in [211] and some of the references therein.

The essence of the prisoner's dilemma game is the two players compete against each other, and each has to decide what best strategy to adopt in order to maximize their payoff. This 2 x 2 matrix assigns the payoff (e.g. reward) to each player on each interaction. My options, as a strategy or, equivalently, as a cell type, are listed along the rows, with row 1 associated with my possible choice to cooperate, or equivalently my cell type being healthy, and row 2 associated with my possible choice to defect, or equivalently my cell type being cancerous. Your options are listed down the columns, with column 1 associated with your choice to cooperate (or you being a healthy cell), and column 2 associated with your choice to defect (or you being a cancer cell). The analysis of a rational player in a prisoner's dilemma game runs as follows. I do not know what strategy you will choose, but suppose you choose to cooperate (column 1). In that case, I am better off defecting (row 2) since I receive a payoff of 5 instead of 3 (if I also cooperate). Suppose instead you choose to defect (column 2). In that case, I am also better off defecting (row 2) since I receive a payoff of 1 instead of 0 (if I were to have cooperated). Therefore, *no matter what you choose, I am better off (from a pure payoff point of view) if I defect*. What makes this game such a useful paradigm for strategic interactions ranging from economics, political science, biology [211], and even psychology [143] is the following additional observation. *You will analyze the game in exactly the same way I did (just switch the roles of me and you in the previous rational analysis), so you will also decide to defect no matter what I do*. The upshot if we both defect is that we will each receive a payoff of 1, instead of each receiving a payoff of 3 if we had both chosen to cooperate. The defect-defect combination is a Nash equilibrium [179], and yet it is sub-optimal for both players and for the system as a whole. Rational thought rules out the cooperate-cooperate combination which would be better for each player (3 points each) and for both players combined (6 points). In fact, the Nash equilibrium strategy of defect-defect is the worst possible system wide choice, yielding a total payoff of 2 points, compared to the cooperate-defect or defect-cooperate combination, which yields a total payoff of 5 points, or the best system-wide strategy of cooperate-cooperate yielding a total payoff of 6 points.

The game becomes even more interesting if it is played repeatedly [179, 143, 201, 10], with each player allowed to decide what strategy to use on each interaction so as to accumulate a higher payoff than the competition over a sequence of N games. In order to analyze this kind of an evolving set-up, a fitness function must be introduced based on the payoff matrix A. Let us now switch our terminology so that the relevance to tumor cell kinetics becomes clear. In this case, we randomly select pairs of cells out of the total population at each step, and subject them to a birth-death process, basing our birth rates and death rates on the prisoner's dilemma payoff matrix. Thus, in our context, it is not the strategies that evolve, as cells cannot change type based on strategy (only based on mutations), but the prevalence of each cell type in the population is evolving, with the winner identified as the sub-type that first reaches fixation in the population. As the populations evolve, the fitness of the two competing sub-populations can be tracked, as

well as the overall fitness of the combined total population of cells.

9.3.3 The fitness landscape

Let us start by laying out the various probabilities of pairs of cells interacting and clearly defining payoffs when there are i cancer cells, and $N-i$ healthy cells in the population. The probability that a healthy cell interacts with another healthy cell is given by $(N-i-1)/(N-1)$, whereas the probability that a healthy cell interacts with a cancer cell is $i/(N-1)$. The probability that a cancer cell interacts with a healthy cell is $(N-i)/(N-1)$, whereas the probability that a cancer cell interacts with another cancer cell is $(i-1)/(N-1)$. The payoffs associated with the healthy cells and cancer cells, obtained by weighting the payoff matrix values with appropriate probabilities, are given by (following notation in [196]):

$$\pi^H = \frac{3(N-i-1)+0i}{N-1}, \tag{127}$$

$$\pi^C = \frac{5(N-i)+1(i-1)}{N-1}. \tag{128}$$

This gives rise to the average payoff associated with the population of cells:

$$\langle \pi \rangle = \frac{\pi^H(N-i)+\pi^C(i)}{N}. \tag{129}$$

Based on these formulas, we define the fitness of the healthy cells as:

$$f^H = 1 - w_H + w_H \pi^H, \tag{130}$$

and the fitness of the cancer cells as:

$$f^C = 1 - w_C + w_C \pi^C. \tag{131}$$

Here, (w_H, w_C) are 'selection strength' parameters, $0 \leq w_H \leq 1, 0 \leq w_C \leq 1$, that measure the strength of selection pressure on each of the population of cells. If $w_H = 0$, there is no natural selection acting on the healthy cell population and the dynamics is driven purely by the Moran process. When $w_H = 1$, the selection pressure on the healthy cell population is strongest and the prisoner's dilemma payoff matrix has maximum effect. Likewise for the parameter w_C and how it controls selection pressure in the cancer cell population. Since therapy imposes selection pressure on different sub-populations of cells, w_H and w_C are the two parameters we alter to administer simulated therapeutic responses. We discuss this in section §3.5.

The expected fitness of each of the sub-populations are:

$$\phi^H = \frac{N-i}{N} f^H, \tag{132}$$

$$\phi^C = \frac{i}{N} f^C, \tag{133}$$

with total expected fitness:

$$\phi = \phi_i^H + \phi^C. \tag{134}$$

From these formulas, we can define the transition probability of going from i to $i+1$ cancer cells on a given step:

$$P_{i,i+1} = \frac{if^C}{if^C + (N-i)f^H} \frac{N-i}{N}. \tag{135}$$

The first term represents that probability that a cancer cell is selected for reproduction (weighted by fitness), and a healthy cell is selected for death. Likewise, the transition probability of going from i to $i-1$ cancer cells on a given step is:

$$P_{i,i-1} = \frac{(N-i)f^H}{if^C + (N-i)f^H} \frac{i}{N}. \tag{136}$$

Here, the first term is the probability healthy cell is selected for reproduction (weighted by fitness), and a cancer cell is selected for death. The remaining transition probabilities are as follows:

$$P_{i,i} = 1 - P_{i,i+1} - P_{i,i-1}; \quad P_{0,0} = 1; \quad P_{N,N} = 1. \tag{137}$$

It is these simple formulas that drive the subsequent dynamics of the competing populations of cells and determine the emergent features of the forming tumor (cancer cell population). A typical set of simulations of the evolving fitness of the healthy cell population, ϕ_H, the cancer cell population ϕ_C, and the total fitness, ϕ, is shown in Figure 59 as the selection parameter varies from 0 to 1 ($w_H = w_C \equiv w$). As the population evolves, the fitness of the healthy cell population decreases, the fitness of the cancer cell population increases (sometimes reaching a maximum point), while the total population fitness decreases.

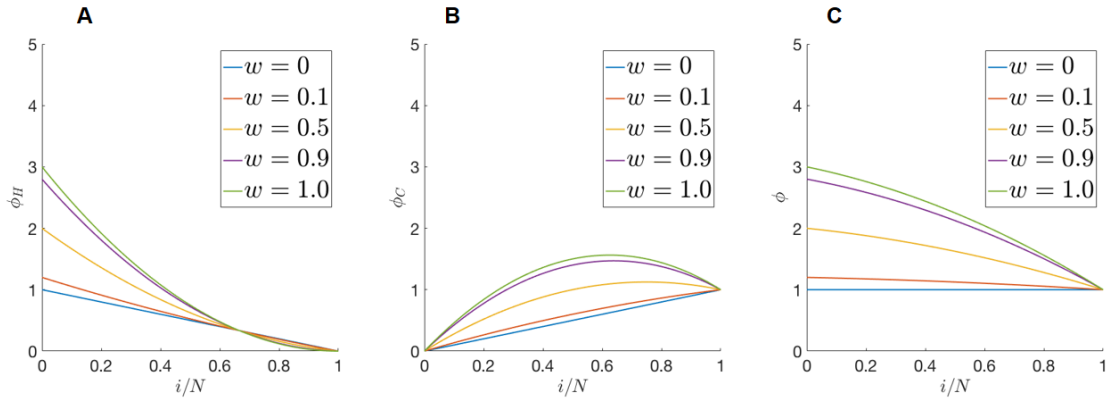


Figure 59: **Fitness as a function of the selection parameter** $w \equiv w_H \equiv w_C$ — (A) Monotonically decreasing fitness of healthy cell sub-population ϕ_H; (B) Fitness of cancer cell sub-population ϕ_C. Note that ϕ_C has a maximum at $i = \frac{N}{2} + \frac{(N-1)}{8w}$, which is between 0 and N for $w > \frac{1}{4}(1 - 1/N)$; (C) Monotonically decreasing fitness of the total population, ϕ.

9.3.4 Passenger and driver mutations

Two remaining parameters in our model are the passenger mutation rate, m_p and the driver mutation rate, m_d [192]. Passenger mutations confer no fitness advantage, hence m_p controls point mutations that act on the digit strings that define the 11 levels of healthy cells 0000-1010, and the 5 levels of cancer cells 1011-1111. A mutation diagram is shown in Figure 60 depicting all of the possible point mutation transitions at each step. Mutations that stay within either of those two ranges do not alter the cell fitness. On the other hand, the driver mutation parameter controls mutations that take a binary string from a healthy cell and, via a point mutation, alter it so that the string becomes a cancer cell[4]. A simple example would be a mutation that alters 1010 (healthy) to 1011 (cancer) by stochastically flipping the first digit from 0 to 1. The interested reader can consult the flow diagram in Figure 67 of the Appendix for more details of the algorithm. The full code is available from the authors upon request.

9.4 Results

Gompertzian growth arising from multicellular systems occurs in many settings with different physical and biological constraints acting in concert. Hence it appears as if this universal growth curve does not depend on specific physical mechanisms (e.g. oxygen diffusion, blood supply, tumor microenvironment, etc.) but more on multi-cellularity and the ability for populations of cells to assume a heterogeneous distribution of functional states, as was described most clearly in Kendal's 1985 paper [100] and documented clinically in

[4]In our simulations, we assume that driver mutations cannot revert to passenger mutations, i.e. once a cancer cell is born, it stays in that category. We do not know of any evidence in the literature that shows the reversion of a cancer cell to a healthy cell, nor is this particularly a focus of this manuscript.

