PROBABILISTIC LEARNING OF TREATMENT TREES IN CANCER

BY TSUNG-HUNG YAO[1], ZHENKE WU[1,*], KARTHIK BHARATH[2,§], JINJU LI[1,†] AND VEERABHADRAN BALADANDAYUTHAPANI[1,‡]

[1]*Department of Biostatistics, University of Michigan at Ann Arbor, yaots@umich.edu;* [*]*zhenkewu@umich.edu;* [†]*lijinju@umich.edu;* [‡]*veerab@umich.edu*

[2]*School of Mathematical Sciences, University of Nottingham,* [§]*Karthik.Bharath@nottingham.ac.uk*

Accurate identification of synergistic treatment combinations and their underlying biological mechanisms is critical across many disease domains, especially cancer. In translational oncology research, preclinical systems such as patient-derived xenografts (PDX) have emerged as a unique study design evaluating multiple treatments administered to samples from the same human tumor implanted into genetically identical mice. In this paper, we propose a novel Bayesian probabilistic tree-based framework for PDX data to investigate the hierarchical relationships between treatments by inferring treatment cluster trees, referred to as treatment trees (R_X-tree). The framework motivates a new metric of mechanistic similarity between two or more treatments accounting for inherent uncertainty in tree estimation; treatments with a high estimated similarity have potentially high mechanistic synergy. Building upon Dirichlet Diffusion Trees, we derive a closed-form marginal likelihood encoding the tree structure, which facilitates computationally efficient posterior inference via a new two-stage algorithm. Simulation studies demonstrate superior performance of the proposed method in recovering the tree structure and treatment similarities. Our analyses of a recently collated PDX dataset produce treatment similarity estimates that show a high degree of concordance with known biological mechanisms across treatments in five different cancers. More importantly, we uncover new and potentially effective combination therapies that confer synergistic regulation of specific downstream biological pathways for future clinical investigations. Our accompanying code, data, and shiny application for visualization of results are available at: `https://github.com/bayesrx/RxTree`.

1. Introduction According to the World Health Organization, cancer is one of the leading causes of death globally, with $\sim$10 million deaths in 2020 (Ferlay et al., 2020). Despite multiple advances over the years, systematic efforts to predict efficacy of cancer treatments have been stymied due to multiple factors, including patient-specific heterogeneity and treatment resistance (Dagogo-Jack and Shaw, 2018; Groisberg and Subbiah, 2021). Given that the evolution of tumors relies on a limited number of biological mechanisms, there has been a recent push towards combining multiple therapeutic agents, referred to as "combination therapy" (Sawyers, 2013; Groisberg and Subbiah, 2021). This is driven by the core hypothesis that combinations of drugs act in synergistic manner, with each drug compensating for the drawbacks of other drugs. However, despite higher response rates and efficacy in certain instances (Bayat Mokhtari et al., 2017), combination therapy can lead to undesired drug-drug interactions, lower efficacy, or severe side effects (Sun, Sanderson and Zheng, 2016). Consequently, it is highly desirable to advance the understanding of underlying mechanisms that confer synergistic drug effects and identify potential favorable drug-drug interaction mechanisms for further investigations.

Keywords and phrases: Approximate Bayesian Computation, Dirichlet Diffusion Trees, Patient Derived Xenograft, Precision Medicine, Tree-Based Clustering.

Given that not all possible drug combinations can be tested on patients in actual clinical trials, cancer researchers rely on preclinical "model" systems to guide the discovery of the most effective combination therapies (note, models have a different contextual meaning here). In translational oncology, preclinical models assess promising treatments and compounds, before they are phased into human clinical trials. The traditional mainstay of such preclinical models has been cell-lines, wherein cell cultures derived from human tumors are grown in an *in vitro* controlled environment. However, it has been argued that they do not accurately reflect the true behavior of the host tumor and, in the process of adapting to *in vitro* growth, lose the original properties of the host tumor, thus leading to limited clinical relevance and successes (Tentler et al., 2012; Bhimani, Ball and Stebbing, 2020). To overcome these challenges, there has been a push towards more clinically relevant model systems that maintain a high degree of fidelity to human tumors. One such preclinical model system is Patient-Derived Xenograft (PDX) wherein tumor fragments obtained from cancer patients are directly transplanted into genetically identical mice (Hidalgo et al., 2014; Lai et al., 2017). Compared to traditional oncology models such as cell-lines (Yoshida, 2020), PDX models maintain key cellular and molecular characteristics, and are thus more likely to mimic human tumors and facilitate precision medicine. More importantly, accumulating evidence suggests responses (e.g. drug sensitivity) to standard therapeutic regimens in PDXs closely correlate with patient clinical data, making PDX an effective and predictive experimental model across multiple cancers (Topp et al., 2014; Nunes et al., 2015).

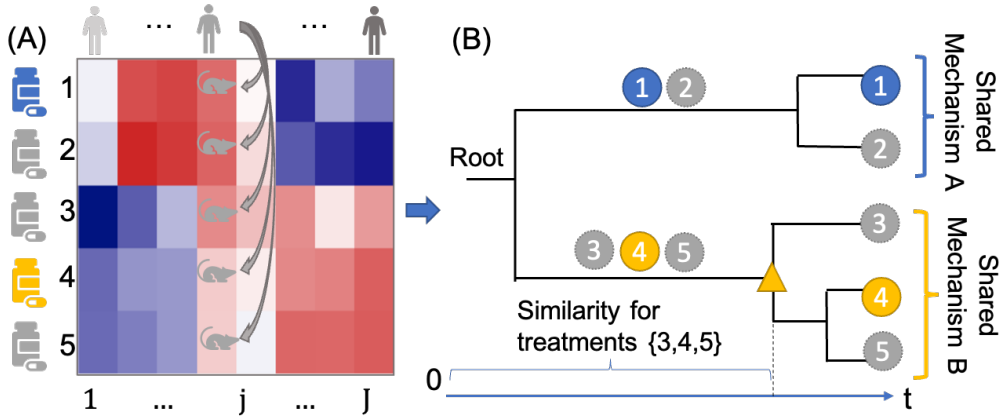

Figure 1: PDX experimental design and tree-based representation. Panel **A**: an illustrative PDX dataset with five treatments (row) and eight patients (column). Mice in a given column are implanted with tumors from the same patient and receive different treatments (across rows). The level of tumor responses are shown along a color gradient. Panel **B**: a tree structure that clusters the treatments and quantifies the similarity among mechanisms. Two treatments (1 and 4) are assumed to have different but known biological mechanisms (in different colors); the rest three treatments (2,3, and 5) have unknown mechanisms (in gray). The tree suggests two treatment groups are present ($\{1,2\}$ and $\{3,4,5\}$) that may correspond to two different known mechanisms. The horizontal position of "$\triangle$" represents the divergence time (defined in Section 2.1) and the mechanism similarity for treatments $\{3,4,5\}$. In a real data analysis, the tree (topology and divergence times) is unknown and is to be inferred from PDX data.

PDX experimental design and key scientific questions. Overall, the PDX experimental design depends on the purpose of the study and we consider a PDX study of the PDX clinical

trial that includes a large number of patients (Abdolahi et al., 2022) and tests a set of common treatments. The PDX experiment then implants the tumor cell to multiple mice and each treatment is given to multiple mice with tumors implanted from the same (matched) patient (see conceptual schema in Figure 1(A)). Treatment responses (e.g. tumor size) are then evaluated, resulting in a data matrix (treatments × patients) as depicted in the heatmap in Figure 1(A). The PDX-based clinical trial is a powerful tool for detecting the drug efficacy and drug sensitivity (Abdolahi et al., 2022) and has been adapted in several studies for different cancers (e.g. Zhang et al., 2013 for the breast cancer and Bertotti et al., 2011 for the colorectal cancer). Due to the relatively high fidelity between PDX models and the human tumors (Oh and Bang, 2020; Abdolahi et al., 2022), a PDX-based clinical trial mirrors a real human clinical trial using mouse "avatars" (Clohessy and Pandolfi, 2015). Thus this protocol serves as a scalable platform to: (a) identify underlying plausible biological mechanisms responsible for tumor growth and resistance, and (b) evaluate the effectiveness of drug combinations based on mechanistic understanding. In this context, the (biological) mechanism refers to the specific mechanism of action of a treatment, which usually represents a specific target, such as an enzyme or a receptor (Grant, Combs and Acosta, 2010). From the perspective of treatment responses as data, responses are the consequences of the downstream biological pathways from the corresponding interaction between a treatment and the target/mechanism.

Ideally, treatments with the same target/mechanism should induce similar responses and engender mechanism-related clustering among treatments. Evidently then, a sensible clustering of treatments would not only partition treatments into clusters but also explicate how the clusters relate to one another; in other words, a hierarchy among treatment clusters is more likely to uncover plausible mechanisms for combinations of treatments with "similar" responses when compared to "flat" clusters (e.g., k-means clustering). Such response-based identification of potential synergistic effects from combinations of treatments will augment understanding from known mechanistic synergy. In our application, using tree-based clustering, we assume known entities at the leaves, i.e., the different treatments. The treatments are assumed to act upon potentially distinct biological pathways, resulting in different levels of responses across the treated mice. In this paper, we use PDX response data on the leaves to infer a hierarchy over treatments that may empirically characterize the similarity in the targeted mechanistic pathways. The primary statistical goals are to (i) define and estimate a general metric measuring the similarity within any subset comprising two or more treatments, and (ii) facilitate (i) by conceptualizing and inferring an unknown hierarchy among treatments.

Tree-based representations for PDX data. To this end, we consider a tree-based construct to explore the hierarchical relationships between treatments, referred to as *treatment tree (R_x-tree*, in short). We view such a tree structure as a representation of clustering of treatments based on mechanisms that confer synergistic effects, wherein similarities between mechanisms are captured through branch lengths. Hierarchy among treatments can be interpreted through branch lengths (from the root) that are potentially reflective of different cancer processes; this would then help identify common mechanisms and point towards treatment combinations disrupting oncological processes if administered simultaneously.

We will focus on rooted trees. The principal ingredients of a rooted tree comprise a root node, terminal nodes (or, leaves), internal nodes and branch lengths. In the context of the R_x-tree for PDX data, the leaves are observed treatment responses, whereas internal nodes and branch lengths are unobserved. Internal nodes are clusters of treatments, and lengths of branches between nodes are indicative of strengths of mechanism similarities. The root is a single cluster consisting of all treatments. This leads to the following interpretation: at the root all treatments share a common target or mechanism; length of path from the root to the internal node (sum of branch lengths) at which two treatments split into different

clusters measures mechanism similarity between the two treatments. Thus treatments that stay clustered "longer" have higher mechanism similarities.

Throughout, we will use 'tree' when describing methodology for an abstract tree (acyclic graphs with distinguished root node) and 'treatment tree' or 'R_x-tree' when referring to the latent tree within the application context.

An illustrative example. A conceptual R_x-tree and its interpretation is illustrated in Figure 1 where five treatments (1 to 5) are applied on eight patients' PDXs (Figure 1(A)) with the corresponding (unknown true) R_x-tree (Figure 1(B)) based on the PDX data. Assume two treatment groups based on different mechanisms – treatments $\{1, 2\}$ and treatments $\{3, 4, 5\}$; further, suppose treatment 4 is approved by the Food and Drug Administration (FDA). The heatmap in Panel (A) visualizes the distinct levels of response profiles to the five treatments so that treatments closer in the tree are more likely to have similar levels of responses. The R_x-tree captures the mechanism similarity by arranging treatments $\{1, 2\}$ and $\{3, 4, 5\}$ to stay in their respective subtrees longer and to separate the two sets of treatment early in the tree. Based on the R_x-tree, treatments $\{3, 5\}$ share high mechanism similarity values with treatment 4; treatment 5 is the closest to the treatment 4, suggesting the most similar synergistic mechanism among all the evaluated treatments 1 to 5.

Existing methods and modeling background. The Pearson correlation is a popular choice to assess mechanism similarity between treatments (Krumbach et al., 2011), but is inappropriate to examine multi-way similarity. A tree-structured approach based on a (binary) dendrogram obtained from hierarchical clustering of cell-line data using the cophenetic distance (Sokal and Rohlf, 1962) was adopted in Narayan et al. (2020); their approach, however, failed to account for uncertainty in the dendrogram, which is highly sensitive to measurement error in the response variables as well distance metrics (we show this via simulations and in real data analyses). Another example with a binary dendrogram of hierarchical clustering was proposed by Rashid et al. (2020), which also utilizes the same PDX dataset as this paper. However, their model uses the tree structure to model the individualized treatment rule for different patients, while our method focuses on the tree structure itself and the corresponding mechanism similarity. In this paper, we consider a model for PDX data parameterized by a tree-structured object representing the R_x-tree. The model is derived from the Dirichlet diffusion tree (DDT) (Neal, 2003) generative model for (hierarchically) clustered data. The DDT engenders a data likelihood and a prior distribution on the tree parameter with support in the space of rooted binary trees. We can then use the posterior distribution to quantify uncertainty about the latent R_x-tree.