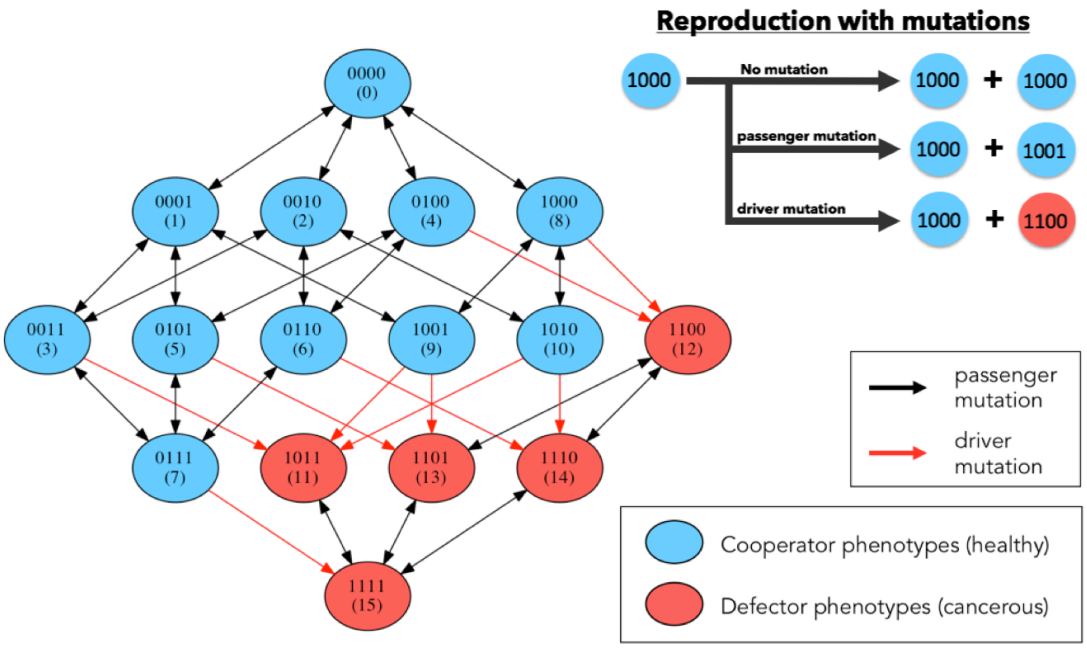


Figure 60: **Markov Point Mutation Diagram** — Left: diagram shows 16 genetic cell types based on 4-digit binary string and the effect of a point mutation on each cell type. Blue indicates healthy cell type (0000 — 1010), red indicates cancerous cell type (1011 — 1111). Black arrows indicate passenger mutations (healthy to healthy or cancer to cancer), red arrows indicate driver mutations (healthy to cancer). Top right: 3 scenarios may occur during the reproduction process: no mutation, passenger mutation, or driver mutation.

breast [1] and other tumor types. Alternative bio-mechanistic models of tumor growth at the cellular level have been developed (see [163, 42, 213, 117, 66, 102]) although do not generally include molecular information or evolutionary effects. Features of the Gompertzian growth curve defined by eqns (138), (139) allow us to clearly describe three distinct growth regimes, the earliest being subclinical and the most critical regime in which to influence future tumor kinetics, the second being the clinical regime where growth measurements are typically obtained [65], and the third being the lethal burden phase where growth saturates. The growth equation, (146), relates tumor heterogeneity to growth rates, and we quantify heterogeneity via the Shannon entropy [39, 134] of the cellular population. One of the main features of our evolutionary simulations is to show how it (i) leads to Gompertzian growth, (ii) how growth is driven by heterogeneity quantitated via Shannon entropy, (iii) how the initiation of heterogeneity and fitness can be tracked via dynamic phylogenetic trees, and (iv) how tumor kinetics can be influenced via therapeutic strategies that target heterogeneity best in earlier growth regimes. In keeping consistent with the notation of the Gompeterzian growth curve, we now represent the tumor growth as the proportion of cancer cells in the population, $n_G(t)$.

9.4.1 Gompertzian tumor growth and three growth regimes

The basic (top-down) equations giving rise to pure Gompertzian growth [63, 109, 194, 118] are the coupled equations:

$$\frac{dn_G}{dt} = \gamma n_G, \tag{138}$$

$$\frac{d\gamma}{dt} = -\alpha\gamma. \tag{139}$$

Here, is the proportion of growing cancer cells in the mixed population, which grows exponentially according to (138), but with a time-dependent growth rate which is exponentially decaying according to (139). It is straightforward to integrate (138) to obtain:

$$n_G(t) = N_0 \exp\left[\left(\frac{1}{t}\int_0^t \gamma dt\right)\cdot t\right]. \tag{140}$$

Then, (139) is solved with:

$$\gamma(t) = \gamma_0 \exp(-\alpha t). \tag{141}$$

Plugging (141) into (140) and integrating yields the Gompertzian curve:

$$n_G(t) = N_0 \exp\left[\frac{\gamma_0}{\alpha}\left(1-\exp(-\alpha t)\right)\right], \tag{142}$$

where in the long-time limit , the population saturates to the value

$$n_\infty = N_0 \exp(\gamma_0/\alpha), \tag{143}$$

which we normalize to one (without loss of generality). The key features of Gompertzian growth are shown in Figure 61. As the cancer cell proportion $n_G(t)$ increases (Figure 61A), there are three distinct growth regimes defined by the inflection point on the n_G growth curve (maximum of $\dot{n}_G$ and $d^2n_G/dt^2 = 0$), and the two inflection points on the growth-velocity curve $\dot{n}_G$ (maximum/minimum of $\ddot{n}_G$ and $d^3n_G/dt^3 = 0$). As shown in Figure 3.1(a), there are three points that divide the growth curve into four distinct regions. For convenience, and symmetry, we lump the second and third regions together and define three basic growth regimes:

- **Regime 1** *(Subclinical)*: Increasing velocity $\dot{n}_G$, increasing acceleration d^2n_G/dt^2. Cell population

and tumor volume grows at an exponential rate;

- **Regime 2** *(Clinical)*: In this regime, $\dot{n}_G$ reaches its maximum value. In the early part of the regime, $\dot{n}_G$ is increasing while d^2n_G/dt^2 decreases. In the later part of the regime, $\dot{n}_G$ is decreasing and d^2n_G/dt^2 becomes negative (deceleration). Growth rates are clinically typically measured as linear;
- **Regime 3** *(Saturation/Lethal)*: Decreasing tumor velocity $\dot{n}_G$ with decreasing deceleration. Growth rate rapidly slows towards full saturation of the cancer cell population.

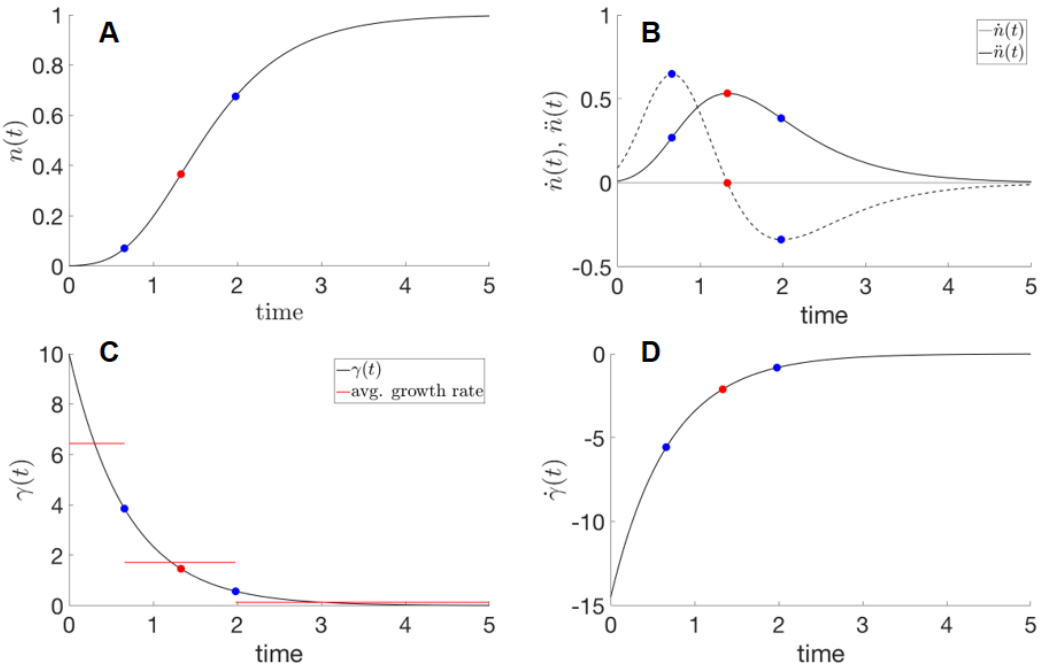


Figure 61: **Gompertzian equation** — Numerical simulation of the Gompertzian equation (138), (139) with parameters $N_0 = 0.001$, $\gamma_0 = 10$, and $\alpha = 0.2895$. The three regimes of tumor growth are demarcated by the blue dots in each subfigure, representing the maximum and minimum of the second-derivative. (A) Cancer cell proportion, n(t), over time; (B) First- and second-derivatives of the tumor growth curve; (C) Growth rate, $\gamma(t)$, over time, with the average growth rate in regimes 1, 2, 3 plotted in red; (D) First derivative of growth rate.

Regime 1, generally speaking, is the subclinical growth regime where the developing tumor has substantially fewer than 10^8 malignant cells with a tumor size smaller than 1 or 2 $\mathtt{mm}^3$. Typically, the clinically measurable regime is Regime 2, while the lethal stage when the tumor saturates is associated with Regime 3. In reality, the boundaries of these regimes are, of course, not sharp and depend on tumor type and location which influence detectability. But the clarity of the pure Gompertzian curve gives a useful framework which delineates the three distinct growth regimes based on clear principles associated with growth, velocity, and

acceleration. The growth rate curve is shown in Figure 61C, with its derivatives shown in Figure 61D. It is most instructive to show the average growth rates defined in each of the three regimes, also shown in the Figure 61C. The average growth rate in the time interval from t_1 to t_2 is defined as:

$$\gamma_{ave} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \gamma(t) dt \tag{144}$$

The subclinical regime 1 has the highest average growth, whereas regime 2, where tumor growth is typically measured, average growth rates are lower, followed by the lowest average growth in the clinically lethal regime 3. This implies that clinically measured growth rates typically *underestimate* growth rates that preceded it in the subclinical stage. It also implies that linear extrapolation back from clinically measured growth rates to estimate tumor initiation times (see [111, 65, 17]) will systematically overestimate the amount of time the tumor has been developing before being measured. While this might generally be seen as good news (since the cancer initiation event was more recent than estimated via linear extrapolation), it also gives the clinician a shorter window of time in which to act.

9.4.2 Heterogeneity and growth via statistical mechanics

Kendal [100] lays out a clear argument of how this growth curve arises from a purely statistical mechanics point of view. In a nutshell, his argument can be explained by considering a population of n cells, let the jth cell ($j = 1, 2, 3, \ldots, n$) have the potential to assume one of q_j possible states. The number of combinations of states possible within the population, P, can be thought of as a measure of intra-neoplastic diversity:

$$P = q_1 q_2 q_3 ... q_n, \tag{145}$$

and is related to the growth rate of a tumor via the equation:

$$\frac{dn}{dt} = \alpha \log P, \tag{146}$$

where $n(t)$ is the number of cells capable of proliferation at a given time t and α is a parameter that sets the timescale of growth [5]. There are two basic cases to consider. First, suppose the cells have no interaction at all, say in the earliest stages of tumor development, and let each of the n cells have the ability to assume

[5] Kendal's formulation [100] assumes a cell population made up of three sub-groups: (1) proliferative cells; (2) nonproliferative and nonclonogentic cells; (3) nonproliferative but clonogenic cells, with an assumption that the neoplasm's growth rate is influenced by the proportion of proliferating to nonproliferating cells and an expression of each clone's growth potential. The log is chosen based on the fact that heterogeneity is measured as the multiplicative combination of achievable states in the tumor, and the requirement that $G(P_1 \cdot P_2) = G(P_1) + G(P_2)$ for any two sub-populations P_1, P_2 and growth function G. The discussion of the relationship between tumor heterogeneity and growth is an ongoing topic in the current literature [133, 123, 85, 71, 186].

m_d	m_p	t_{Emax}	t_{SAT}	n_d	n_p	$\gamma_{ave,1}$	$\gamma_{ave,2}$	$\gamma_{ave,3}$
0.4	0.1	5.50e+5	1.830e+6	1.289e+4	4.68e+4	3.14e-5	3.68e-6	1.448e-7
0.3	0.2	4.88e+5	1.753e+6	1.682e+4	8.26e+4	4.04e-5	4.31e-6	1.677e-7
0.2	0.3	4.85e+5	1.761e+6	1.715e+4	1.230e+4	3.86e-5	4.41e-6	1.729e-7
0.1	0.4	5.40e+5	1.426e+6	1.362e+4	1.836e+4	3.04e-5	3.81e-6	1.658e-7

Table 22: m_d: driver mutation rate; m_p : passenger mutation rate; t_{Emax}: time to maximum entropy; t_{SAT}: time to saturation; n_d: number of driver mutations; n_p: number of passenger mutations; $\gamma_{ave,1}$: average growth rate in regime 1; $\gamma_{ave,2}$: average growth rate in regime 2; $\gamma_{ave,3}$: average growth rate in regime 3.

one of m possible states. Then, $P = m^n$, and the growth equation becomes

$$\frac{dn}{dt} = \alpha n \log m = (\alpha \log m)n. \tag{147}$$

The solution to this equation is the exponentially growing population:

$$n(t) = N_0 \exp((\alpha \log m)t). \tag{148}$$

Thus, early stage development is characterized by exponential growth (regime 1), with a growth rate proportional to the log of the number of assumable states of the cells comprising the tumor population. This stage is characterized by the Gompertzian curve shown in Figure 61A to the left of the first blue dot, in regime 1. Contrast this with later stages of tumor growth, when the sub-populations of cells communicate and influence each other's growth characteristics, either via competition, or cooperation (regime 3) within the tumor microenvironment. In effect, this will constrain (reduce) the number of assumable states of each cell, since the population is effectively coupled. In the extreme, suppose $P = m^n/n^n$. In other words, suppose P is now inversely related to the total number of possible intercellular interactions. Inserting this into (146) yields

$$\frac{dn}{dt} = \alpha \log \left((\frac{m}{n})^n \right) = \alpha n \left[\log m - \log n \right]. \tag{149}$$

The solution to this equation is exactly the Gompertzian growth curve (142) and accounts for regimes 2 and 3 previously discussed in which tumor growth slows down. The growth equation (146) which relates cancer cell population growth to tumor heterogeneity is capable of producing a family of growth curves, depending on details of intercellular coupling, which of course is influenced by details of the biological and physical constraints influencing the tumor microenvironment. Thus, the growth equation (146) has the ability to produce different detailed shapes based on assumptions associated with intercellular coupling. Table 22 shows the average growth rates in the three regimes as a function of the key parameters in the model.

9.4.3 Quantitative measures of tumor heterogeneity and growth

For our purposes, we measure heterogeneity using the Shannon entropy from information theory [39]:

$$E(t) = -\sum_{i=1}^{N} p_i \log_2 p_i, \tag{150}$$

(here, log is defined as base 2). The probability p_i measures the proportion of cells of type i, with $i = 1, ..., 16$ representing the distribution of binary strings ranging from 0000 to 1111. We then course-grain this distribution further so that cells having strings ranging from 0000 up to 1010 are called "healthy", while those ranging from 1011 to 1111 are "cancerous"[6]. The growth equation (146) then becomes

$$\frac{dn_E}{dt} = \alpha E(t). \tag{151}$$

It follows from (151) that the cancer cell proportion $n_E(t)$ can be written in terms of entropy as:

$$n_E(t) = \alpha \int_0^t E(t)dt. \tag{152}$$

The panel in Figure 62 shows the results from our cell-based simulations. Figure 62A shows the Gompertzian curve associated with the proportion of cancer cells in the population, while Figure 62B shows the velocity and accelerations associated with growth, and can be compared with Figure 61B. In Figure 62C we show the entropy during a typical simulation, marking the maximum entropy point which peaks relatively early in the simulation before the entropy returns back down to zero, reflecting the fact that cancer cells have reached fixation and have saturated the population. Figure 62D shows the fitness of the cancer cell sub-population, healthy cell sub-population, and the overall tumor fitness ($w_H = w_C \equiv w = 0.5$). As a typical simulation proceeds, the cancer cell sub-population fitness increases, the healthy cell sub-population fitness decreases, while the overall tumor fitness decreases. Figure 62E, 62F shows the Gompertzian growth curves as the selection pressure increases (Figure 62E) and as the mutation rate increases (Figure 62F). High values for either of these parameters leads to a very steep growth curve, as is expected.

Figure 63 shows the growth curves linearly extrapolated back to give a prediction of when the first driver mutation occurred that initiated tumor growth. The growth rates from regime 2 (linear regime) are used to extrapolate back to the initiation event. Since the actual growth rate in regime 1 is much faster than linear, the linear extrapolation extends the event too far back in time as compared to when the event actually occurred. The inset of Figure 63 shows histograms of the average growth rates in each of the three regimes as a function of the mutation rate m (here, we take $m_p = m_d = m$).

[6] Our results are relatively insensitive to where we draw the dividing line between healthy and cancerous.

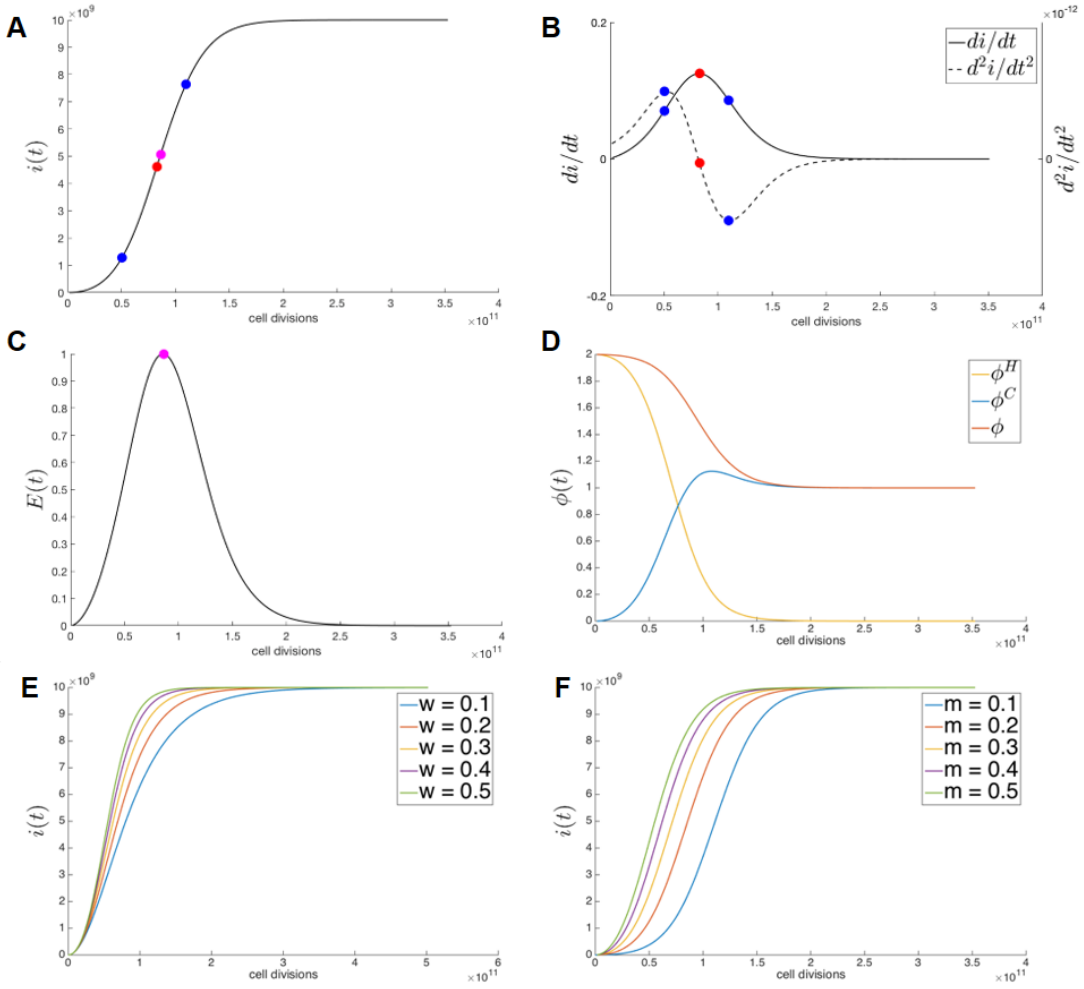


Figure 62: **Moran birth-death process with selection** — (A) Cancer cell population, $i(t)$ ($w = 0.5$, $m = 0.2$, $N = 10^{10}$) plotted with a spline curve connecting 200 data points from a single stochastic simulation; (B) First- and second-derivatives of the tumor growth curve in (A) are plotted with maximum and minimum of second-derivative indicated (blue); (C) Entropy of the cell population from eqn. (150) as it relates to the growth equation (151); (D) Fitness of healthy cell population and cancer cell population and total fitness as defined by eqns. (132), (133), (134); (E) Simulations of cancer cell population, $i(t)$, for a range of selection parameter values; (F) Simulations of cancer cell population, $i(t)$, for a range of mutation rate values.

A typical stochastic simulation showing the evolution of all 16 possible cell types is shown in Figure 64. We also show $E(t)$, where entropy is computed using the most extreme course-grained two-state system comprised of the two sub-populations of healthy cells and cancer cells. We compare in Figure 64 the Gompertzian growth curve (eqn. (142)) and the corresponding curve obtained from eqn. (152) to the stochastic simulation and the agreement is excellent. Likewise, we also show a comparison of dn/dt with eqn. (151) and eqn. (138) with $E(t)$ normalized so that limiting values match the stochastic simulation,

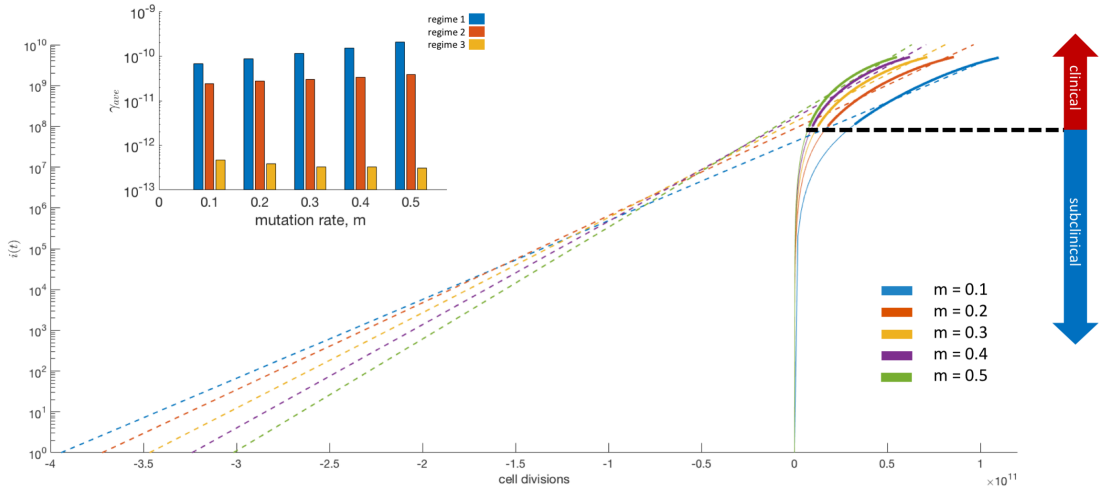


Figure 63: **Tumor initiation prediction** — Five sample stochastic simulations of tumor growth ($N = 10^{10}$ cells, $w = 0.5$, $m = 0.1, 0.2, 0.3, 0.4, 0.5$) plotted on a log-linear graph where the model output ($i(t)$, solid lines) is fit in the clinical regime (greater than 10^8 cells) using an exponential growth equation and extrapolated backwards in simulation time (dashed lines). The inset bar graph shows the average growth rate in each regime.

and the agreement is also excellent. In the beginning, entropy is zero, since the population consists purely of healthy cells, and in the end of the simulation, entropy is again zero as the population consists purely of cancer cells. Entropy peaks somewhere early in the simulation when the mixture of cell types is equally distributed over cancer and healthy types. It is this intermediate but important heterogeneously distributed state that is the key driver of growth, as is clear from eqn. (151).