Summary of novel contributions and organization of the article. Our approach based on the DDT model for PDX data results in three main novel contributions:

(a) *Derivation of a closed-form likelihood that encodes the tree structure.* The DDT specification results in a joint distribution on PDX data, treatment tree parameters and other model parameters. By marginalizing over unobserved data that correspond to internal nodes of the tree, we obtain a new multivariate Gaussian likelihood with a special tree-structured covariance matrix, which completely characterizes the treatment tree (Proposition 1 and Lemma 1).

(b) *Efficient two-stage algorithm for posterior sampling*. Motivated by the form of marginal data likelihood in (a), we decouple the Euclidean and tree parameters and propose a two-stage algorithm that combines an approximate Bayesian computation (ABC) procedure (for Euclidean parameters) with a Metropolis-Hasting (MH) step (for tree parameters). We demonstrate via multiple simulation studies the superiority of our hybrid approach over approaches based on classical single-stage MH algorithms (Sections 4.2 and 4.1).

(c) *Corroborating existing, and uncovering new, synergistic combination therapies.* We define and infer a new similarity measure that accounts for inherent uncertainty in estimating a latent hierarchy among treatments. As a result, the *maximum a posteriori* R_x-tree and the related mechanism similarity show high concordance with known existing biological mechanisms for monotherapies and uncover new and potentially useful combination therapies (Sections 5.3 and 5.4).

Of particular note is contribution (c), where we leverage a recently collated PDX dataset from the Novartis Institutes for BioMedical Research - PDX Encyclopedia [NIBR-PDXE, (Gao et al., 2015)] that interrogated multiple targeted therapies across five different cancers. Our pan-cancer analyses of the NIBR-PDXE dataset show a high degree of concordance with known existing biological mechanisms across different cancers; for example, a high mechanistic similarity is suggested between two agents currently in clinical trials: CGM097 and HDM201 in breast cancer and colorectal cancer, known to target the same gene MDM2 (Konopleva et al., 2020). In addition, our model uncovers new and potentially effective combination therapies. For example, exploiting knowledge of the combination therapy of a class of agents targeting the PI3K-MAPK-CDK pathway axes – PI3K-CDK for breast cancer, PI3K-ERBB3 for colorectal cancer and BRAF-PI3K for melanoma – confers possible synergistic regulation for prioritization in future clinical studies.

The rest of the paper is organized as follows: we first review our probabilistic formulation for PDX data based on the DDT model and present the marginal data likelihood and computational implications in Section 2. In Section 3, we derive the posterior inference algorithm based on a two-stage algorithm. In Section 4, we conduct two sets of simulations to evaluate the operating characteristics of the model and algorithm. A detailed analysis of the NIBR-PDXE dataset, results, biological interpretations and implications are summarized in Section 5. The paper concludes by discussing implications of the findings, limitations, and future directions.

2. Modeling R_x-tree via Dirichlet Diffusion Trees Given a PDX experiment with I correlated treatments and J independent patients, we focus on the setting with $1 \times 1 \times 1$ design (one animal per PDX model per treatment) with no replicate response for each treatment and patient. A PDX experiment produces an observed data matrix $\mathbf{X}_{I\times J} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_I]^\mathsf{T}$ where $\boldsymbol{X}_i = [X_{i1}, \ldots, X_{iJ}]^\mathsf{T}$ is data under treatment i across J patients; let the observed response column for each patient be $\boldsymbol{X}_{\cdot,j} = [x_{1j}, \ldots, x_{Ij}]^\mathsf{T} \in \mathbb{R}^I, j = 1, \ldots, J$.

In this paper, the observed treatment responses are continuous and we model the responses through a generative model that results in a Gaussian likelihood with a structured covariance:

$$\boldsymbol{X}_{\cdot,j} | \boldsymbol{\Sigma}^{\mathcal{T}}, \sigma^2 \overset{iid}{\sim} \mathsf{N}_I(0, \boldsymbol{\Sigma}^{\mathcal{T}}), \quad j = 1, \ldots, J, \tag{1}$$

where the $\boldsymbol{\Sigma}^{\mathcal{T}}$ is a tree-structured covariance matrix that encodes the tree $\mathcal{T}$. In particular, $\boldsymbol{\Sigma}^{\mathcal{T}} = \left\{ \boldsymbol{\Sigma}^{\mathcal{T}}_{i,i'}, i, i' = 1, \ldots, I \right\}$ encodes the tree $\mathcal{T}$ through two constraints (Lapointe and Legendre, 1991; McCullagh, 2006):

$$\boldsymbol{\Sigma}^{\mathcal{T}}_{i',i} = \boldsymbol{\Sigma}^{\mathcal{T}}_{i,i'} \geq 0; \boldsymbol{\Sigma}^{\mathcal{T}}_{i,i} \geq \boldsymbol{\Sigma}^{\mathcal{T}}_{i,i'}, \tag{2}$$

$$\boldsymbol{\Sigma}^{\mathcal{T}}_{i,i'} \geq \min\{\boldsymbol{\Sigma}^{\mathcal{T}}_{i,i''}\ ,\ \boldsymbol{\Sigma}^{\mathcal{T}}_{i',i''}\} \text{ for all } i \neq i' \neq i''. \tag{3}$$

Each element $\boldsymbol{\Sigma}^{\mathcal{T}}_{i,i}$ is the covariance between treatments i and i' and measures their similarity. The inequality (2) imposes the symmetry of covariance matrix and ensures the divergence of all leaves. The tree structure is characterized by the ultrametric inequality (3) that ensures $\boldsymbol{\Sigma}^{\mathcal{T}}$ bijectively maps to a tree $\mathcal{T}$; for more details on the relationship between the covariance $\boldsymbol{\Sigma}^{\mathcal{T}}$ and the tree $\mathcal{T}$ see McCullagh (2006) and Bravo et al. (2009). Of note, mean parameterized

models (e.g. mixed effects models) are inappropriate for uncovering the tree parameter under the given data structure since the latent tree is completely encoded in covariance matrix $\mathbf{\Sigma}^{\mathcal{T}}$.

A Bayesian formulation requires an explicit prior distribution on $\mathbf{\Sigma}^{\mathcal{T}}$ which satisfies constrains (2) and (3); this requirement is far from straightforward since the set of tree-structured matrices is complicated (e.g., it is not a manifold (McCullagh, 2006)). We instead consider the Dirichlet Diffusion tree (DDT) model (Neal, 2003) for hierarchically clustered data which provides two useful ingredients:

1. a prior is implicitly specified on the latent treatment tree, comprising the root, internal nodes, leaves, and branch lengths;
2. upon integrating out the internal nodes, a tractable Gaussian likelihood on PDX data with tree-structured covariance is specified.

We first provide a brief description of the DDT model proposed by Neal (2003) and its joint density on data and tree (Section 2.1). Subsequently, we derive an expression for the likelihood and demonstrate how it can be profitably employed to develop a generative model for PDX data and carry out $\mathrm{R_x}$-tree estimation (Section 2.2 and 2.3).

2.1. *The Generative Process of DDT* The DDT prescribes a fragmentary, top-down mechanism to generate a binary tree (acyclic graph with a preferred node or vertex referred to as the root), starting from a root, containing J-dimensional observed responses $\boldsymbol{X}_i$ at I leaves/terminal nodes; each node in the tree has either 0 or 2 children excepting the root which has a solitary child. This prescription manipulates dynamics of a system of I independent Brownian motions $B_1, \ldots, B_I$ on $\mathbb{R}^J$ in a common time interval $t \in [0, 1]$. As shown in Figure 2(A), all Brownian motions $B_i(t)$ start at the same point at time $t = 0$, location of which is the root $\mathbf{0} \in \mathbb{R}^J$, and diverge at time points in $[0, 1]$ and locations in $\mathbb{R}^J$ before stopping at the time $t = 1$ at locations $\boldsymbol{X}_i$. The Brownian trajectories and their divergences engender the tree structure as shown in Figure 2(A).

Specifics on when and how the Brownian motions diverge are as follows: the first Brownian motion $B_1(t)$ starts at $t = 0$ and generates $\boldsymbol{X}_1$ at $t = 1$; a second independent Brownian motion $B_2(t)$ starts at the same point at $t = 0$, branches out from the first Brownian motion at some time t, after which it generates $\boldsymbol{X}_2$ at time 1. The probability of divergence in a small interval $[t, t + dt]$ is given by a *divergence function* $t \mapsto a(t)$, assumed as in Neal (2003) to be of the form $a(t) = c(1 - t)^{-1}$ for some divergence parameter $c > 0$. Inductively then, the vector of observed responses to treatment i, $\boldsymbol{X}_i$, is generated by $B_i(t)$, which follows the path of previous ones. If at time t, $B_i(t)$ has not diverged and meets the previous divergent point, it will follow one of the existing path with the probability proportional to the number of data points that have previously traversed along each path. Eventually, given $B_i(t)$ has not diverged at time t, it will do so in $[t, t + dt]$ with probability $a(t)dt/m$, where m is the number of data points that have previously traversed the current path.

From the illustration in panel (A) of Figure 2, we note that B_3 diverges from the B_1 and B_2 at time t_1 at location $\boldsymbol{X}_1'$ and at $t = 1$ is at location $\boldsymbol{X}_3$, which is the J-dimensional response vector for treatment 3; this creates a solitary branch of length t_1 from the root and an unobserved internal node at location $\boldsymbol{X}_1'$. Continuing, given three Brownian motions B_1, B_2 and B_3, B_4 does not diverge before t_1 and meet the previous divergent point t_1. B_4 chooses to follow the path of B_3 with probability $1/3$ at t_1 and finally diverges from B_3 at time $t_2 > t_1$ at location $\boldsymbol{X}_2'$; this results in observation $\boldsymbol{X}_4$ for treatment 4 and an unobserved internal node at $\boldsymbol{X}_2'$, and so on. As a consequence, the binary tree that arises from the DDT comprises of:

(i) an unobserved root at the origin in $\mathbb{R}^J$ at time $t = 0$;
(ii) observed data $\mathbf{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_I]^{\mathsf{T}} \in \mathbb{R}^{I \times J}$ situated at the leaves of the tree;

(iii) unobserved internal nodes $\mathbf{X}^I = [\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_{I-1}]^\mathsf{T} \in \mathbb{R}^{(I-1)\times J}$;
(iv) unobserved times $\boldsymbol{t} = (t_1, \ldots, t_{I-1})^\mathsf{T} \in [0,1]^{I-1}$ that characterize lengths of branches;
(v) unobserved topology $\mathcal{T}$ that links (i)-(iv) into a tree structure, determined by the number of data points $\boldsymbol{X}_i$ that have traversed through each segment or branch.

Conceptually, observed data at the leaves $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_I$ collectively form the observed PDX responses generated through a process involving a few parameters: tree-related parameters $(\mathcal{T}, \boldsymbol{t})$ and the locations of internal nodes $\boldsymbol{X}'_i$. The tree $\mathcal{T}$ clusters I treatments as a hierarchy of $(I-1)$ levels (excluding the last level containing leaves). At level $0 < d \leq I-1$ of the hierarchy, characterized by the pair $(\boldsymbol{X}'_d, t_d)$, the I treatments are clustered into $d+1$ groups; a measure of similarity (or dissimilarity) between treatment clusters at levels d and $d+1$ is given by the branch length $t_{d+1} - t_d$.

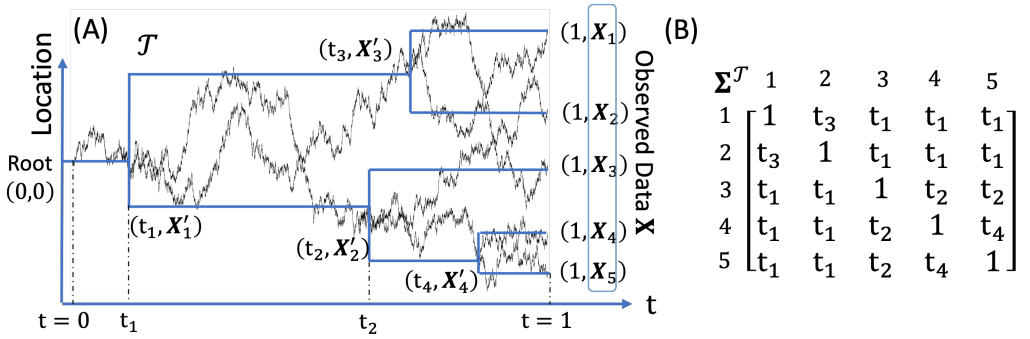

Figure 2: (A) A binary tree with $I = 5$ leaves underlying the diffusion dynamics. The observed response vector $\boldsymbol{X}_i, i = 1, \ldots, I$ is generated by the Brownian motion up to $t = 1$. The unobserved response vector $\boldsymbol{X}'_d, d = 1, \ldots, (I-1)$ at the divergence is generated by the Brownian motion at time t_d. (B) A tree-structured matrix $\boldsymbol{\Sigma}^{\mathcal{T}}$ that encapsulates the tree $\mathcal{T}$. See the Proposition 1 for the definition of $\boldsymbol{\Sigma}^{\mathcal{T}}$.