9.4.4 Dynamic phylogenetic trees and evolution of fitness

To track the initiation of cellular heterogeneity from an initially homogeneous state, we follow all of the mutations that take place during the course of a simulation, and organize this in the form of a phylogenetic tree in Figure 65 showing the typical size of the genotypic space and the evolution of the genotypic landscape. As the simulation proceeds, the phylogenetic tree dynamically branches out into an increasingly complex structure, with fitness characteristics color coded in Figure 65A. We also show the bins associated with each of the 16 cell types, the number of cancer cells $i(t)$, and the entropy associated with the sub-population of cell types as a simulation proceeds, in Figure 65B. Knowing exactly the types of cells comprising the tumor at any given time allows us to target cell distributions for simulated therapies to test different strategies, which we describe next.

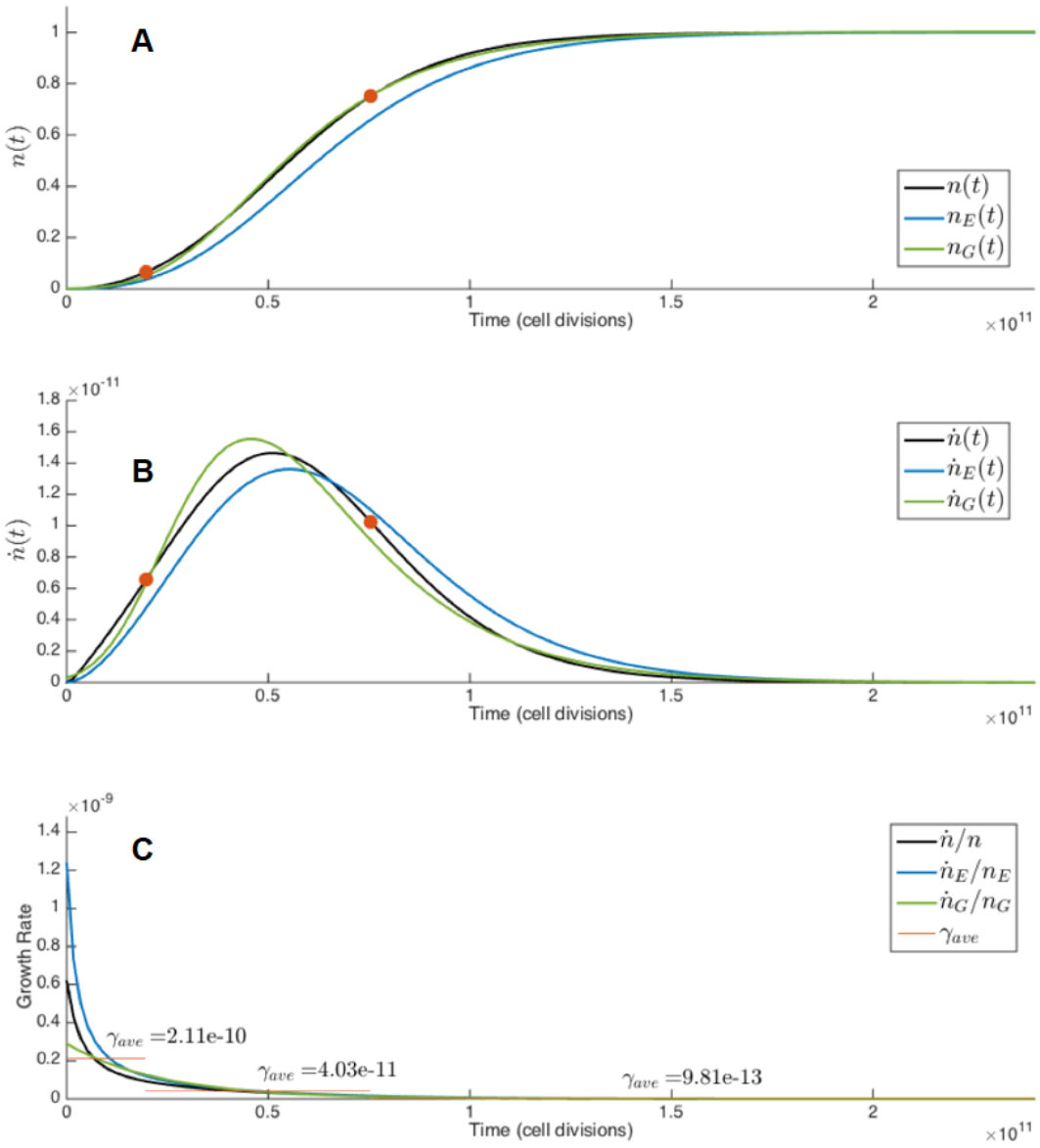


Figure 64: **Comparison of stochastic Moran birth-death process, Gompertzian, and Shannon entropy growth curves** — (a) A single stochastic simulation ($N = 10^{10}$ cells, $m = 0.5$, $w = 0.5$, $m_p = m_d = 0.25$) growth curve, $n(t)$, compared with the Gompertzian growth curve, $n_G(t)$, eqn. (142), and Shannon entropy growth curve, $n_E(t)$, eqn. (152). Growth curves $n_G(t)$ and $n_E(t)$ are normalized to equal one in the limit; (b) Comparison of first-derivatives of $n(t)$, $n_E(t)$, $n_G(t)$; (c) Comparison of growth rates associated with $n(t)$, $n_E(t)$, $n_G(t)$, with average growth rates of $n(t)$ plotted for each regime, eqn. (144).

9.4.5 A comparison of early vs. late therapy

In Figure 66 we show the results from asking the simple question of how early therapy (administered in regime 1) compares with therapy in the middle stages of tumor development (regime 2), or in the later

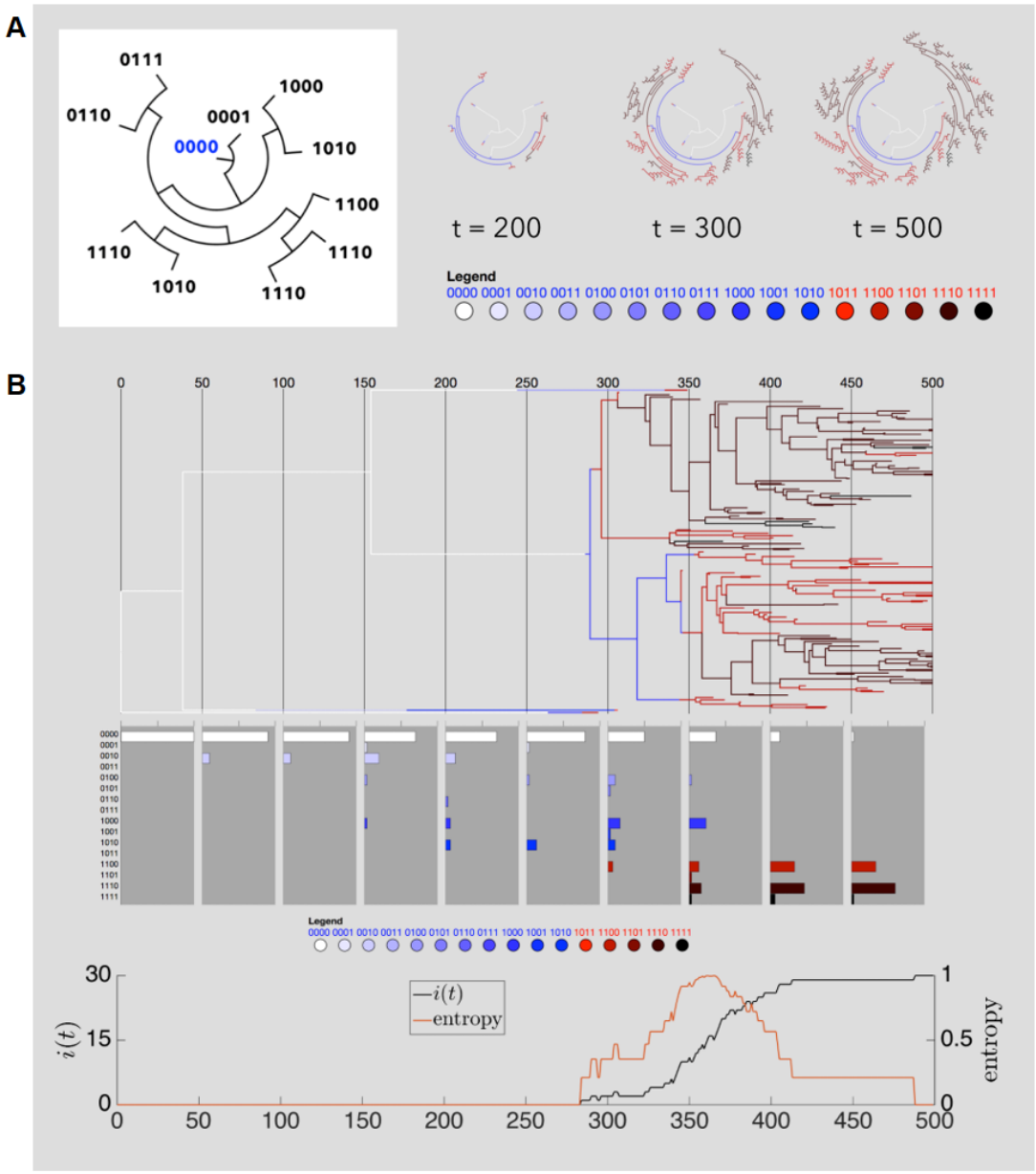


Figure 65: **Emergence of genetic heterogeneity** — (A) Left: sample dendritic phylogenetic tree tracking point mutations as time extends radially. Right: three snapshots in time of a dendritic tree in a simulation of 30 cells with strong selection ($w = 1$, $m_p = 0.1$, $m_d = 0.2$). Pathways are color coded to indicate genetic cell type; (B) Linear phylogenetic tree of the same stochastic simulation shown in (a) along with histogram plots of the distribution of genetic cell types and a plot of the cancer cell population $i(t)$ and entropy.

stages of development (regime 3). Eqns (135), (136) are the governing equations controlling birth/death rates of the cancer cell, healthy cell sub-populations as natural selection plays out. Since the proliferation of cancer cells can be thought of as an imbalance of selection pressures on the competing sub-populations in favor of the cancer sub-population, the goal of any therapeutic intervention is to alter this complex imbalance

in favor of the healthy cell sub-population. We implement this by adjusting the selection pressure parameters (w_H, w_C) in the formulas (130), (131). In particular, when therapy is 'on', we choose $w_C = 0$, and $w_H = 1$, tilting the selection pressure in favor of the healthy cell sub-population. When therapy is 'off', the two parameters return to their original baseline values, which here we take as $w_H = 0.1$, $w_C = 0.1$. Figure 66 depicts the proportion of cancer cells in the population both in the absence of therapy, and when therapy is administered. As a comparative tool, in each case, we administer the therapy until a fixed number of cancer cells remains (in each case, we take this threshold number to be 25 cancer cells), and we compare the amount of time, Δt, it takes to achieve this low level. The figure clearly shows $\Delta t_1 < \Delta t_2 < \Delta t_3 < \Delta t_4$, while if therapy is administered too late, as in Δt_5, the low threshold is never achieved. The simulations show that a shorter therapeutic time-period is needed if administered earlier to gain the same level of success. The topic of how best to optimize computational therapies is complex, and these simulations are only meant as a confirmation and quantification of how early stage therapy is more effective than late stage therapy.