We now give a brief description of how the joint density of $(\mathbf{X}, \mathbf{X}^I, \boldsymbol{t}, \mathcal{T})$ can be derived; for more details we direct the reader to Neal (2003) and Knowles and Ghahramani (2015). For a fixed $c > 0$ that governs the divergence function $a(t) = c(1-t)^{-1}$, probabilities associated with the independent Brownian motions $B_1, \ldots, B_I$ induce a joint (Lebesgue) density on the generated tree. Note that the binary tree arising from the DDT is encoded by the triples $\{(t_d, \boldsymbol{X}'_d, \boldsymbol{X}_i), d = 1, \ldots, I-1; i = 1, \ldots, I\}$. An internal node at $\boldsymbol{X}'_d$ contains l_d and r_d leaves below to its left and right with $m_d = l_d + r_d$. If each of the Brownian motions is scaled by $\sigma^2 > 0$, then given $\mathcal{T}$ and a branch with endpoints $(t_u, \boldsymbol{X}'_u)$ and $(t_v, \boldsymbol{X}'_v)$ with $0 < t_u < t_v < 1$, from properties of a Brownian motion we see that $\boldsymbol{X}'_v \sim N_J(\boldsymbol{X}'_u, \sigma^2(t_v - t_u)\boldsymbol{I}_J)$, and the (Lebesgue) density of $\mathcal{T}$ can be expressed as the product of contributions from its branches. Then the joint density of all nodes, times and the tree topology is given by

(4)
$$P(\mathbf{X}, \mathbf{X}^I, \boldsymbol{t}, \mathcal{T} | c, \sigma^2) = \Pi_{[u,v]\in\mathcal{S}(\mathcal{T})} \frac{(l_v - 1)!(r_v - 1)!}{(l_v + r_v - 1)!} c(1-t_v)^{cJ_{l_v, r_v}-1} N_J(\boldsymbol{X}'_u, \sigma^2(t_v - t_u)\boldsymbol{I}_J)$$

where $\mathcal{S}(\mathcal{T})$ is the collection of branches and $\mathbf{X}^I_{(I-1)\times J} = [\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_{(I-1)}]^\mathsf{T}$ are unobserved locations of the internal nodes. On each branch $[u, v]$, the first term $\frac{(l_v-1)!(r_v-1)!}{(l_v+r_v-1)!}$ represents the chance the branch containing l_v and r_v leaves to its left and right respectively; $c(1-t_v)^{cJ_{l_v,r_v}-1}$ represents the probability of diverging at t_v with l_v and r_v leaves, where $J_{l+r} = H_{l_v+r_v-1} - H_{l_v-1} - H_{r_v-1}$ with $H_n = \sum_{i=1}^n 1/i$ is the nth harmonic number.

The joint density is hence parameterized by (c, σ^2), where c plays a crucial role in determining the topology $\mathcal{T}$: through the divergence function $a(t)$, it determines the propensity of the Brownian motion to diverge from its predecessors; consequently, a small c engenders later divergence and a higher degree of similarity among treatments in PDX. The latent tree has two components: (i) topology $\mathcal{T}$ and (ii) vector of divergence times $\boldsymbol{t}$ determining branch lengths. We refer to (c, σ^2) as the *Euclidean parameters* and $(\mathcal{T}, \boldsymbol{t})$ as *tree parameters*.

2.2. *Prior on tree and closed-form likelihood* The joint density in (4) factors into a prior $P(\boldsymbol{t}, \mathcal{T}|c, \sigma^2)$ on the tree parameter through $(\mathcal{T}, \boldsymbol{t})$ and a density $P(\mathbf{X}, \mathbf{X}^I|\boldsymbol{t}, \mathcal{T}, c, \sigma^2)$ that is a product of J-dimensional Gaussians on the internal nodes and leaves. The prior distribution on the latent tree is thus implicitly defined through the Brownian dynamics and is parameterized by $(\mathcal{T}, \boldsymbol{t})$ with hyperparameters (c, σ^2). In (4) the product is over the set of branches $\mathcal{S}(\mathcal{T})$, and the contribution to the prior $P(\mathcal{T}, \boldsymbol{t}|c, \sigma^2)$ from each branch $[u, v]$ is $\frac{(l_v-1)!(r_v-1)!}{(l_v+r_v-1)!}c(1-t_v)^{cJ_{l_v,r_v}-1}$, which is free of σ^2; on the other hand, the contribution to $P(\mathbf{X}, \mathbf{X}^I|\boldsymbol{t}, \mathcal{T}, c, \sigma^2)$ from $[u, v]$ is the J-dimensional $N_J(\boldsymbol{X}'_u, \sigma^2(t_v - t_u)\boldsymbol{I}_J)$, which is independent of c. The likelihood function based on the observed $\mathbf{X}$ is thus obtained by integrating out the unobserved internal nodes $\mathbf{X}^I$ from $P(\mathbf{X}, \mathbf{X}^I|\boldsymbol{t}, \mathcal{T}, \sigma^2)$. Accordingly, our first contribution is to derive a closed-form likelihood function for efficient posterior computations; to our knowledge, this task is currently achieved only through sampling-based or variational methods (Neal, 2003; Knowles and Ghahramani, 2015).

Denote as $\mathsf{MN}_{I\times J}(M, U, V)$ the matrix normal distribution of an $I \times J$ random matrix with mean matrix M, row covariance U, and column covariance V, and let $\boldsymbol{I}_k$ denote the $k \times k$ identity matrix. Evidently, $\mathbf{X}$ follows a matrix normal distribution since Gaussian laws of the Brownian motions imply that $[\mathbf{X}, \mathbf{X}^I] = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_I, \boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_{(I-1)}]^\mathsf{T}$ follow a matrix normal distribution.

PROPOSITION 1. *Under the assumption that the root is located at the origin in $\mathbb{R}^J$, the data likelihood $\mathbf{X}|\sigma^2, \mathcal{T}, \boldsymbol{t} \sim \mathsf{MN}_{I\times J}(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}^\mathcal{T}, \boldsymbol{I}_J)$, where $\boldsymbol{\Sigma}^\mathcal{T} = \left(\boldsymbol{\Sigma}^\mathcal{T}_{i,i'}\right)$ is an $I \times I$ tree-structured covariance matrix satisfying* (2) *and* (3) *with $\boldsymbol{\Sigma}^\mathcal{T}_{i,i} = 1$ and $\boldsymbol{\Sigma}^\mathcal{T}_{i,i'} = t_d$, for $i \neq i'$ where $i, i' = 1, \ldots, I$ and $d = 1, \ldots, I-1$.*

Proposition 1 asserts that use of the DDT model leads to a centered Gaussian likelihood on PDX data $\boldsymbol{X}$ with a tree-structured covariance matrix. Proposition 1 also implies that each patient independently follows the normal distribution of (1) with an additional scale parameter (σ^2) from the Brownian motion:

$$\boldsymbol{X}_{\cdot,j}|\boldsymbol{\Sigma}^\mathcal{T}, \sigma^2 \overset{iid}{\sim} \mathsf{N}_I(0, \sigma^2\boldsymbol{\Sigma}^\mathcal{T}), \quad j = 1, \ldots, J. \tag{5}$$

By setting $\boldsymbol{\Sigma}^\mathcal{T}_{i,i'} = t_{i,i'}$ as the divergence time of i and i', $\boldsymbol{\Sigma}^\mathcal{T}$ satisfies (2) and (3) and encodes the tree $\mathcal{T}$. For example, consider a three-leaf tree with $\boldsymbol{\Sigma}^\mathcal{T}_{i,i'} = t_{i,i'}$, inequality (3) implies that for the three leaves, say, i, i' and i'', one of the following conditions must hold: (i) $t_{i',i''} \geq t_{i,i'} = t_{i,i''}$; (ii) $t_{i,i''} \geq t_{i,i'} = t_{i',i''}$; (iii) $t_{i,i'} \geq t_{i,i''} = t_{i',i''}$. We then obtain a tree containing 1) a subtree of two leaves with a higher similarity and 2) a singleton clade with a lower similarity between the singleton leaf and the two leaves in the first subtree. In particular, if $t_{i',i''} \geq t_{i,i'} = t_{i,i''}$ holds, the three-leaf tree has leaf i diverging earlier before the subtree of (i', i'').

2.3. *Decoupling Tree and Euclidean Parameters for Efficient Sampling.* In the full joint density in (4) the Euclidean and tree parameters are confounded across row and column dimensions of $\mathbf{X}$, and this may result in slow mixing of chains using traditional MCMC algorithms (Turner et al., 2013). State-of-the-art posterior inference on $(c, \sigma^2, \mathcal{T}, \boldsymbol{t})$ can be broadly classified into sampling-based approaches (e.g., Knowles and Ghahramani, 2015) and deterministic approaches based on variational message passing (e.g., Knowles, Gael and Ghahramani, 2011, VMP). Variational algorithms can introduce approximation errors to the joint posterior via factorization assumptions (e.g.,mean-field) and choice of algorithm is typically determined by the speed-accuracy trade-off tailored for particular applications. On the other hand, in classical MCMC-based algorithms for DDT we observed slow convergence in the sampling chains for c and σ^2 with high autocorrelations for the corresponding chains, owing to possibly the high mutual dependence between c in the divergence function and the tree topology $\mathcal{T}$, resulting in slow local movements in the joint parameter space of model and tree parameters (Simulation II in Section 4.2).

Notwithstanding absence of the parameter c in the Gaussian likelihood, the dependence, and information about, c is implicit: the distribution of divergence times $\boldsymbol{t}$ that populate $\boldsymbol{\Sigma}^{\mathcal{T}}$ are completely determined by the divergence function $t \mapsto c(1-t)^{-1}$. In other words, c can indeed be estimated from treatment responses $\{\boldsymbol{X}_{\cdot,j}\}$ using the likelihood. From a sampling perspective, however, form of the likelihood obtained by integrating out the internal nodes $\mathbf{X}^I$, suggests an efficient two-stage sampling strategy that resembles the classical collapsed sampling (Liu, 1994) strategy in MCMC literature: first draw posterior samples of (c, σ^2) and then proceed to draw posterior samples of $(\mathcal{T}, \boldsymbol{t})$ conditioned on each sample of (c, σ^2).

3. $\mathbf{R_x}$-tree Estimation and Posterior Inference In line with the preceding discussion, we consider a two-stage sampler for Euclidean and tree parameters. While in principle MCMC techniques could be used in both stages, we propose to use a hybrid ABC-MH algorithm. Specifically, we use an approximate Bayesian computation (ABC) scheme to draw weighted samples of (c, σ^2) in the first stage followed by a Metropolis-Hastings (MH) step that samples $(\mathcal{T}, \boldsymbol{t})$ given ABC samples of (c, σ^2) in the second stage. Motivation for using ABC in the first stage stems from: (i) availability of informative statistics; (ii) generation of better quality samples of the tree (compared to a single-stage MH); and (iii) better computational efficiency. We refer to Section 4.2 for more details.

3.1. *Hybrid ABC-MH Algorithm* ABC is a family of inference techniques that are designed to estimate the posterior density $\text{pr}(\theta|\mathcal{D})$ of parameters θ given data $\mathcal{D}$ when the corresponding likelihood $\text{pr}(\mathcal{D}|\theta)$ is intractable but fairly simple to sample from. Summarily, ABC approximates $\text{pr}(\theta|\mathcal{D})$ by $\text{pr}(\theta|\boldsymbol{S}_{obs})$ where $\boldsymbol{S}_{obs}$ is a d-dimensional summary statistic that ideally captures most information about θ. In the special case where $\boldsymbol{S}_{obs}$ is a vector of sufficient statistics, it is well known that $\text{pr}(\theta \mid \mathcal{D}) = \text{pr}(\theta \mid \boldsymbol{S}_{obs})$. To generate a sample from the partial posterior distribution $\text{pr}(\theta \mid \boldsymbol{S}_{obs})$, ABC with rejection sampling proceeds by: (i) simulating N^{syn} values $\theta_l, l = 1, \ldots, N^{\text{syn}}$ from the prior distribution $\text{pr}(\theta)$; (ii) simulating datasets D_l from $\text{pr}(\mathcal{D}|\theta_l)$; (iii) computing summary statistics $\boldsymbol{S}_l, l = 1, \ldots, N^{\text{syn}}$ from $\mathcal{D}_l$; (iv) retaining a subset of $\{\theta_{l_s}, s = 1 \ldots, k\}$ of size $k < N^{\text{syn}}$ that corresponds to 'small' $\|\boldsymbol{S}_{l_s} - \boldsymbol{S}_{obs}\|$ values based on some threshold. Given pairs $\{(\theta_{l_s}, \boldsymbol{S}_{l_s})\}$, the task of estimating the partial posterior translates to a problem of conditional density estimation, e.g., based on Nadaraya-Waston type estimators and local regression adjustment variants to correct for the fact that $\boldsymbol{S}_{l_s}$ may not be exactly $\boldsymbol{S}_{obs}$; see Sisson, Fan and Beaumont (2018) for a comprehensive review. To implement ABC, the choice of summary statistics is central.