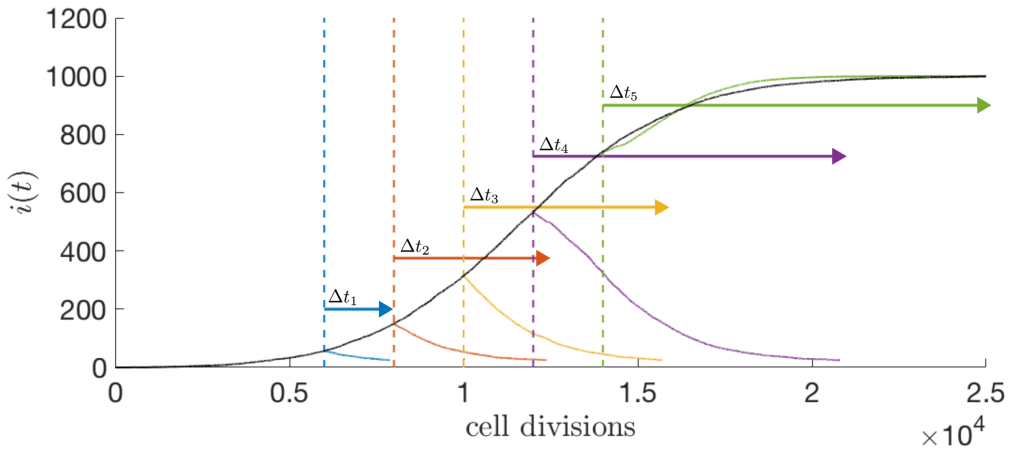


Figure 66: **Simulated therapy** — An average of 25 stochastic simulations ($N = 10^3$ cells, $w = 0.5$, $m = 0.1$) where therapy ($w_H = 1$, $w_C = 0$) is administered at different time points ($t = 6000, 8000, 10000, 12000, 14000$ cell divisions) until all cancer cells are eliminated below a small threshold value (25 cells). Time required (Δt) for tumor elimination increases as the tumor volume increases (i.e. $\Delta t_1 < \Delta t_2 < \Delta t_3 < \Delta t_4$, blue, red, yellow, purple arrows respectively), until, at later simulation time points, therapy is unable to regress tumor size (Δt_5, green arrow).

9.5 Discussion

To summarize the main points forming the framework of our model:

(i) A tumor is a complex Darwinian ecosystem of competing cells operating on an adaptive fitness landscape driven by mutational dynamics and shaped by evolutionary pressures;

(ii) The basic competitors in an evolutionary game theory model of tumor development are cell populations with a broad distribution of fitness characteristics course grained into two types: healthy cells (cooperators) and cancer cells (defectors). Each of these cell sub-populations attempts to maximize its own fitness;

(iii) Cell fitness is associated with reproductive prowess and in this respect, healthy cells are less fit than cancer cells;

(iv) Primary tumors initiate from a single malignant cell that has undergone the appropriate mutational steps and subsequently undergoes clonal and sub-clonal expansion. Polyclonality and heterogeneity are thus seen as emergent features of tumor development;

(v) Parameters and distributions measured in the detectable range of tumor growth, such as tumor growth rates and fixation probabilities, are emergent features that have developed from a monoclonal state via cell kinetics and evolutionary development taking place in the subclinical regime;

(vi) Tumor growth is driven by molecular heterogeneity of the cell population comprising the tumor and reflected in the growth equation (146);

(vii) Tumor cell populations are more amenable to therapeutic strategies in the early stages of development, before selection for growth and survival have shaped the environment.

We believe the simple evolutionary model described in this paper, driven by a Moran process and shaped by heritable mutations with a fitness landscape based on the prisoner's dilemma evolutionary game is useful in helping to understand early stage tumor growth and how it is influenced by the interplay of a few select small number of key parameters. When a malignant tumor cell population has already exceeded $O(10^8 - 10^{10})$ cells, some of which may have entered the circulation or lymphatics and migrated to other sites, the opportunity to control or even shape future events may be limited. Attacking tumor heterogeneity as soon as it develops seems to be a useful strategy, particularly if heterogeneity is the driver of growth, as in eqn. (146). Whether these concepts can be developed in the more general context when cell dissemination to other sites is included in the model, and then translated into actionable clinical strategies is a challenge for the future.

9.6 Appendix

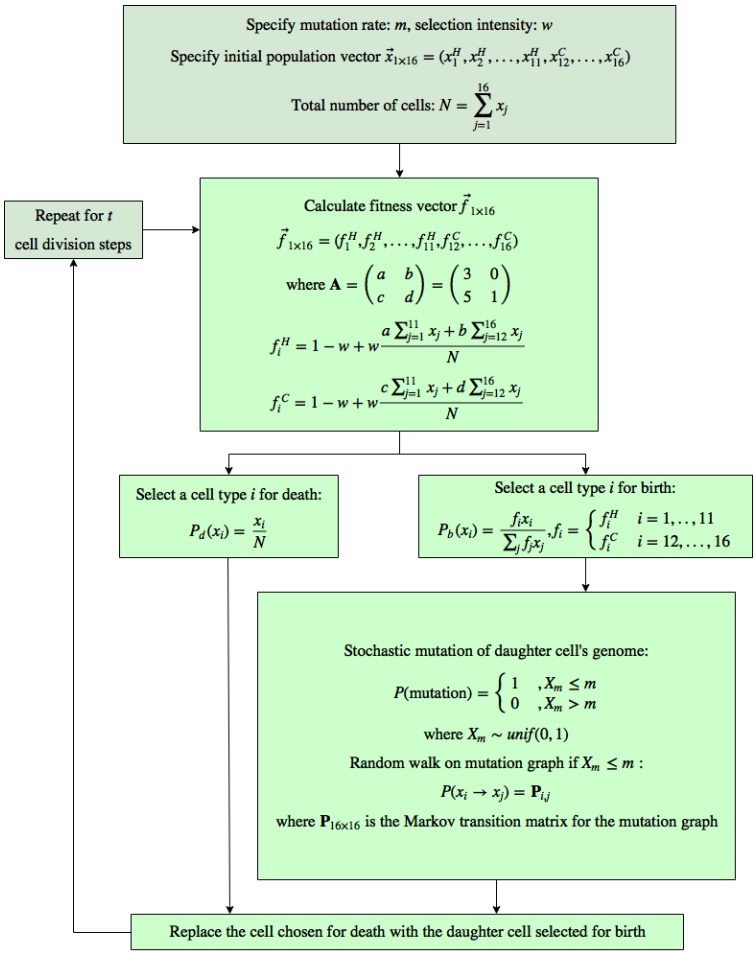


Figure 67: **A flow chart of the Moran process with selection and mutation algorithm** — Box 1: mutation rate m (where $m = m_p + m_d$), selection pressure w and the initial state vector x containing N total cells are the inputs for a simulation. Box 2: the prisoner's dilemma game ($a = 3$; $b = 0$; $c = 5$; $d = 1$) is used to calculate the fitness of each healthy and cancer cell type, which is a function of the payoff values and the state vector, x. Box 3, 4: a single cell is chosen for death according to the relative proportion of the cell type in the cell population. Simultaneously, a single cell is selected for birth according to the relative proportion, weighted by cell fitness. Box 5: During the replication process, the daughter cell inherits a replica of the parent cell's genetic string, with errors occurring at a rate of m. A single bit of the daughter cell's genetic string may flip during each cell division. The possible mutations can be thought of as a single step random walk on the Markov diagram shown in Figure 60.

10 Publications

In-progress

Submitted

Completed

10.1 First author publications

Journal article: Low-dimensional characterization of human performance of cancer patients using motion data [81]

Conference paper: Mining Human Mobility to Quantify Performance Status [136]

Journal article: Machine learning models for long-term outcome prediction in bladder cancer

Journal article: Feature and model based characterization of spatial and temporal organoid dynamics

Journal article: Kinematics from clinical exercises differentiate patients by unexpected hospitalizations and physical activity

Journal article: Temporal differentiation of phenotypes using a Hidden Markov model of organoid shape dynamics

10.2 Second author publications

Journal article: The prisoner's dilemma as a cancer model [211]

Journal article: An evolutionary model of tumor cell kinetics and the emergence of molecular heterogeneity driving Gompertzian growth [210]

Journal article: Pathways of metastatic bladder cancer from a longitudinal patient data set

11 References

[1] LH Abbott and F Michor. "Mathematical models of targeted cancer therapy". In: *British Journal of Cancer* 95.9 (2006), pp. 1136–1141.

[2] Robert G Abbott, Stephanie Forrest, and Kenneth J Pienta. "Simulating the Hallmarks of Cancer". In: *Artificial Life* 12.4 (2006), pp. 617–634.

[3] John Adam and Nicola Bellomo. *A survey of models for tumor-immune system dynamics*. Springer Science & Business Media, 2012.

[4] "Age Standards for Survival". In: (2018). URL: `https://seer.cancer.gov/stdpopulations/survival.html`.

[5] Alkiviadis G Akritas and Gennadi I Malaschonok. "Applications of singular-value decomposition (SVD)". In: *Mathematics and computers in simulation* 67.1-2 (2004), pp. 15–31.

[6] M Ando et al. "Prognostic value of performance status assessed by patients themselves, nurses, and oncologists in advanced non-small cell lung cancer". In: *British journal of cancer* 85.11 (2001), p. 1634.

[7] Robyn P Araujo and DL Sean McElwain. "A history of the study of solid tumour growth: the contribution of mathematical modelling". In: *Bulletin of mathematical biology* 66.5 (2004), p. 1039.

[8] Camille Stephan-Otto Attolini and Franziska Michor. "Evolutionary theory of cancer". In: *Annals of the New York Academy of Sciences* 1168.1 (2009), pp. 23–51.

[9] Robert M Axelrod. *The evolution of cooperation*. Basic books, 2006.

[10] Robert Axelrod, David E Axelrod, and Kenneth J Pienta. "Evolution of cooperation among tumor cells". In: *Proceedings of the National Academy of Sciences* 103.36 (2006), pp. 13474–13479.

[11] Pia Bader et al. "Disease progression and survival of patients with positive lymph nodes after radical prostatectomy. Is there a chance of cure?" In: *The Journal of urology* 169.3 (2003), pp. 849–854.

[12] David Basanta and Andreas Deutsch. "A Game Theoretical Perspective on the Somatic Evolution of Cancer". In: *Modeling and Simulation in Science, Engineering and Technology* (2008), p. 97.

[13] Michéle Basseville. "Divergence measures for statistical data processing—An annotated bibliography". In: *Signal Processing* 93.4 (2013), pp. 621–633.

[14] Gustavo EAPA Batista, Xiaoyue Wang, and Eamonn J Keogh. "A complexity-invariant distance measure for time series". In: *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM. 2011, pp. 699–710.

[15] Leonard Baum. “An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process”. In: *Inequalities* 3 (1972), pp. 1–8.

[16] Riccardo Bellazzi and Blaz Zupan. “Predictive data mining in clinical medicine: current issues and guidelines”. In: *International journal of medical informatics* 77.2 (2008), pp. 81–97.

[17] Sébastien Benzekry et al. “Classical mathematical models for description and prediction of experimental tumor growth”. In: *PLoS Comput Biol* 10.8 (2014), e1003800.

[18] Donald J Berndt and James Clifford. “Using dynamic time warping to find patterns in time series.” In: *KDD workshop*. Vol. 10. 16. Seattle, WA. 1994, pp. 359–370.

[19] Swati Bhonsle et al. *Centroid Distance Function and the Fourier Descriptor with Applications to Cancer Cell Clustering*. 2011.

[20] SP Blagden et al. “Performance status score: do patients and their oncologists agree?” In: *British journal of cancer* 89.6 (2003), p. 1022.

[21] Bernard H Bochner, Michael W Kattan, and Kinjal C Vora. “Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer.” In: *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 24.24 (2006), pp. 3967–3972.

[22] Katrin Bohl et al. “Evolutionary game theory: molecules as players”. In: *Molecular BioSystems* 10.12 (2014), pp. 3066–3074.

[23] Sylvia F Boj et al. “Organoid models of human and mouse ductal pancreatic cancer”. In: *Cell* 160.1 (2015), pp. 324–338.

[24] Wayne Brewer, Brian T Swanson, and Alexis Ortiz. “Validity of Fitbit’s active minutes as compared with a research-grade accelerometer and self-reported measures”. In: *BMJ open sport & exercise medicine* 3.1 (2017), e000254.

[25] G Buccheri, D Ferrigno, and M Tamburini. “Karnofsky and ECOG performance status scoring in lung cancer: a prospective, longitudinal study of 536 patients from a single institution”. In: *European Journal of Cancer* 32.7 (1996), pp. 1135–1141.

[26] Thomas A Burke, Tami Wisniewski, and Frank R Ernst. “Resource utilization and costs associated with chemotherapy-induced nausea and vomiting (CINV) following highly or moderately emetogenic chemotherapy administered in the US outpatient hospital setting”. In: *Supportive Care in Cancer* 19.1 (2011), pp. 131–140.

[27] Rebecca A Burrell et al. “The causes and consequences of genetic heterogeneity in cancer evolution”. In: *Nature* 501.7467 (2013), pp. 338–345.

[28] RJ Butland et al. “Two-, six-, and 12-minute walking tests in respiratory disease.” In: *British medical journal (Clinical research ed.)* 284.6329 (1982), p. 1607.

[29] Atul J Butte and Isaac S Kohane. “Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements”. In: *Biocomputing 2000*. World Scientific, 1999, pp. 418–429.

[30] *Cancer Facts & Figures 2018.* Atlanta: American Cancer Society. 2018.

[31] Sung-Hyuk Cha. “Comprehensive survey on distance/similarity measures between probability density functions”. In: *City* 1.2 (2007), p. 1.

[32] Liang Cheng et al. “Biomarkers in bladder cancer: translational and clinical implications”. In: *Critical reviews in oncology/hematology* 89.1 (2014), pp. 73–111.

[33] Sierra Cheng et al. “Do patients with reduced or excellent performance status derive the same clinical benefit from novel systemic cancer therapies? A systematic review and meta-analysis”. In: *ESMO open* 2.4 (2017), e000225.

[34] Rebecca Clark-Snow, Mary Lou Affronti, and Cynthia N Rittenberg. “Chemotherapy-induced nausea and vomiting (CINV) and adherence to antiemetic guidelines: results of a survey of oncology nurses”. In: *Supportive Care in Cancer* 26.2 (2018), pp. 557–564.

[35] David J Cook et al. “Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology”. In: *The Annals of thoracic surgery* 96.3 (2013), pp. 1057–1061.

[36] Timothy F Cootes et al. “Active shape models-their training and application”. In: *Computer Vision and Image Understanding* 61.1 (1995), pp. 38–59.

[37] Timothy F Cootes, Chris J Taylor, et al. *Statistical models of appearance for computer vision*. 2004.

[38] Romain Corre et al. “Use of a comprehensive geriatric assessment for the management of elderly patients with advanced non–small-cell lung cancer: The phase III randomized ESOGIA-GFPC-GECP 08-02 study”. In: *Journal of Clinical Oncology* 34.13 (2016), pp. 1476–1483.

[39] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[40] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[41] Bernard Crespi and Kyle Summers. “Evolutionary biology of cancer”. In: *Trends in Ecology & Evolution* 20.10 (2005), pp. 545–552.

[42] Vittorio Cristini, John Lowengrub, and Qing Nie. “Nonlinear simulation of tumor growth”. In: *Journal of mathematical biology* 46.3 (2003), pp. 191–224.

[43] Joseph A Cruz and David S Wishart. "Applications of machine learning in cancer prediction and prognosis". In: *Cancer informatics* 2 (2006), p. 117693510600200030.

[44] Richard Dawkins. *The selfish gene*. Oxford university press, 2016.

[45] Peter Dear. *The intelligibility of nature: How science makes sense of the world*. University of Chicago Press, 2008.

[46] Kaat Desloovere et al. "Do dynamic and static clinical measurements correlate with gait analysis parameters in children with cerebral palsy?" In: *Gait & posture* 24.3 (2006), pp. 302–313.

[47] Jakub Dobruch et al. "Gender and bladder cancer: a collaborative review of etiology, biology, and outcomes". In: *European urology* 69.2 (2016), pp. 300–310.

[48] Michael Doebeli and Christoph Hauert. "Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game". In: *Ecology letters* 8.7 (2005), pp. 748–766.

[49] Michael Doebeli, Christoph Hauert, and Timothy Killingback. "The evolutionary origin of cooperators and defectors". In: *science* 306.5697 (2004), pp. 859–862.

[50] Pedro Domingos. "A few useful things to know about machine learning". In: *Communications of the ACM* 55.10 (2012), pp. 78–87.

[51] David Donaire-Gonzalez et al. "Benefits of physical activity on COPD hospitalisation depend on intensity". In: *European Respiratory Journal* 46.5 (2015), pp. 1281–1289.

[52] Jarno Drost and Hans Clevers. "Organoids in cancer research". In: *Nature Reviews Cancer* (2018), p. 1.

[53] Michael T Durheim et al. "Six-minute-walk distance and accelerometry predict outcomes in chronic obstructive pulmonary disease independent of Global Initiative for Chronic Obstructive Lung Disease 2011 Group". In: *Annals of the American Thoracic Society* 12.3 (2015), pp. 349–356.

[54] Richard Durrett and Stephen Moseley. "Evolution of resistance and progression to disease during clonal expansion of cancer". In: *Theoretical population biology* 77.1 (2010), pp. 42–48.

[55] Rick Durrett et al. "Intratumor heterogeneity in evolutionary models of tumor progression". In: *Genetics* 188.2 (2011), pp. 461–477.

[56] Andrew J Ewald et al. "Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis". In: *Developmental Cell* 14.4 (2008), pp. 570–581.

[57] Martine Extermann et al. “Predicting the risk of chemotherapy toxicity in older patients: The Chemotherapy Risk Assessment Scale for High-Age Patients (CRASH) score”. In: *Cancer* 118.13 (2012), pp. 3377–3386.

[58] M Extermann et al. “MAX2—a convenient index to estimate the average per patient risk for chemotherapy toxicity: validation in ECOG trials”. In: *European Journal of Cancer* 40.8 (2004), pp. 1193–1198.

[59] A Farrell et al. “Nature milestones: cancer”. In: *Nat* 440 (2006), S7–S23.

[60] Steve R Fisher et al. “Mobility after hospital discharge as a marker for 30-day readmission”. In: *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 68.7 (2012), pp. 805–810.

[61] Steven A Frank. *Dynamics of cancer: incidence, inheritance, and evolution*. Princeton University Press, 2007.

[62] Steven A Frank and Martin A Nowak. “Problems of somatic mutation and cancer”. In: *Bioessays* 26.3 (2004), pp. 291–299.

[63] CL Frenzen and JD Murray. “A cell kinetics justification for Gompertz’equation”. In: *SIAM Journal on Applied Mathematics* 46.4 (1986), pp. 614–629.

[64] G Freyer et al. “Comprehensive geriatric assessment predicts tolerance to chemotherapy and survival in elderly patients with advanced ovarian carcinoma: a GINECO study”. In: *Annals of oncology* 16.11 (2005), pp. 1795–1800.

[65] Sten Friberg and Stefan Mattson. “On the growth rates of human malignant tumors: implications for medical decision making”. In: *Journal of surgical oncology* 65.4 (1997), pp. 284–297.

[66] Hermann B Frieboes et al. “Three-dimensional multispecies nonlinear tumor growth—II: tumor invasion and angiogenesis”. In: *Journal of theoretical biology* 264.4 (2010), pp. 1254–1278.

[67] Carlo Frigo and Paolo Crenna. “Multichannel SEMG in clinical gait analysis: a review and state-of-the-art”. In: *Clinical Biomechanics* 24.3 (2009), pp. 236–245.

[68] Tak-chung Fu. “A review on time series data mining”. In: *Engineering Applications of Artificial Intelligence* 24.1 (2011), pp. 164–181.

[69] Robert A Gatenby and Thomas L Vincent. “An evolutionary model of carcinogenesis”. In: *Cancer Research* 63.19 (2003), pp. 6212–6220.

[70] Robert A Gatenby and Thomas L Vincent. “Application of quantitative models from population biology and evolutionary game theory to tumor therapeutic strategies”. In: *Molecular cancer therapeutics* 2.9 (2003), pp. 919–927.

[71] Marco Gerlinger et al. “Intratumor heterogeneity and branched evolution revealed by multiregion sequencing”. In: *N Engl j Med* 2012.366 (2012), pp. 883–892.

[72] Rafael Giusti and Gustavo EAPA Batista. “An empirical comparison of dissimilarity measures for time series classification”. In: *Intelligent Systems (BRACIS), 2013 Brazilian Conference on*. IEEE. 2013, pp. 82–88.

[73] James H Goldie and Andrew J Coldman. *Drug resistance in cancer: mechanisms and models*. Cambridge University Press, 2009.

[74] James H Goldie and Andrew J Coldman. *Drug resistance in cancer: mechanisms and models*. Cambridge University Press, 2009.

[75] Mel Greaves and Carlo C Maley. “Clonal evolution in cancer”. In: *Nature* 481.7381 (2012), pp. 306–313.

[76] Cesare Gridelli and John Hainsworth. “Meeting the chemotherapy needs of elderly and poor performance status patients with NSCLC”. In: *Lung cancer* 38 (2002), pp. 37–41.

[77] Arjun Gupta et al. “Feasibility of wearable physical activity monitors in patients with cancer”. In: *JCO Clinical Cancer Informatics* 2 (2018), pp. 1–10.

[78] John D Hainsworth et al. “Weekly combination chemotherapy with docetaxel and gemcitabine as first-line treatment for elderly patients and patients with poor performance status who have extensive-stage small cell lung carcinoma: A Minnie Pearl Cancer Research Network Phase II Trial”. In: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 100.11 (2004), pp. 2437–2441.

[79] ME Hamaker, MC Prins, and R Stauder. “The relevance of a geriatric assessment for elderly patients with a haematological malignancy–a systematic review”. In: *Leukemia research* 38.3 (2014), pp. 275–283.

[80] Nathan R Handley, Lynn M Schuchter, and Justin E Bekelman. “Best Practices for Reducing Unplanned Acute Care for Patients With Cancer”. In: *Journal of oncology practice* 14.5 (2018), pp. 306–313.

[81] Zaki Hasnain et al. “Low-dimensional dynamical characterization of human performance of cancer patients using motion data”. In: *Clinical Biomechanics* 56 (2018), pp. 61–69.

[82] Md Rafiul Hassan and Baikunth Nath. “Stock market forecasting using hidden Markov model: a new approach”. In: *Intelligent Systems Design and Applications, 2005. ISDA’05. Proceedings. 5th International Conference on*. IEEE. 2005, pp. 192–196.