We detail the specialization of ABC to the marginal posterior distributions of c and σ^2 in Section 3.1.1. Given any pair of (c, σ^2), we can sample trees from a density function

up to an unknown normalizing constant based on an existing MH algorithm (Knowles and Ghahramani, 2015). Our proposal is to condition on the posterior median of (c, σ^2) of ABC-weighted samples from the first stage, when sampling the trees in the second stage; clearly, other choices are also available. This strategy produced comparable MAP trees and inference of other tree-derived results relative to tree samples based on full ABC samples of c and σ^2.

Pseudo code for the two-stage algorithm is presented in the Supplementary Material Algorithm S1. We briefly describe below its key components.

3.1.1. *Stage 1: Sampling Euclidean Parameters (c, σ^2) using ABC* Accuracy and efficiency of the ABC procedure is linked to two competing desiderata on the summary statistics: (i) informative, or ideally sufficient; (ii) low-dimensional.

Summary statistic for σ^2. From the closed-form likelihood in Equation (5), a sufficient statistic of $\sigma^2\boldsymbol{\Sigma}^{\mathcal{T}}$ is easily available, using which we construct a summary statistics for σ^2.

LEMMA 1. *With $\mathbf{X}$ as the observed data, the statistic $\boldsymbol{T} := \sum_j \boldsymbol{X}_{\cdot,j}\boldsymbol{X}_{\cdot,j}^{\mathsf{T}}$ is sufficient for $\sigma^2\boldsymbol{\Sigma}^{\mathcal{T}}$ and follows a Wishart distribution $W_I(J, \sigma^2\boldsymbol{\Sigma}^{\mathcal{T}})$, where $\boldsymbol{X}_{\cdot,j} = [x_{1j}, \ldots, x_{Ij}] \in \mathbb{R}^I$. Then with $S^{(\sigma^2)} := \frac{tr(\boldsymbol{T})}{IJ}$ we have $E[S^{(\sigma^2)}] = \sigma^2$ and $Var[S^{(\sigma^2)}] = \frac{2\sigma^4 tr((\boldsymbol{\Sigma}^{\mathcal{T}})^2)}{I^2J}$.*

Due to the normality of $\mathbf{X}$ in (5), and the Factorization theorem (Casella and Berger, 2001), we see that $\boldsymbol{T}$ is complete and sufficient for $\sigma^2\boldsymbol{\Sigma}^{\mathcal{T}}$ and $\boldsymbol{T} \sim W_I(J, \sigma^2\boldsymbol{\Sigma}^{\mathcal{T}})$. Well-known results about the trace and determinant of $\mathbf{X}$ (see for e.g. Mathai (1980)) provide the stated results on the mean and variance of $tr(\boldsymbol{T})$. Owing to its unbiasedness, we choose $S^{(\sigma^2)} = tr(\boldsymbol{T})/IJ$ as the summary statistic for σ^2 and examine its performance through simulations in Section 4; other choices are assessed in the Supplementary Material Section S4.1.

Summary statistic for c. Based on the matrix normal distribution of Proposition 1, the divergence parameter c does not appear in the observed data likelihood. Any statistic based on the entire observed data set $\mathbf{X}$ is sufficient, but not necessarily informative about c. In DDT, the prior distribution of the vector of branching times $\boldsymbol{t}$ is governed by divergence parameter c via the divergence function $a(t; c)$. Thus an informative summary statistic for c can be chosen by assessing its information about $\boldsymbol{t}$. For example, tighter observed clusters indicate small c (e.g., $c < 1$), where the level of tightness is indicated by the branch lengths from leaves to their respective parents. We construct summary statistics for c based on a dendrogram estimated via hierarchical clustering of $\mathbf{X}$ based on pairwise distances $\delta_{i,i'} := \|\boldsymbol{X}_i - \boldsymbol{X}_{i'}\|, i \neq i'$. The summary statistics $\boldsymbol{S}^{(c)}$ we choose is a ten-dimensional concatenated vector comprising the 10th, 25th, 50th, 75th and 90th percentiles of empirical distribution of: (i) $\delta_{i,i'}$; (ii) branch lengths associated with leaves of the dendrogram. Other candidate summary statistics for c are examined in Supplementary Material Section S4.1.

3.1.2. *Stage 2: Sampling Tree Parameters $(\mathcal{T}, \boldsymbol{t})$ using Metropolis-Hastings* For the second stage, we proceed by choosing a representative value (c_0, σ_0^2) chosen from the posterior sample of (c, σ^2), which in our case is the posterior median. Then a Metropolis-Hastings (MH) algorithm to sample from $\text{pr}((\mathcal{T}, \boldsymbol{t})|c_0, \sigma_0^2, \mathbf{X})$; recall that the $\text{R}_\text{x}$tree is characterized by both the topology $\mathcal{T}$ and divergence times $\boldsymbol{t}$. In particular, after initialization (e.g., the dendrogram obtained from hierarchical clustering), we first generate a candidate tree $(\mathcal{T}', \boldsymbol{t}')$ from the current tree $(\mathcal{T}, \boldsymbol{t})$ in two steps: (i) detaching a subtree from the original tree; (ii) reattaching the subtree back to the remaining tree. Acceptance probabilities for a candidate tree can be computed exactly and directly using the explicit likelihood in (5), without which they would have to be calculated iteratively (Neal, 2003; Knowles and Ghahramani, 2015). See Supplementary Material Section S2.2 for details of the proposal function and the acceptance probabilities.

REMARK 1. In order to use the explicit likelihood in (5) from Proposition 1 to generate observed data $\mathbf{X}$, a tree-structured covariance $\mathbf{\Sigma}^{\mathcal{T}}$ needs to be specified, whose entries in-turn depend on the parameter c through the divergence function. It is not straightforward to fix or sample a $\mathbf{\Sigma}^{\mathcal{T}}$ since its entries need to satisfy the inequalities (3). It is easier to generate data $\mathbf{X}$ directly using the DDT generative mechanism in the ABC stage, and this is the approach we follow and is described in Supplementary Section S2.

Summarily, there are three main advantages to using the explicit likelihood from Proposition 1: (i) decoupling of Euclidean and tree parameters to enable an efficient two-stage sampling algorithm; (ii) direct and exact computation of tree acceptance probabilities in MH stage; (iii) determination of informative sufficient statistic for σ^2 (Lemma 1).

REMARK 2. From the computational aspect, the calculation of the explicit Gaussian likelihood of (5) in Proposition 1 through the matrix decomposition is slower (e.g. Cholesky decomposition with $\mathcal{O}(I^3)$) than the message passing (e.g. the belief propagation with $\mathcal{O}(I)$ (Mezard and Montanari, 2009)) in terms of the big O notation (Knuth, 1976). However, the computation speed also depends on the implementation. For this paper, we implemented our algorithm in `R` and found that the matrix decomposition is faster than the message passing on `R`. We offer more details with a simulation study in Supplementary Material Section S5.3.

3.2. *Posterior Summary of R_x-Tree, $(\mathcal{T}, \boldsymbol{t})$* While quantifying uncertainty concerning the tree parameters $(\mathcal{T}, \boldsymbol{t})$ is of main interest, we note that, from definition of the DDT, this is influenced by uncertainty in the model parameters. In particular, the first stage of ABC-MH produces weighted samples and we calculate the posterior median by fitting an intercept-only quantile regression with weights (see details in the Supplementary Material Section S2.1). For the R_x-tree, we consider global and local tree posterior summaries that capture uncertainty in the latent hierarchy among all and subsets of treatments.

Flexible posterior inference is readily available based on L posterior samples of $(\mathcal{T}, \boldsymbol{t})$ from the MH step. It is possible to construct correspond tree-structured covariance matrices $\mathbf{\Sigma}^{\mathcal{T}}$ from sample $(\mathcal{T}, \boldsymbol{t})$. Instead, we compute:

(a) a global *maximum a posteriori* (MAP) estimate of the R_x-tree that represents the overall hierarchy underlying the treatment responses;
(b) local uncertainty estimates of co-clustering probabilities among a subset $\mathcal{A} \subset \{1, \ldots, I\}$ of treatments based on posterior samples of the corresponding subset of divergence times.

Posterior co-clustering probability functions. We elaborate on the local summary (b). Suppose $\mathcal{A} = \{i, i', i''\}$ consists of three treatments. Given a tree topology $\mathcal{T}$, note that at every $t \in [0, 1]$ a clustering of all I treatments is available and the clustering changes only at times $0 < t_1 < \cdots < t_{I-1}$. Consequently, for a given tree topology $\mathcal{T}$ drawn from its posterior, we can compute for every level $t \in [0, 1]$ a posterior probability that i, i' and i'' belong to the same cluster. Such a posterior probability can be approximated using Monte Carlo on the L posterior samples. Accordingly, we define the estimated posterior co-clustering probability (PCP) function associated with $\mathcal{A}$ as,

$$\text{PCP}_{\mathcal{A}}(t) = \frac{\sum_{l=1}^{L} \mathbb{I}_{[0, t^{(l)}_{i,i',i''})}(t)}{L},$$

where $\mathbb{I}_B$ is the indicator function on the set B and $t^{(l)}_{i,i',i''}$ is the divergence time of $\mathcal{A} = \{i, i', i''\}$ in the l-th tree sample. Essentially, the $\text{PCP}_{\mathcal{A}}(t)$ can be viewed as the proportion of tree samples with $\{i, i', i''\}$ having the most recent common ancestor later than t.

For every subset $\mathcal{A}$, the function $[0, 1] \ni t \mapsto \text{PCP}_{\mathcal{A}}(t) \in [0, 1]$ is non-increasing starting at 1 and ending at 0, and reveals propensity among treatments in $\mathcal{A}$ to cluster as one traverses

down an (estimate of) R_x-tree starting at the root: a curve that remains flat and drops quickly near 1 indicates higher relative similarity among the treatments in $\mathcal{A}$ relative to the rest of the treatments. A scalar summary of $\text{PCP}_{\mathcal{A}}(t)$ is the area under its curve known as integrated PCP $\text{iPCP}_{\mathcal{A}}$, which owing to the definition of $\text{PCP}_{\mathcal{A}}(t)$, can be interpreted as the expected (or average) chance of co-clustering for treatments in $\mathcal{A}$.

Figure 3 illustrates an example of a three-way $\text{iPCP}_{\mathcal{A}}$ with $\mathcal{A} = \{i, i', i''\}$ for a PDX data with I treatments and J patients (Figure 3(A)). Given $L = 3$ posterior trees samples (Figure 3(B)) drawn from the PDX data, we first calculate the whole $\text{PCP}_{\mathcal{A}}(t)$ function by moving the time t from 0 to 1. Starting from time $t = 0$, no treatment diverges at time $t = 0$ and the $\text{PCP}_{\mathcal{A}}(t)$ is 1. At time t', treatments diverge in one out of the three posterior trees and $\text{PCP}_{\mathcal{A}}(t)$ therefore drops from 1 to $2/3$. Moving the time toward $t = 1$, treatments diverge in all trees and the $\text{PCP}_{\mathcal{A}}(t)$ drops to 0. The $\text{iPCP}_{\mathcal{A}}$ then can be obtained by the area under the $\text{PCP}_{\mathcal{A}}(t)$.

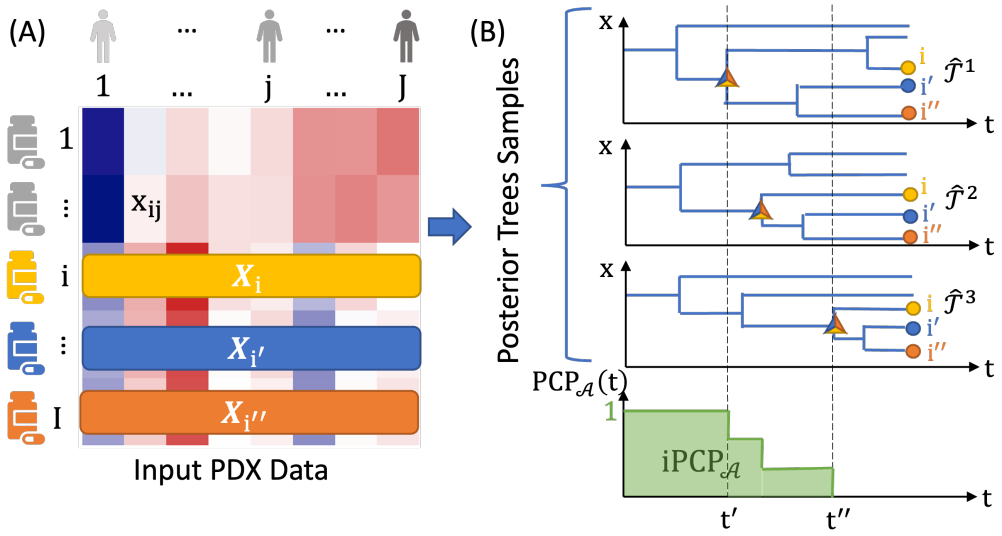

Figure 3: Posterior tree summaries. (A) The input PDX data with I treatments and J patients, and treatments $\mathcal{A} = \{i, i', i''\}$ are of interest. (B) $\text{PCP}_{\mathcal{A}}(t)$ and $\text{iPCP}_{\mathcal{A}}$ for treatments $\mathcal{A}$ based on $L = 3$ posterior trees. The relevant divergence times are represented by a "△" in each posterior tree sample. For example, at time t', the treatments in $\mathcal{A}$ diverge in one out of the three trees. Because $\text{PCP}_{\mathcal{A}}(t')$ is defined by the proportion of posterior tree samples in which $\mathcal{A}$ has *not* diverged up to and including t', it drops from 1 to $2/3$.

REMARK 3. In the special case of $\mathcal{A} = \{i, i'\}$ for two treatments, the definition of $\text{iPCP}_{\mathcal{A}}$ can be related to the cophenetic distance (Sokal and Rohlf, 1962; Cardona et al., 2013) and, moreover, extends definition of the cophenetic distance to multiple trees. Given two treatments i and i' in a single tree, let t_d be the time at which their corresponding Brownian paths diverge. Then $\text{PCP}_{\mathcal{A}}(t) = \mathbb{I}_{[0,t_d)}(t)$ and $\text{iPCP}_{\mathcal{A}} = t_d$; this implies that the cophenetic distance is $2(1 - t_d)$ and thus $\text{iPCP}_{\mathcal{A}}$ and the cophenetic distances uniquely determines the same tree structure. For $L > 1$ trees, a Carlo average of divergence times of L trees leads to the corresponding $\text{iPCP}_{\mathcal{A}}$.

REMARK 4. Given I treatments, since pairwise cophenetic distances from one tree determines a tree (Lapointe and Legendre, 1991; McCullagh, 2006), one might consider summarizing and represent posterior trees in terms of an $I \times I$ matrix $\mathbf{\Sigma}$ consisting of entries $\text{iPCP}_{\{i,i'\}}$ for every pair of treatments of (i, i'), estimated from the posterior sample of trees. However, $\mathbf{\Sigma}$ need not to be a tree-structured matrix that uniquely encodes a tree. It is possible to project $\mathbf{\Sigma}$ on to the space of tree-structured matrices (see for e.g., Bravo et al. (2009)) but the projection might result in a non-binary tree structure. We discuss this issue and its resolution in Supplementary Material Section S3.

4. Simulations Accurate characterization of similarities among any subset of treatments is central to our scientific interest in identifying the promising treatment subsets for further investigation. In addition, we have introduced a two-stage algorithm to improve our ability to efficiently draw tree samples from the posterior distribution (similarly for the Euclidean parameters). To demonstrate the modeling and computational advantages, we conduct two sets of simulations. The first simulation shows that the proposed model estimates the similarity (via iPCP) better than alternatives, even when the true data generating mechanisms deviate from DDT assumptions in terms of the form of divergence function, prior distribution for the unknown tree, and normality of the responses. The second simulation illustrates the computational efficiency of the proposed two-stage algorithm in producing higher quality posterior samples of Euclidean parameters, resulting in more accurate subsequent estimation of an unknown tree and iPCPs, two key quantities to our interpretation of real data results.

4.1. *Simulation I: Estimating Treatment Similarities* We first show that iPCPs estimated by DDT are closer to the true similarities (operationalized by functions of elements in the true divergence times in $\mathbf{\Sigma}^{\mathcal{T}}$) under different true data generating mechanisms that may follow or deviate from the DDT model assumptions in three distinct aspects (the form of divergence function, the prior distribution over the unknown tree, and normality).

Simulation setup. We simulate data by mimicking the PDX breast cancer data (see Section 5) with $I = 20$ treatments and $J = 38$ patients. We set the true scale parameter as the posterior median σ_0^2 and the true tree $\mathcal{T}_0$ as the MAP tree that are estimated from the breast cancer data; We consider four scenarios to represent different levels of deviation from the DDT model assumptions:

(i) No deviation of the true data generating mechanism from the fitted DDT models: given σ_0^2 and $\mathcal{T}_0$, simulate data based on the DDT marginal data distribution (Equation (5));

The true data generating mechanism deviates from the fitted DDT in terms of:

(ii) divergence function: same as in (i), but the true tree is a random tree from DDT with misspecified divergence function, $a(t; r) = \frac{r}{(1-t)^2}, r = 0.5$;

(iii) prior for tree topology: same as in (i), but the true tree is a random tree from the coalescence model (generated by function `rcoal` in R package `ape`), and,

(iv) marginal data distribution: same as in (i), but the marginal likelihood is a centered multivariate t distribution with degree-of-freedom four and scaled by $\sigma_0^2 \mathbf{\Sigma}^{\mathcal{T}_0}$.

For each of four true data generating mechanisms above, we simulate $B = 50$ replicate data sets. In the following, we use the DDT model and the two-stage algorithm for all estimation regardless of the true data generating mechanisms. For DDT, we ran the two-stage algorithm where the second stage is implemented with five parallel chains. For each chain, we ran $10,000$ iterations, discarded first $9,000$ trees and combined five chains with a total of $5,000$ posterior tree samples.

First, we compute the iPCPs for all pairs of treatment combinations following the definition of $\text{iPCP}_{\mathcal{A}}$ where $\mathcal{A} = \{i, i'\}, 1 \leq i < i' \leq I$. Two alternative approaches to defining

and estimating similarities between treatments are considered: (i) similarity derived from agglomerative hierarchical clustering, and (ii) empirical Pearson correlation of the two vectors of responses $\boldsymbol{X}_i$ and $\boldsymbol{X}_{i'}$, for $i \neq i'$. In particular, for (i), we considered five different linkage methods (Ward, Ward's D2, single, complete and Mcquitty) with Euclidean distances. Given an estimated dendrogram from hierarchical clustering, the similarity for a pair of treatments is defined by first normalizing the sum of branch lengths from the root to leaf as 1, and then calculating the area under of the co-clustering curve (AUC) obtained by cutting the dendrogram at various levels from 0 to 1. For three- or higher-way comparisons, (i) can still produce an AUC based on a dendrogram obtained from hierarchical clustering, while the empirical Pearson correlation in (ii) is undefined hence not viable as a comparator beyond assessing pairwise treatment similarities.

Performance metrics. For treatment pairs $\mathcal{A} = \{i, i'\}$, to assess the quality of estimated treatment similarities for each of the methods above (DDT-based, hierarchical-clustering-based, and empirical Pearson correlation), we compare the estimated values against the true branching time $\Sigma^{\mathcal{T}_0}_{i,i'}$; similarly when assessing recovery of three-way treatment similarities, e.g., $\mathcal{A} = \{i, i', i''\}$, $\Sigma^{\mathcal{T}_0}_{i,i',i''}$ is defined as the time when $\{i, i', i''\}$ first branches in the true tree $\mathcal{T}_0$. In particular, for replication data set $b = 1, \ldots, B$, let $\widehat{\Sigma}^{(b)}_{i,i'}$ generically represent the pairwise similarities for treatment subsets (i, i') that can be based on DDT, hierarchical clustering or empirical pairwise Pearson correlation. For three-way comparisons, let $\widehat{\Sigma}^{(b)}_{i,i',i''}$ generically represent the three-way similarities for treatment subset (i, i', i'') that can be based on DDT, or hierarchical clustering.

We assess the goodness of recovery by computing $\sqrt{\sum_{i,i'}(\widehat{\Sigma}^{(b)}_{i,i'} - \Sigma^{\mathcal{T}_0}_{i,i'})^2}$, the Frobenious norm of the matrix in recovering the entire $\Sigma^{\mathcal{T}_0}$. We compute $\max_{i,i',i''} |\widehat{\Sigma}^{(b)}_{i,i',i''} - \Sigma^{\mathcal{T}_0}_{i,i',i''}|$, the max-norm of the matrix in recovering the true three-way similarities. For a given method and treatment subset $\mathcal{A}$, the above procedure results in B values, the distribution of which can be compared across methods; smaller values indicate better recovery of the true similarities.

Alternatively, for each method and each treatment subset, we also compute the Pearson correlation between the estimated similarities and the true branching times across replicates for pairwise or three-way treatment subsets: $\widehat{\text{Cor}}\left((\widehat{\Sigma}^{(b)}_{i,i'}, \Sigma^{\mathcal{T}_0}_{i,i'}), b = 1, \ldots, B = 50\right)$, for treatments $i < i'$ and $\widehat{\text{Cor}}\left((\widehat{\Sigma}^{(b)}_{i,i',i''}, \Sigma^{\mathcal{T}_0}_{i,i',i''}), b = 1, \ldots, B = 50\right)$, for treatments $i < i' < i''$. We refer to this metric as "Correlation of correlations" (the latter uses the fact that the entries in the true $\Sigma^{\mathcal{T}_0}$ being correlations; see Equation (5)); higher values indicate better recovery of the true similarities.

Simulation results. We observe that DDT better estimates the treatment similarities even under misspecified models. In particular, under scenarios where the true data generating mechanisms deviate from the fitted DDT model assumptions (ii-iv), the DDT captures the true pairwise and three-way treatment similarities the best by higher values in correlation of correlations (left panels, Figure 4) and lower matrix/array distances (right panels, Figure 4). In particular, the fitted DDT with divergence function $a(t) = c/(1-t)$ under Scenario i, ii and iii performed similarly well indicating the relative insensitivity to the DDT modeling assumptions with respect to divergence function and the tree generative model. Under Scenario iv where the marginal likelihood assumption deviates from Gaussian with heavier tails, the similarity estimates from all methods deteriorate relative to Scenarios i-iii. Comparing between methods, the similarities derived from hierarchical clustering with single linkage is comparable to DDT model when evaluated by correlation of correlation, but worse than DDT when evaluated by the matrix norm.

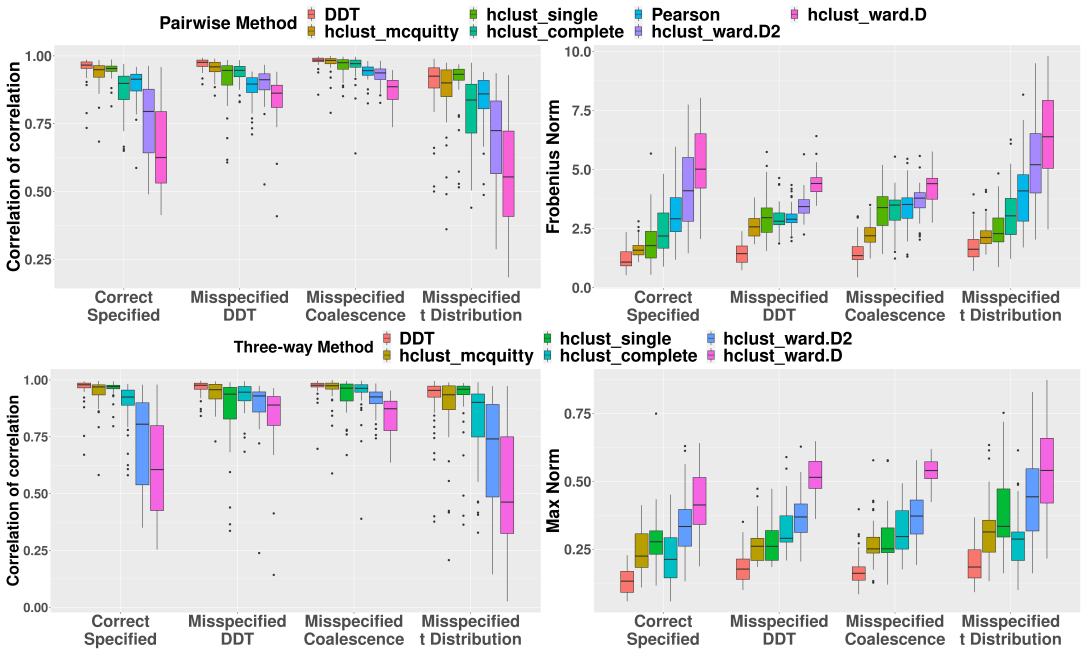

Figure 4: Simulation studies for comparing the quality of estimated treatment similarities based on DDT, hierarchical clustering, and empirical Pearson correlation. Two performance metrics are used: (Left) Correlation of correlation (higher values are better); (Right) Matrix distances with Frobenius norm for pairwise similarity and max norm for three-way similarity (lower values are better). DDT captures both true pairwise (upper panels) and three-way (lower panels) similarity best under four levels of misspecification scenarios.