[83] C Hauert. “Evolution from cellular to social scales”. In: *Evolutionary Dynamics* (2008), pp. 11–44.

[84] Christoph Hauert and György Szabó. "Game theory and physics". In: *American Journal of Physics* 73.5 (2005), pp. 405–414.

[85] Gloria H Heppner. "Tumor heterogeneity". In: *Cancer research* 44.6 (1984), pp. 2259–2265.

[86] Gloria H Heppner and Bonnie E Miller. "Tumor heterogeneity: biological implications and therapeutic consequences". In: *Cancer and Metastasis Reviews* 2.1 (1983), pp. 5–23.

[87] Harry W Herr et al. "Surgical factors influence bladder cancer outcomes: a cooperative group report". In: *Journal of clinical oncology* 22.14 (2004), pp. 2781–2789.

[88] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[89] Gineke A ten Holt, Marcel JT Reinders, and EA Hendriks. "Multi-dimensional dynamic time warping for gesture recognition". In: *Thirteenth annual conference of the Advanced School for Computing and Imaging*. Vol. 300. 2007.

[90] Steve Horvath. *Weighted network analysis: applications in genomics and systems biology*. Springer Science & Business Media, 2011.

[91] Sabine Hummert et al. "Evolutionary game theory: cells as players". In: *Molecular BioSystems* 10.12 (2014), pp. 3044–3065.

[92] Arti Hurria et al. "Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study". In: *Journal of Clinical Oncology* 29.25 (2011), p. 3457.

[93] Yoh Iwasa, Martin A Nowak, and Franziska Michor. "Evolution of resistance during clonal expansion". In: *Genetics* 172.4 (2006), pp. 2557–2566.

[94] Dan Jurafsky and James H Martin. *Speech and language processing*. Vol. 3. Pearson London, 2014.

[95] Irina Kareva. "Prisoner's dilemma in cancer metabolism". In: *PloS one* 6.12 (2011), e28576.

[96] DA Karnofsky and JH Burchenal. "The evaluation of chemotherapeutic agents against neoplastic disease". In: *Cancer Research*. Vol. 8. AMER ASSOC CANCER RESEARCH PO BOX 11806, BIRMINGHAM, AL 35202. 1948, pp. 388–389.

[97] David A Karnofsky, Rose Ruth Ellison, and Robert B Golbey. "Selection of patients for evaluation of chemotherapeutic procedures in advanced cancer". In: *Journal of Clinical Epidemiology* 15.3 (1962), pp. 243–249.

[98] Tomoya Kawaguchi et al. “Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC”. In: *Journal of Thoracic Oncology* 5.5 (2010), pp. 620–630.

[99] Ciara M Kelly and Armin Shahrokni. “Moving beyond Karnofsky and ECOG performance status assessments with new technologies”. In: *Journal of oncology* 2016 (2016).

[100] WS Kendal. “Gompertzian growth as a consequence of tumor heterogeneity”. In: *Mathematical Biosciences* 73.1 (1985), pp. 103–107.

[101] Eamonn J Keogh and Michael J Pazzani. “Derivative dynamic time warping”. In: *Proceedings of the 2001 SIAM International Conference on Data Mining*. SIAM. 2001, pp. 1–11.

[102] Yangjin Kim, Magdalena A Stolarska, and Hans G Othmer. “The role of the microenvironment in tumor growth and invasion”. In: *Progress in biophysics and molecular biology* 106.2 (2011), pp. 353–379.

[103] Michael Kirby. *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. John Wiley & Sons, 2001.

[104] Alfred G Knudson. “Mutation and cancer: statistical study of retinoblastoma”. In: *Proceedings of the National Academy of Sciences* 68.4 (1971), pp. 820–823.

[105] Natalia Komarova. “Stochastic modeling of drug resistance in cancer”. In: *Journal of theoretical biology* 239.3 (2006), pp. 351–366.

[106] Natalia L Komarova and Dominik Wodarz. “Drug resistance in cancer: principles of emergence and prevention”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.27 (2005), pp. 9714–9719.

[107] Igor Kononenko. “Machine learning for medical diagnosis: history, state of the art and perspective”. In: *Artificial Intelligence in medicine* 23.1 (2001), pp. 89–109.

[108] Konstantina Kourou et al. “Machine learning applications in cancer prognosis and prediction”. In: *Computational and structural biotechnology journal* 13 (2015), pp. 8–17.

[109] Frank Kozusko and Željko Bajzer. “Combining Gompertzian growth and cell population dynamics”. In: *Mathematical Biosciences* 185.2 (2003), pp. 153–167.

[110] Fabian L Kriegel et al. “Cell shape characterization and classification with discrete Fourier transforms and self-organizing maps”. In: *Cytometry Part A* 93.3 (2018), pp. 323–333.

[111] Anna Kane Laird. “Dynamics of tumour growth”. In: *British journal of cancer* 18.3 (1964), p. 490.

[112] Anna Kane Laird. "Dynamics of tumour growth: comparison of growth rates and extrapolation of growth curve to one cell". In: *British Journal of Cancer* 19.2 (1965), p. 278.

[113] Madeline A Lancaster and Juergen A Knoblich. "Organogenesis in a dish: modeling development and disease using organoid technologies". In: *Science* 345.6194 (2014), p. 1247125.

[114] Keun-Wook Lee et al. "Weekly low-dose docetaxel for salvage chemotherapy in pretreated elderly or poor performance status patients with non-small cell lung cancer". In: *Journal of Korean medical science* 23.6 (2008), pp. 992–998.

[115] TH Li. *Robust divergence measures for time series discrimination*. Tech. rep. Tech. Rept. 237, Statistics Dept., Texas A&M Univ., College Station, 1995.

[116] Erez Lieberman, Christoph Hauert, and Martin A Nowak. "Evolutionary dynamics on graphs". In: *Nature* 433.7023 (2005), pp. 312–316.

[117] Paul Macklin and John Lowengrub. "Nonlinear simulation of the effect of microenvironment on tumor growth". In: *Journal of theoretical biology* 245.4 (2007), pp. 677–704.

[118] Thomas Robert Malthus. *An essay on the principle of population: or, A view of its past and present effects on human happiness*. Reeves & Turner, 1888.

[119] Carlos R Marti-Figueroa and Randolph S Ashton. "The case for applying tissue engineering methodologies to instruct human organoid morphogenesis". In: *Acta biomaterialia* 54 (2017), pp. 35–44.

[120] Alexander S Martin et al. *Objective metrics of patient activity: Use of wearable trackers and patient reported outcomes in predicting unexpected healthcare events in cancer patients undergoing highly emetogenic chemotherapy*. 2018.

[121] LC Martin et al. "Using information theory to search for co-evolving residues in proteins". In: *Bioinformatics* 21.22 (2005), pp. 4116–4124.

[122] Iñigo Martincorena and Peter J Campbell. "Somatic mutation in cancer and normal cells". In: *Science* 349.6255 (2015), pp. 1483–1489.

[123] Andriy Marusyk and Kornelia Polyak. "Tumor heterogeneity: causes and consequences". In: *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1805.1 (2010), pp. 105–117.

[124] Eric Meijering, Oleh Dzyubachyk, and Ihor Smal. "Methods for Cell and Particle Tracking". In: *Methods in Enzymology Vol. 504*. Elsevier, 2012, pp. 183–200.

[125] Erik Meijering. "Cell segmentation: 50 years down the road". In: *IEEE Signal Processing Magazine* 140 (2012).

[126] Lauren MF Merlo et al. "Cancer as an evolutionary and ecological process". In: *Nature Reviews Cancer* 6.12 (2006), pp. 924–935.

[127] Masaru Moriyama et al. "Relevance network between chemosensitivity and transcriptome in human hepatoma cells1". In: *Molecular Cancer Therapeutics* 2.2 (2003), pp. 199–205.

[128] Meinard Müller. *Information retrieval for music and motion*. Vol. 2. Springer, 2007.

[129] Lillian M Nail. "My get up and go got up and went: fatigue in people with cancer". In: *JNCI Monographs* 2004.32 (2004), pp. 72–75.

[130] John Nash. "Non-cooperative games". In: *Annals of mathematics* (1951), pp. 286–295.

[131] John F Nash et al. "Equilibrium points in n-person games". In: *Proceedings of the national academy of sciences* 36.1 (1950), pp. 48–49.

[132] Nasser M Nasrabadi. "Pattern recognition and machine learning". In: *Journal of electronic imaging* 16.4 (2007), p. 049901.

[133] Nicholas Navin et al. "Inferring tumor progression from genomic heterogeneity". In: *Genome research* 20.1 (2010), pp. 68–80.

[134] Paul K Newton et al. "Entropy, complexity, and Markov diagrams for random walk cancer models". In: *Scientific reports* 4 (2014).

[135] M. N. B. Nguyen et al. "Mining Human Mobility to Quantify Performance Status". In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. Nov. 2017, pp. 1172–1177. DOI: 10.1109/ICDMW.2017.168.

[136] Minh NB Nguyen et al. "Mining human mobility to quantify performance status". In: *Data Mining Workshops (ICDMW), 2017 IEEE International Conference on*. IEEE. 2017, pp. 1172–1177.

[137] Kim-Vy Nguyen-Ngoc et al. "3D culture assays of murine mammary branching morphogenesis and epithelial invasion". In: *Tissue Morphogenesis*. Springer, 2015, pp. 135–162.

[138] CO Nordling. "A new theory on the cancer-inducing mechanism". In: *British journal of cancer* 7.1 (1953), p. 68.

[139] L Norton and R Simon. "Tumor size, sensitivity to therapy, and design of treatment schedules". In: *Cancer Treatment Reports* 61.7 (1977), p. 1307.

[140] Larry Norton and Richard Simon. "Growth curve of an experimental solid tumor following radiotherapy". In: *Journal of National Cancer Institute* 58.6 (1977).

[141] M Nowak. "EvolutionaryDynamics: ExploringtheEquationsofLife". In: *Cambridge: HarvardUniversityPress* (2006).

[142] Martin Nowak. "Stochastic strategies in the prisoner's dilemma". In: *Theoretical Population Biology* 38.1 (1990), pp. 93–112.

[143] Martin A Nowak. *Evolutionary Dynamics*. Harvard University Press, 2006.

[144] Martin A Nowak and Robert M May. "Evolutionary games and spatial chaos". In: *Nature* 359.6398 (1992), pp. 826–829.

[145] Martin A Nowak and Karl Sigmund. "Evolutionary dynamics of biological games". In: *science* 303.5659 (2004), pp. 793–799.

[146] Peter C Nowell. "The clonal evolution of tumor cell populations". In: *Science* 194.4260 (1976), pp. 23–28.

[147] Martin M Oken et al. "Toxicity and response criteria of the Eastern Cooperative Oncology Group." In: *American journal of clinical oncology* 5.6 (1982), pp. 649–656.

[148] Sai-Hong Ignatius Ou and Jason A Zell. "Validation study of the proposed IASLC staging revisions of the T4 and M non-small cell lung cancer descriptors using data from 23,583 patients in the California Cancer Registry". In: *Journal of Thoracic Oncology* 3.3 (2008), pp. 216–227.

[149] Fabian Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *Journal of machine learning research* 12.Oct (2011), pp. 2825–2830.

[150] François Petitjean, Alain Ketterlin, and Pierre Gançarski. "A global averaging method for dynamic time warping, with applications to clustering". In: *Pattern Recognition* 44.3 (2011), pp. 678–693.