Additional simulations. Another alternative to bring the information of the posterior samples of c and σ^2 is to use the whole posterior samples instead of the fixed representative statistics only. Following the same set-up, we offer another simulation result to empirically compare the inference performance from the algorithm with the posterior median only and the the whole posterior samples. See more details in Supplementary Material Section S5.4.

4.2. *Simulation II: Comparison with Single-Stage MCMC Algorithms* We have also conducted extensive simulation studies that focus on the computational aspect of the proposed algorithms and demonstrate the advantage of the proposed two-stage algorithm in producing higher quality posterior samples of the unknown tree than classical single-stage MCMC algorithms. In particular, we demonstrate that the proposed algorithm produces (i) MAP trees that are closer to the true tree than alternatives (hierarchical clustering, single-stage MH with default hierarchical clustering or the true tree at initialization) and (ii) more accurate estimation of pairwise treatment similarities compared to single-stage MCMC algorithms. See Supplementary Material Section S5 for further details.

Additional simulations and sensitivity analyses. Aside from the simulations above focusing on the tree structure and the divergence time, Supplementary Material S4 offers additional details for Euclidean parameters including the parameter inference, algorithm diagnostics, and sensitivity analysis for the number of the synthetic data. In particular, we empirically show that current $\boldsymbol{S}^{(c)}$ and $S^{(\sigma^2)}$ outperform other candidate summary statistics in terms of bias in Section S4.1. In Section S4.2, we present additional simulation results that demonstrate that the two-stage algorithm (i) enjoys stable effective sample size (ESS) for (c, σ^2); (ii) leads to similar or better inference on (c, σ^2), as ascertained using credible intervals. In

Section S4.3, we check the convergence of MH and the goodness of fit for ABC. A sensitivity analysis for the number of the synthetic data providing the possible acceleration for ABC is shown in Section S4.4.

5. Treatment Trees in Cancer using PDX Data

5.1. *Dataset Overview and Key Scientific Questions* We leverage a recently collated PDX dataset from the Novartis Institutes for BioMedical Research - PDX Encyclopedia [NIBR-PDXE, (Gao et al., 2015)] that interrogated multiple targeted therapies across different cancers and established that PDX systems provide a more accurate measure of the response of a population of patients than traditional preclinical models. Briefly, the NIBR-PDXE consists of $> 1,000$ PDX lines across a range of human cancers and uses a $1 \times 1 \times 1$ design (one animal per PDX model per treatment); i.e., each PDX line from a given patient was treated simultaneously with multiple treatments allowing for direct assessments of treatment hierarchies and responses. In this paper, we focus on our analyses on a subset of PDX lines with complete responses across five common human cancers: Breast cancer (BRCA), Cutaneous Melanoma (CM, skin cancer), Colorectal cancer (CRC), Non-small Cell Lung Carcinoma (NSCLC), and Pancreatic Ductal Adenocarcinoma (PDAC). After re-scaling data and missing data imputation, different numbers of treatments, I, and PDX models, J, presented in the five cancers were, (I, J): BRCA, $(20, 38)$; CRC, $(20, 40)$; CM, $(14, 32)$; NSCLC, $(21, 25)$; and PDAC, $(20, 36)$. (See Supplementary Material Table S7 for treatment names and Section S6.1 for details of pre-processing procedures.)

In our analysis, we used the best average response (BAR) as the main response, by taking the untreated group as the reference group and using the tumor size difference before and after administration of the treatment(s) following Rashid et al. (2020). Positive values of BAR indicate the treatment(s) shrunk the tumor more than the untreated group with higher values indicative of (higher) treatment efficacy. To apply the Proposition 1, we also checked the distributional assumption for each cancer (see Supplementary Material Section S6.2). The treatments included both drugs administered individually with established mechanisms (referred to as "monotherapy") and multiple drugs combined with potentially unknown synergistic effects (referred to as "combination therapy"). Our key scientific questions were as follows: (a) identify plausible biological mechanisms that characterize treatment responses for monotherapies within and between cancers; (b) evaluate the effectiveness of combination therapies based on biological mechanisms. Due to a potentially better outcome and lower resistance, combination therapy with synergistic mechanism is highly desirable (Bayat Mokhtari et al., 2017).

DDT model setup. For all analyses we followed the setup in the Section 4.1 and obtained $N^{\text{syn}} = 600,000$ synthetic datasets from the ABC algorithm (Section 3.1.1) with prior $c \sim \mathsf{Gamma}(2, 2)$ and $1/\sigma^2 \sim \mathsf{Gamma}(1, 1)$ and took the first 0.5% ($d = 0.5\%$) closest data in terms of $\boldsymbol{S}^{(c)}$ and $S^{(\sigma^2)}$. We calculated the posterior median of (c, σ^2) as described in Section 3.2. For the second-stage MH, we ran five chains of the two-stage algorithm with (c, σ^2) fixed at the posterior median by 10,000 iterations and discarded the first 9,000 trees, which resulted in 5,000 posterior trees in total. Finally, we calculated the $\mathrm{R_x}$-tree (MAP) and iPCP based on 5,000 posterior trees for all subsequent analyses and interpretations. All computations were divided on multiple different CPUs (see the Supplementary Table S5 for the full list of CPUs). For the BRCA data with $I = 20$ and $J = 38$, we divided the ABC stage into 34 compute cores with a total of 141 CPU hours and maximum 4.7 hours in real time. For the MH stage and the single-stage MCMC, we split the computation on 5 compute cores with a total of 8.6 and 12 CPU hours, and a maximum 1.7 and 2.5 hours in real time, respectively.

Our results are organized as follows: we provide a summary of the R_x-tree estimation and treatment clusters in Section 5.2 followed by specific biological and translational interpretations in Sections 5.3 and 5.4 for monotherapy and combination therapy, respectively. Additional results can be accessed and visualized using our companion `R`-shiny application (see Supplementary Material Section S6.6 for details).

5.2. *R_x-Tree Estimation and Treatment Clusters* We focus our discussion on three cancers: BRCA, CRC and CM here – see Supplementary Materials Section S6.5 for NSCLC and PDAC. In Figure 5, R_x-tree, pairwise iPCP and (scaled) Pearson correlation are shown in the left, middle and right panels, respectively. Focusing on the left two panels, we observe that the R_x-tree and the pairwise iPCP matrix show the similar clustering patterns. For example, three combination therapies in CM form a tight subtree and are labeled by a box in the R_x-tree of Figure 5 and a block with higher values of iPCP among three combination therapies also shows up in the corresponding iPCP matrix with a box labeled. In our analysis, the treatments predominantly target six oncogenic pathways that are closely related to the cell proliferation and cell cycle: (i) phosphoinositide 3-kinases, PI3K; (ii) mitogen-activated protein kinases, MAPK; (iii) cyclin-dependent kinases, CDK; (iv) murine double minute 2, MDM2; (v) janus kinase, JAK; (vi) serine/threonine-protein kinase B-Raf, BRAF. We label targeting pathways above for monotherapies with solid dots and further group PI3K, MAPK and CDK due to the common downstream mechanisms (e.g., Repetto et al., 2018; Kurtzeborn, Kwon and Kuure, 2019). Roughly, the R_x-tree from our model clusters monotherapies targeting oncogenic processes above and largely agrees with common and established biology mechanisms. For example, all PI3K-MAPK-CDK inhibitors (solid square) belong to a tighter subtree in three cancers; two MDM2 monotherapies (solid triangle) are closest in both BRCA and CRC. While visual inspection of the MAP R_x-tree agrees with known biology, iPCP further quantifies the similarity by assimilating the information across multiple trees from our MCMC samples. For the ensuing interpretations in Sections 5.3 and 5.4, we focus on iPCP and verify our model through monotherapies with known biology, since our a priori hypothesis is that monotherapies that share the same downstream pathways should exhibit higher iPCP values. Furthermore, we extend our work to identify combination therapies with synergy and discover several combination therapies for each cancer.

5.3. *Biological Mechanisms in Monotherapy* Our estimation procedure exhibits a high level of concordance between known biological mechanisms and established monotherapies for multiple key signalling pathways. From the R_x-tree in Figure 5, aside from the oncogenic process (solid dots) introduced above, monotherapies also target receptors (hollow circles) or other non-kinase targets (e.g. tubulin; crosses). We summarize our key findings and interpretations along with their implications in monotherapy across different cancers for PI3K-MAPK-CDK in this section and list the rest signaling pathways and their regulatory axes, namely, MDM2 from cell cycle regulatory pathways, human epidermal growth factor receptor 3 (ERBB3) from receptor pathways, and tubulin from non-kinase pathways in Supplementary Material Section S6.4. For the following sections, because we wish to conduct fully-exploratory analyses where we do not assume prior knowledge about treatment mechanism, we set the threshold of the co-clustering at the 75-th percentile of all pairwise iPCPs. Specifically, we set the cut-off at 0.753, 0.687 and 0.801 for BRCA, CRC and CM, respectively. See Supplementary Material Section S6.3 for more details about cut-off choices under full and partially exploratory settings related to prior knowledge about monotherapies.

PI3K-MAPK-CDK inhibitors. For treatments targeting PI3K, MAPK and CDK, treatments have the same target share high iPCP. In the NIBR-PDXE dataset, three PI3K inhibitors (BKM120, BYL719 and CLR457), two MAPK inhibitors (binimetinib and CKX620) and

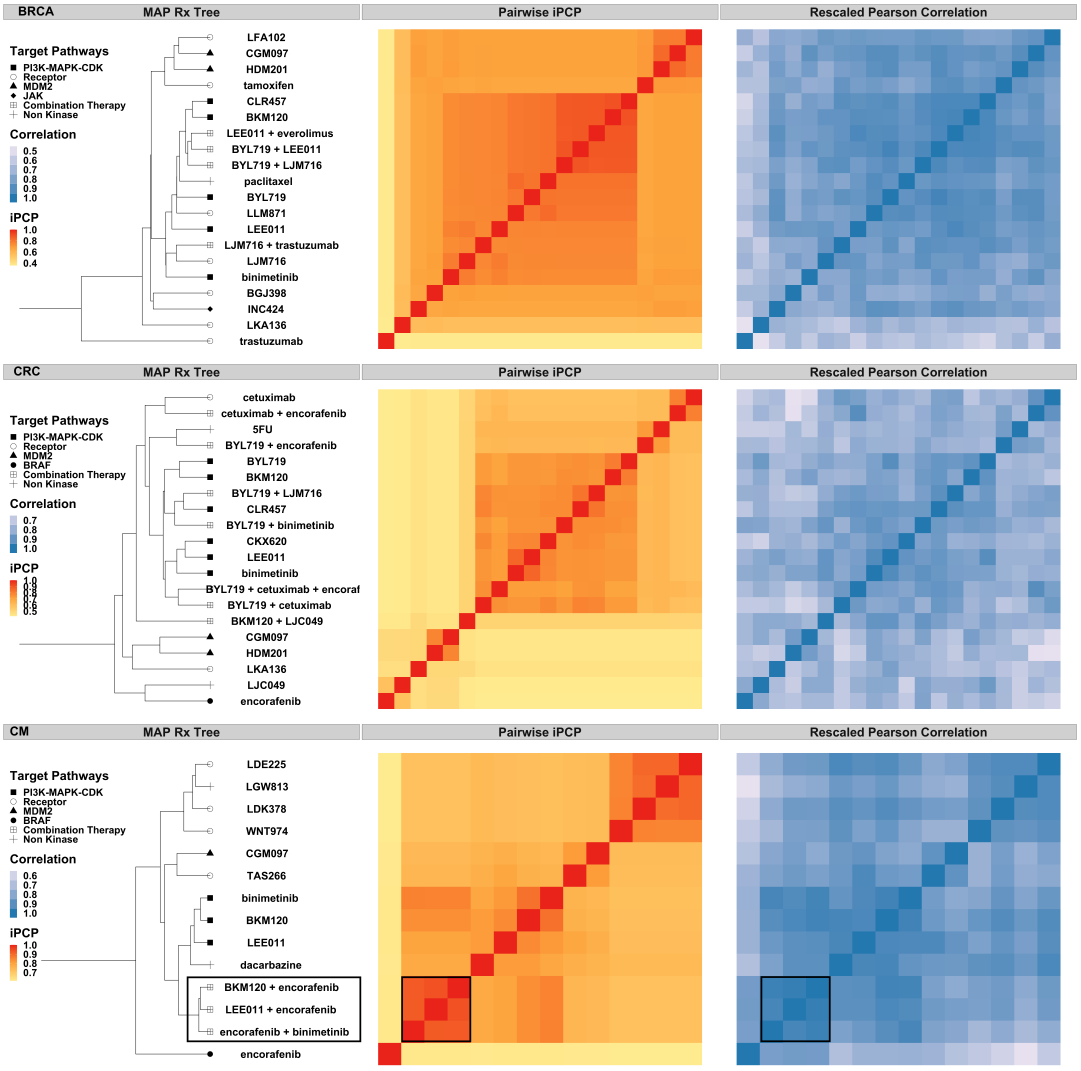

Figure 5: The R_X-tree and iPCP for breast cancer (BRCA, top row), colorectal cancer (CRC, middle row) and melanoma (CM, lower row). Three panels in each row represent: (left) estimated R_X-tree (MAP); distinct external target pathway information is shown in distinct shapes for groups of treatments on the leaves; (middle) estimated pairwise iPCP, i.e., the posterior mean divergence time for pairs of entities on the leaves (see the result paragraph for definition for any subset of entities); (right) scaled Pearson correlation for each pair of treatments. Note that the MAP visualizes the hierarchy among treatments; the iPCP is not calculated based on the MAP, but based on posterior tree samples (see definition in Section 3.2)