[151] Dominik Péus, Nicolas Newcomb, and Silvia Hofer. "Appraisal of the Karnofsky Performance Status and proposal of a simple algorithmic system for its evaluation". In: *BMC medical informatics and decision making* 13.1 (2013), p. 72.

[152] Zachary Pincus and JA Theriot. "Comparison of quantitative methods for cell-shape analysis". In: *Journal of Microscopy* 227.2 (2007), pp. 140–156.

[153] William F Pirl et al. *Actigraphy as an objective measure of performance status in patients with advanced cancer*. 2015.

[154] Jennifer Pohle et al. "Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement". In: *Journal of Agricultural, Biological and Environmental Statistics* 22.3 (2017), pp. 270–293.

[155] Marcus L Quek et al. “Prognostic significance of lymphovascular invasion of bladder cancer treated with radical cystectomy”. In: *The Journal of urology* 174.1 (2005), pp. 103–106.

[156] Lawrence R Rabiner. “A tutorial on hidden Markov models and selected applications in speech recognition”. In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.

[157] Lawrence R Rabiner and Biing-Hwang Juang. “An introduction to hidden Markov models”. In: *ieee assp magazine* 3.1 (1986), pp. 4–16.

[158] E Radzikowska, P Głaz, and K Roszkowski. “Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases”. In: *Annals of oncology* 13.7 (2002), pp. 1087–1093.

[159] Aliya Ramjaun et al. “Improved targeting of cancer care for older patients: a systematic review of the utility of comprehensive geriatric assessment”. In: *Journal of geriatric oncology* 4.3 (2013), pp. 271–281.

[160] Chotirat Ann Ratanamahatana and Eamonn Keogh. “Everything you know about dynamic time warping is wrong”. In: *Third Workshop on Mining Temporal and Sequential Data*. Citeseer. 2004, pp. 22–25.

[161] Chotirat Ann Ratanamahatana and Eamonn Keogh. “Three myths about dynamic time warping data mining”. In: *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM. 2005, pp. 506–510.

[162] Robert A Redden and Edward J Doolin. “Microgravity assay of neuroblastoma: in vitro aggregation kinetics and organoid morphology correlate with MYCN expression”. In: *In Vitro Cellular & Developmental Biology-Animal* 47.4 (2011), pp. 312–317.

[163] Katarzyna A Rejniak and Alexander RA Anderson. “Hybrid models of tumor growth”. In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 3.1 (2011), pp. 115–125.

[164] Eric Roeland et al. *Hospitalization costs for nausea and vomiting: A savings opportunity*. 2017.

[165] Sheldon M Ross. *Introduction to Probability Models: Solutions Manual for Introduction to Probability Models. Solu, 4*. Acad. Press, 1989.

[166] Charles M Rubin. “The genetic basis of human cancer”. In: *Annals of Internal Medicine* 129.9 (1998), pp. 759–759.

[167] Graeme D Ruxton. “The unequal variance t-test is an underused alternative to Student’s t-test and the Mann–Whitney U test”. In: *Behavioral Ecology* 17.4 (2006), pp. 688–690.

[168] Yara E Sánchez-Corrales et al. “Morphometrics of complex cell shapes: lobe contribution elliptic Fourier analysis (LOCO-EFA)”. In: *Development* (2018), dev–156778.

[169] Doruk Sart et al. “Accelerating dynamic time warping subsequence search with GPUs and FPGAs”. In: *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE. 2010, pp. 1001–1006.

[170] Randall T Schuh. *Biological systematics: principles and applications*. Cornell University Press, 2000.

[171] Stefan Schuster et al. “Use of game-theoretical methods in biochemistry and biophysics”. In: *Journal of biological physics* 34.1-2 (2008), pp. 1–17.

[172] Eliah R Shamir and Andrew J Ewald. “Three-dimensional organotypic culture: experimental models of mammalian biology and disease”. In: *Nature Reviews Molecular Cell Biology* 15.10 (2014), p. 647.

[173] Michail Shipitsin et al. “Molecular definition of breast tumor heterogeneity”. In: *Cancer cell* 11.3 (2007), pp. 259–273.

[174] Mohammad Shokoohi-Yekta et al. “Generalizing DTW to the multi-dimensional case requires an adaptive approach”. In: *Data Mining and Knowledge Discovery* 31.1 (2017), pp. 1–31.

[175] Rebecca L Siegel et al. “Colorectal cancer statistics, 2017”. In: *CA: a cancer journal for clinicians* 67.3 (2017), pp. 177–193.

[176] Sheldon R Simon. “Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems”. In: *Journal of biomechanics* 37.12 (2004), pp. 1869–1880.

[177] Howard E Skipper. “Laboratory models: some historical perspective”. In: *Cancer Treatment Reports* 70.1 (1986), pp. 3–7.

[178] J Maynard Smith. “Evolutionary game theory”. In: *Vito Volterra symposium on mathematical models in biology*. Springer. 1980, pp. 73–81.

[179] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[180] John S Spratt and John A Spratt. “What is breast cancer doing before we can detect it?” In: *Journal of surgical oncology* 30.3 (1985), pp. 156–160.

[181] Mario Stanke and Stephan Waack. “Gene prediction with a hidden Markov model and a new intron submodel”. In: *Bioinformatics* 19.suppl_2 (2003), pp. ii215–ii225.

[182] John P Stein et al. “Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients”. In: *Journal of clinical oncology* 19.3 (2001), pp. 666–675.

[183] Arnulf Stenzl et al. “Treatment of muscle-invasive and metastatic bladder cancer: update of the EAU guidelines”. In: *European urology* 59.6 (2011), pp. 1009–1018.

[184] Sang-Yeon Suh et al. “Longitudinal patient-reported performance status assessment in the cancer clinic is feasible and prognostic”. In: *Journal of oncology practice* 7.6 (2011), pp. 374–381.

[185] Robert S Svatek et al. “Discrepancy between clinical and pathological stage: external validation of the impact on prognosis in an international radical cystectomy cohort”. In: *BJU international* 107.6 (2011), pp. 898–904.

[186] Charles Swanton. “Intratumor heterogeneity: evolution through space and time”. In: *Cancer Research* 72.19 (2012), pp. 4875–4882.

[187] Christopher J Sweeney et al. “Outcome of patients with a performance status of 2 in eastern cooperative oncology group study E1594: a phase III trial in patients with metastatic nonsmall cell lung carcinoma”. In: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 92.10 (2001), pp. 2639–2647.

[188] Tetsuya Takahashi et al. “In-patient step count predicts re-hospitalization after cardiac surgery”. In: *Journal of cardiology* 66.4 (2015), pp. 286–291.

[189] Asoke K Talukder et al. “Tracking Cancer Genetic Evolution using OncoTrack”. In: *Scientific reports* 6 (2016), srep29647.

[190] Weijun Tao et al. “Gait analysis using wearable sensors”. In: *Sensors* 12.2 (2012), pp. 2255–2283.

[191] ANNE E Taylor et al. “Observer error in grading performance status in cancer patients”. In: *Supportive Care in cancer* 7.5 (1999), pp. 332–335.

[192] Ian PM Tomlinson, MR Novelli, and WF Bodmer. “The mutation rate and cancer”. In: *Proceedings of the National Academy of Sciences* 93.25 (1996), pp. 14800–14803.

[193] IPM Tomlinson. “Game-theory models of interactions between tumour cells”. In: *European Journal of Cancer* 33.9 (1997), pp. 1495–1500.

[194] P Tracqui. “Biophysical models of tumour growth”. In: *Reports on Progress in Physics* 72.5 (2009), p. 056701.

[195] Tiffany A Traina et al. “Optimizing chemotherapy dose and schedule by Norton-Simon mathematical modeling”. In: *Breast Disease* 31.1 (2010), pp. 7–18.

[196] Arne Traulsen, Jens Christian Claussen, and Christoph Hauert. “Coevolutionary dynamics: from finite to infinite populations”. In: *Physical review letters* 95.23 (2005), p. 238701.

[197] Arne Traulsen and Christoph Hauert. “Stochastic evolutionary game dynamics”. In: *Reviews of nonlinear dynamics and complexity* 2 (2009), pp. 25–61.

[198] Katia Turcot et al. “New accelerometric method to discriminate between asymptomatic subjects and patients with medial knee osteoarthritis during 3-D gait”. In: *IEEE Transactions on Biomedical Engineering* 55.4 (2008), pp. 1415–1422.

[199] A_P Varga and RK Moore. “Hidden Markov model decomposition of speech and noise”. In: *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*. IEEE. 1990, pp. 845–848.

[200] Andrew J Vickers et al. “Clinical benefits of a multivariate prediction model for bladder cancer: a decision analytic approach”. In: *Cancer* 115.23 (2009), pp. 5460–5469.

[201] Thomas L Vincent. “Carcinogenesis as an evolutionary game”. In: *Advances in Complex Systems* 9.04 (2006), pp. 369–382.

[202] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton university press, 2007.

[203] James C Wall et al. “The Timed Get-up-and-Go test revisited: measurement of the component tasks”. In: *Journal of rehabilitation research and development* 37.1 (2000).

[204] J Walsh, J Hussey, and D O’Donnell. “A pilot study comparing objective physical activity to the physical component of the Eastern Cooperative Oncology Group (ECOG) performance status scale”. In: *Journal of Clinical Oncology* 27.15S (2009), e20501–e20501.

[205] Zhen Wang et al. “Universal scaling for the dilemma strength in evolutionary games”. In: *Physics of life reviews* 14 (2015), pp. 1–30.

[206] Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.

[207] Robert Weinberg. *The biology of cancer*. Garland science, 2013.

[208] Robert Allan Weinberg. *One renegade cell: how cancer begins*. Basic Books, 1998.

[209] Geoffrey B West, James H Brown, and Brian J Enquist. “A general model for ontogenetic growth”. In: *Nature* 413.6856 (2001), pp. 628–631.

[210] Jeffrey West et al. “An evolutionary model of tumor cell kinetics and the emergence of molecular heterogeneity driving gompertzian growth”. In: *SIAM Review* 58.4 (2016), pp. 716–736.

[211] Jeffrey West et al. “The prisoner’s dilemma as a cancer model”. In: *Convergent science physical oncology* 2.3 (2016), p. 035002.

[212] EP Wigner. “The unreasonable effectiveness of mathematics in the natural sciences”. In: *Philosophical Reflections and Syntheses*. Springer, 1995, pp. 534–549.

[213] Steven M Wise et al. “Three-dimensional multispecies nonlinear tumor growth—I: model and numerical method”. In: *Journal of theoretical biology* 253.3 (2008), pp. 524–543.

[214] Junji Yamato, Jun Ohya, and Kenichiro Ishii. “Recognizing human action in time-sequential images using hidden markov model”. In: *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*. IEEE. 1992, pp. 379–385.

[215] Lucy R Yates and Peter J Campbell. “Evolution of the cancer genome”. In: *Nature Reviews Genetics* 13.11 (2012), pp. 795–806.

[216] Illhoi Yoo et al. “Data mining in healthcare and biomedicine: a survey of the literature”. In: *Journal of medical systems* 36.4 (2012), pp. 2431–2448.

[217] Jiaping Zhao, Zerong Xi, and Laurent Itti. “metricDTW: local distance metric learning in Dynamic Time Warping”. In: *arXiv preprint arXiv:1606.03628* (2016).

[218] Yunyue Zhu and Dennis Shasha. “Warping indexes with envelope transforms for query by humming”. In: *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*. ACM. 2003, pp. 181–192.

[219] Christophe Zimmer et al. “Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: A tool for cell based drug testing”. In: *IEEE Trans. on Medical Imaging* 21.10 (2002), pp. 1212–1221.