one CDK inhibitor (LEE011) were tested, but different cancers contain different numbers of treatments. Specifically, all three PI3K inhibitors present in BRCA and CRC, but only BKM120 is tested in CM; CRC contains two MAPK inhibitors while BRCA and CM only have binimetinib; LEE011 is tested in all three cancers. In Figure 6, BKM120, BYL719 and CLR457 share high pairwise iPCPs (box (1)) and all target PI3K for BRCA and CRC (BRCA, (BKM120, CLR457): 0.8986, (BKM120, BYL719): 0.8002, (BYL719, CLR457): 0.8002; CRC, (BKM120, CLR457): 0.7555, (BKM120, BYL719): 0.8041, (BYL719, CLR457): 0.7597); MAPK (box (2)) inhibitors, binimetinib and CKX620, show a high pairwise iPCP in CRC (0.7792). Asides from the pairwise iPCPs, our model also suggests high multi-way

iPCPs among PI3K inhibitors in BRCA (0.8002) and CRC (0.7513). Among these inhibitors, PI3K inhibitor of BYL719 was approved by FDA for breast cancer; MAPK inhibitor of binimetinib was approved by FDA for BRAF mutant melanoma in combination with encorafenib; and CDK inhibitor of LEE011 was approved for breast cancer.

Our model suggests treatments targeting different pathways also share high iPCP values across different cancers. Monotherapies targeting different cell cycle regulatory pathways (PI3K, MAPK and CDK) exhibit high iPCPs. CDK inhibitor, LEE011, and MAPK inhibitors share high pairwise iPCP values in BRCA ((LEE011, binimetinib): 0.7709), CRC ((LEE011, binimetinib): 0.8617, (LEE011, CKX620): 0.7820) and CM ((LEE011, binimetinib): 0.8210) in the Figure 6 with box (3). High iPCP among MAPK and CDK inhibitors agree with biology, since it is known that CDK and MAPK collaboratively regulate downstream pathways such as Ste5 (Repetto et al., 2018). High pairwise iPCP values between PI3K and MAPK inhibitors were observed in box (3) in the Figure 6. Specifically, our model suggests high pairwise iPCPs as follows: (i) BRCA, (binimetinib, BKM120): 0.7427, (binimetinib, BYL719): 0.7441, (binimetinib, CLR457): 0.7427)); (ii) CRC, (binimetinib, BKM120): 0.7374, (binimetinib, BYL719): 0.7388, (binimetinib, CLR457): 0.7541, (CKX620, BKM120): 0.7366, (CKX620, BYL719): 0.7357, (CKX620, CLR457): 0.7676)); (iii) CM, (binimetinib, BKM120): 0.8882. Aside from the pairwise iPCPs above, high multiway iPCPs in BRCA (0.7422), CRC (0.7300) and CM (0.8882) also show the similar information. From the existing literature, both PI3K and MAPK can be induced by ERBB3 phosphorylation (Balko et al., 2012) and it is not surprising to see high iPCPs between PI3K and MAPK inhibitors.

5.4. *Implications in Combination Therapy* Based on the concordance between the monotherapy and biology mechanism, we further investigate combination therapies to identify mechanisms with synergistic effect. In NIBR-PDXE, 21 combination therapies were tested and only one of them includes three monotherapies (BYL719 + cetuximab + encorafenib in CRC) and the rest contain two monotherapies. Out of 21 combination therapies, only three do not target any cell cycle (PI3K, MAPK, CDK, MDM2, JAK and BRAF) pathways (see Supplementary Material Table S8 for the full list of combination therapies). From the R_x-tree in Figure 5, combination therapies tend to form a tighter subtree and are closer to monotherapies targeting PI3K-MAPK-CDK, which implies that the mechanisms under combination therapies are similar to each other and are closer to the PI3K-MAPK-CDK pathways. We identified several combination therapies with known synergistic effects and provide a brief description for each of the cancers in the following paragraphs.

Breast cancer. Four combination therapies were tested in BRCA and three therapies targeting PI3K-MAPK-CDK (BYL719 + LJM716, BYL719 + LEE011 and LEE011 + everolimus) form a subtree in R_x-tree with a high three-way iPCP (0.8719). Among these combination therapies, PI3K-CDK inhibitor, BYL719 + LEE011, suggests a possible synergistic regulation (Vora et al., 2014; Bonelli et al., 2017; Yuan et al., 2019). Based on the high iPCP between BYL719 + LEE011 and the rest two therapies, we suggest synergistic effect for combination therapies targeting PI3K-ERBB3 (BYL719 + LJM716), and CDK-MTOR (LEE011 + everolimus) for future investigation.

Colorectal cancer. Our model suggests a high three-way iPCP (0.7437) among PI3K-EGFR (BYL719 + cetuximab), PI3K-EGFR-BRAF (BYL719 + cetuximab + encorafenib) and PI3K-ERBB3 (BYL719 + LJM716) inhibitors. Since the triple therapy (BYL719 + cetuximab + encorafenib) enters the phase I clinical trial with synergy (Geel et al., 2014), our model proposes the potential synergistic effect for PI3K-ERBB3 based on iPCP for future investigation. Of note, we found a modest iPCP (0.6280) between the FDA-approved combination therapy EGFR-BRAF (cetuximab + encorafenib) and PI3K-EGFR-BRAF (BYL719 +

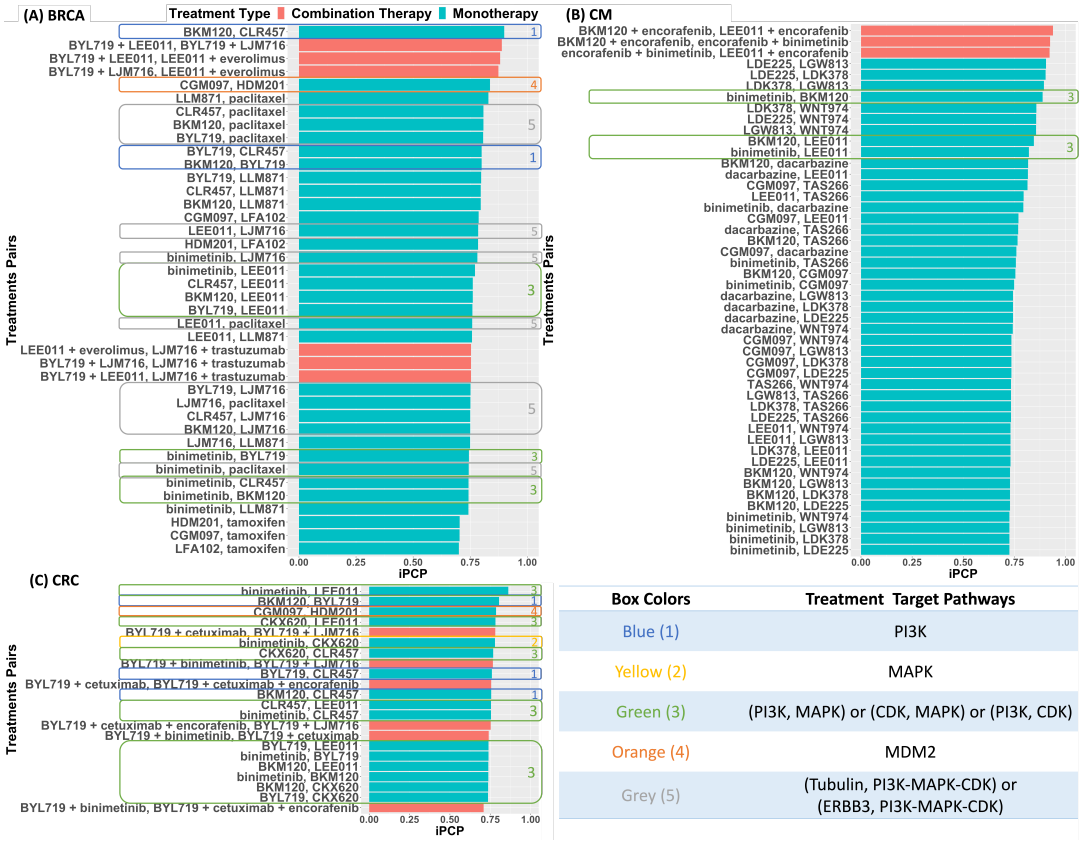

Figure 6: Bar plot of iPCPs for pairs of combination therapies (red bars) and pairs of monotherapies (green bars): (A) breast cancer, (B) colorectal cancer and (C) melanoma. The bar plots are sorted by the iPCP values (high to low); pairs of treatments are shown only if the estimated iPCP is greater than 0.7. Monotherapies have different known targets which are listed in the bottom-right table (see Section 5.3 for more details and discussion on monotherapies).

cetuximab + encorafenib) and the modest iPCP can be explained by an additional drug-drug interaction between BYL719 and encorafenib in triple-combined therapy (van Geel et al., 2017).

Melanoma. In NIBR-PDXE, three combination therapies were tested in CM, and all of them consist one monotherapy targeting PI3K-MAPK-CDK and the other one targeting BRAF. A tight subtree is observed in the R_x-tree and our model also suggests a high iPCP (0.9222) among three combination therapies. Since PI3K, MAPK and CDK work closely and share a high iPCP (0.8204) among monotherapies in CM, a high iPCP (0.9222) among three combination therapies is not surprising. Since two combination therapies of BRAF-MAPK (dabrafenib + trametinib and encorafenib + binimetinib) are approved by FDA for BRAF-mutant metastatic melanoma (Dummer et al., 2018a,b; Robert et al., 2019), we recommend the synergy for BRAF-PI3K (encorafenib + BKM120) and BRAF-CDK (encorafenib + LEE011) inhibitors.

Comparison to alternative approaches. Unlike the probabilistic generative modeling approach proposed in this paper, standard distance-based agglomerative hierarchical clustering and Pearson correlation can also be applied to the PDX data to estimate the similarity. However, simple pairwise similarities can be potentially noisy and the uncertainty in the estimation is not fully incorporated due to the absence of a generative model. As we showed

in the Section 4.1 (Simulation I) that agglomerative hierarchical clustering and the Pearson correlation leads to inferior recovery of the true branching times and the true tree structure under different data generating mechanisms mimicking the real data. As further evidence, we compute pairwise similarities based on Pearson correlation (other distance metrics show similar patterns) in the right panel of Figure 5. By mapping the original Pearson correlation $\rho \in [-1, 1]$ through a linear function $\frac{\rho+1}{2}$, we make the range of iPCP and Pearson correlation comparable. We observe that pairwise iPCP estimated through the DDT model is less noisy than Pearson correlation. For example, both iPCP and Pearson correlation in CM show higher similarities among combination therapy framed by a box, but iPCP exhibits a clearer pattern than Pearson correlation.

6. Summary and Discussion In translational oncology research, PDX studies have emerged as a unique study design that evaluates multiple treatments when applied to samples from the same human tumor implanted into genetically identical mice. PDX systems are promising tools for large-scale screening to evaluate a large number of FDA-approved and novel cancer therapies. However, there remain scientific questions concerning how distinct treatments may be synergistic in inducing similar efficacious responses, and how to identify promising subsets of treatments for further clinical evaluation. To this end, in this paper, we propose a probabilistic framework to learn treatment trees (R_x-trees) from PDX data to identify promising treatment combinations and plausible biological mechanisms that confer synergistic effect(s). In particular, in a Bayesian framework based on the Dirichlet Diffusion Tree, we estimate a *maximum a posteriori* rooted binary tree with the treatments on the leaves and propose a posterior uncertainty-aware similarity measure (iPCP) for any subset of treatments. The divergence times of the DDT encode the tree topology and are profitably interpreted within the context of an underlying plausible biological mechanism of treatment actions.

From the class of probabilistic models with an unknown tree structure component, we have chosen the DDT mainly owing to the availability of a closed-form marginal likelihood that directly links the tree topological structure to the covariance structure of the observed PDX data, which additionally decouples the Euclidean and tree parameters; to the best of our knowledge this method has not been proposed or explored hitherto for the DDT. The decoupling leads to efficient posterior inference via a two-stage algorithm that confers several advantages. The algorithm generates posterior samples of Euclidean parameters through approximate Bayesian computation and passes the posterior medians to a second stage classical Metropolis-Hastings algorithm for sampling from the conditional posterior distribution of the tree given all other quantities. Through simulation studies, we show that the proposed two-stage algorithm generates better posterior tree samples and captures the true similarity among treatments better than alternatives such as single-stage MCMC and naive Pearson correlations. The posterior samples of trees are summarized by iPCP, which we propose to measure the empirical mechanistic similarity for multiple treatments incorporating uncertainty.

Using the proposed methodology on NIBR-PDXE data, we estimate R_x-trees and iPCPs for five cancers. Among the monotherapies, iPCP is highly concordant with known biology across different cancers. For example, BKM120 and BYL719 show a high iPCP value among treatments in breast and colorectal cancer, which corroborates known mechanisms, since both monotherapies target the same biological pathway, PI3K, and BYL719 was approved by FDA for breast cancer. The proposed iPCP can also suggest improvements upon an existing combination therapy. We first identify a combination therapy with known synergy (not based on the our data) and then determine which additional therapies (monotherapies or combination therapies) have high iPCPs when considered together with the existing combination therapy. Based on the NIBR-PDXE data, for each cancer, we suggest potential synergies between PI3K-ERBB3 and CDK-MTOR for breast cancer, PI3K-ERBB3 for colorectal cancer,

and BRAF-PI3K and BRAF-CDK for melanoma that could be potentially explored in future translational studies.

Our current analysis infers treatment trees based on the drug responses from the NIBR-PDXE dataset which provides treatment similarity information that may be used to guide potential treatment strategies. However, there are a few limitations. First, the PDX experiments may fail to capture the difference in the microenvironment between the human and the immunodeficient mouse (Dobrolecki et al., 2016), which must be considered in disease contexts when findings are generalized to human. As PDX technology matures, this can be compensated by better PDX experiments that capture the tumor microenvironment more precisely. For example, one can use the genetically engineered mice to reconstruct the human immune system (Abdolahi et al., 2022), and some studies have started to adapt this method in the context of immunotherapies (Zhao et al., 2018). Second, on experimental design, current literature points to the potential advantage of designs with multiple animals per treatment and patient (Abdolahi et al., 2022). We can incorporate the random effects in the current model of (4) for the multiple-animal-per-patient design and we refer the reader to the Supplementary Material Section S7 for more details. Also, to evaluate PDX designs with fewer treatments and patients that is common in co-clinical trials (e.g., Koga and Ochiai, 2019), we conducted a simulation for two datasets with a smaller dimension ($(I, J) = (5, 5)$ and $(10, 15)$) which confirmed the advantage of the proposed method in terms of recovering treatment similarities (see Supplementary Material Section S5.5). Finally, from a statistical perspective, we have assumed independent patients without using the underlying patient-specific genomic information that is also available in the NIBR-PDXE. By including patient-specific genomic information, we may further improve our ability to identify synergistic treatments that may be specific to a subset of patients. One approach to utilizing genomic information could be to extend the DDT model to incorporate patient-specific genomic information in the mean structure or the column covariance of the marginal likelihood of Equation (4). In addition, models with non-Gaussian marginal likelihood and non-binary treatment tree in principle can be defined by considering generative tree models based on general diffusion processes (Heaukulani, Knowles and Ghahramani, 2014; Knowles and Ghahramani, 2015). Both extensions raise significant, non-trivial methodological and computation issues (e.g., deriving tractable likelihoods; finding low-dimensional summary statistics for new parameters) and constitute the foundation for future work.

Code and data availability We also provide a general purpose code in `R` that accompanies this manuscript along with all the necessary documentation and datasets required to replicate our results (see `https://github.com/bayesrx/RxTree`). Furthermore, to aid access and visualization of the results, we have also developed an `R`-shiny application (see Supplementary Material Section S6.6).

REFERENCES

ABDOLAHI, S., GHAZVINIAN, Z., MUHAMMADNEJAD, S., SALEH, M., ASADZADEH AGHDAEI, H. and BAGHAEI, K. (2022). Patient-derived xenograft (PDX) models, applications and challenges in cancer research. *J Transl Med* **20** 206.

BALKO, J. M., MILLER, T. W., MORRISON, M. M., HUTCHINSON, K., YOUNG, C., RINEHART, C., SÁNCHEZ, V., JEE, D., POLYAK, K., PRAT, A., PEROU, C. M., ARTEAGA, C. L. and COOK, R. S. (2012). The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium. *Proc Natl Acad Sci U S A* **109** 221–226.

BAYAT MOKHTARI, R., HOMAYOUNI, T. S., BALUCH, N., MORGATSKAYA, E., KUMAR, S., DAS, B. and YEGER, H. (2017). Combination therapy in combating cancer. *Oncotarget* **8** 38022–38043.

BERTOTTI, A., MIGLIARDI, G., GALIMI, F., SASSI, F., TORTI, D., ISELLA, C., CORÀ, D., DI NICOLANTONIO, F., BUSCARINO, M., PETTI, C., RIBERO, D., RUSSOLILLO, N., MURATORE, A., MASSUCCO, P., PISACANE, A., MOLINARO, L., VALTORTA, E., SARTORE-BIANCHI, A., RISIO, M., CAPUSSOTTI, L.,

GAMBACORTA, M., SIENA, S., MEDICO, E., SAPINO, A., MARSONI, S., COMOGLIO, P. M., BARDELLI, A. and TRUSOLINO, L. (2011). A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov* **1** 508–523.

BHIMANI, J., BALL, K. and STEBBING, J. (2020). Patient-derived xenograft models-the future of personalised cancer treatment. *Br J Cancer* **122** 601–602.

BONELLI, M. A., DIGIACOMO, G., FUMAROLA, C., ALFIERI, R., QUAINI, F., FALCO, A., MADEDDU, D., LA MONICA, S., CRETELLA, D., RAVELLI, A., ULIVI, P., TEBALDI, M., CALISTRI, D., DELMONTE, A., AMPOLLINI, L., CARBOGNANI, P., TISEO, M., CAVAZZONI, A. and PETRONINI, P. G. (2017). Combined inhibition of CDK4/6 and PI3K/AKT/mTOR pathways induces a synergistic anti-tumor effect in malignant pleural mesothelioma cells. *Neoplasia* **19** 637–648.

BRAVO, H. C., WRIGHT, S., ENG, K. H., KELES, S. and WAHBA, G. (2009). Estimating tree-structured covariance matrices via mixed-integer programming. *J Mach Learn Res* **5** 41–48.

CARDONA, G., MIR, A., ROSSELLÓ, F., ROTGER, L. and SÁNCHEZ, D. (2013). Cophenetic metrics for phylogenetic trees, after Sokal and Rohlf. *BMC Bioinformatics* **14** 3.

CASELLA, G. and BERGER, R. (2001). *Statistical Inference*. Duxbury Resource Center.

CLOHESSY, J. G. and PANDOLFI, P. P. (2015). Mouse hospital and co-clinical trial project–from bench to bedside. *Nat Rev Clin Oncol* **12** 491–498.

DAGOGO-JACK, I. and SHAW, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15** 81–94.

DOBROLECKI, L. E., AIRHART, S. D., ALFEREZ, D. G., APARICIO, S., BEHBOD, F., BENTIRES-ALJ, M., BRISKEN, C., BULT, C. J., CAI, S., CLARKE, R. B., DOWST, H., ELLIS, M. J., GONZALEZ-SUAREZ, E., IGGO, R. D., KABOS, P., LI, S., LINDEMAN, G. J., MARANGONI, E., MCCOY, A., MERIC-BERNSTAM, F., PIWNICA-WORMS, H., POUPON, M. F., REIS-FILHO, J., SARTORIUS, C. A., SCABIA, V., SFLOMOS, G., TU, Y., VAILLANT, F., VISVADER, J. E., WELM, A., WICHA, M. S. and LEWIS, M. T. (2016). Patient-derived xenograft (PDX) models in basic and translational breast cancer research. *Cancer Metastasis Rev* **35** 547–573.

DUMMER, R., ASCIERTO, P. A., GOGAS, H. J., ARANCE, A., MANDALA, M., LISZKAY, G., GARBE, C., SCHADENDORF, D., KRAJSOVA, I., GUTZMER, R., CHIARION-SILENI, V., DUTRIAUX, C., DE GROOT, J. W. B., YAMAZAKI, N., LOQUAI, C., MOUTOUH-DE PARSEVAL, L. A., PICKARD, M. D., SANDOR, V., ROBERT, C. and FLAHERTY, K. T. (2018a). Encorafenib plus binimetinib versus vemurafenib or encorafenib in patients with BRAF-mutant melanoma (COLUMBUS): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* **19** 603–615.

DUMMER, R., ASCIERTO, P. A., GOGAS, H. J., ARANCE, A., MANDALA, M., LISZKAY, G., GARBE, C., SCHADENDORF, D., KRAJSOVA, I., GUTZMER, R., CHIARION SILENI, V., DUTRIAUX, C., DE GROOT, J. W. B., YAMAZAKI, N., LOQUAI, C., MOUTOUH-DE PARSEVAL, L. A., PICKARD, M. D., SANDOR, V., ROBERT, C. and FLAHERTY, K. T. (2018b). Overall survival in patients with BRAF-mutant melanoma receiving encorafenib plus binimetinib versus vemurafenib or encorafenib (COLUMBUS): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol* **19** 1315–1327.

FERLAY, J., ERVIK, M., LAM, F., COLOMBET, M., MERY, L., PIÑEROS, M., ZNAOR, A., SOERJOMATARAM, I. and BRAY, F. (2020). *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer Available from: `https://gco.iarc.fr/today`, accessed 05.28.2021.

GAO, H., KORN, J. M., FERRETTI, S., MONAHAN, J. E., WANG, Y., SINGH, M., ZHANG, C., SCHNELL, C., YANG, G., ZHANG, Y., BALBIN, O. A., BARBE, S., CAI, H., CASEY, F., CHATTERJEE, S., CHIANG, D. Y., CHUAI, S., COGAN, S. M., COLLINS, S. D., DAMMASSA, E., EBEL, N., EMBRY, M., GREEN, J., KAUFFMANN, A., KOWAL, C., LEARY, R. J., LEHAR, J., LIANG, Y., LOO, A., LORENZANA, E., ROBERT MCDONALD, E., MCLAUGHLIN, M. E., MERKIN, J., MEYER, R., NAYLOR, T. L., PATAWARAN, M., REDDY, A., RÖELLI, C., RUDDY, D. A., SALANGSANG, F., SANTACROCE, F., SINGH, A. P., TANG, Y., TINETTO, W., TOBLER, S., VELAZQUEZ, R., VENKATESAN, K., VON ARX, F., WANG, H. Q., WANG, Z., WIESMANN, M., WYSS, D., XU, F., BITTER, H., ATADJA, P., LEES, E., HOFMANN, F., LI, E., KEEN, N., COZENS, R., JENSEN, M. R., PRYER, N. K., WILLIAMS, J. A. and SELLERS, W. R. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21** 1318–1325.

GEEL, R. V., ELEZ, E., BENDELL, J. C., FARIS, J. E., LOLKEMA, M. P. J. K., ESKENS, F., SPREAFICO, A., KAVAN, P., DELORD, J.-P., SCHULER, M. H., WAINBERG, Z. A., YAMADA, Y., YOSHINO, T., DEMUTH, T., AVSAR, E., CHATTERJEE, A., ZHU, P., BERNARDS, R., TABERNERO, J. and SCHELLENS, J. H. (2014). Phase I study of the selective BRAFV600 inhibitor encorafenib (LGX818) combined with cetuximab and with or without the α-specific PI3K inhibitor BYL719 in patients with advanced BRAF-mutant colorectal cancer. *Journal of Clinical Oncology* **32** 3514-3514.

GRANT, R. L., COMBS, A. B. and ACOSTA, D. (2010). Experimental models for the investigation of toxicological mechanisms. In *Comprehensive Toxicology (Second Edition)* second edition ed. (C. A. McQueen, ed.) 203-224. Elsevier, Oxford.

GROISBERG, R. and SUBBIAH, V. (2021). Combination therapies for precision oncology: the ultimate whack-a-mole game. *Clin Cancer Res* **27** 2672–2674.

HEAUKULANI, C., KNOWLES, D. A. and GHAHRAMANI, Z. (2014). Beta diffusion trees. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. *ICML'14* II–1809–II–1817. JMLR.org.

HIDALGO, M., AMANT, F., BIANKIN, A. V., BUDINSKÁ, E., BYRNE, A. T., CALDAS, C., CLARKE, R. B., DE JONG, S., JONKERS, J., MÆLANDSMO, G. M., ROMAN-ROMAN, S., SEOANE, J., TRUSOLINO, L. and VILLANUEVA, A. (2014). Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov* **4** 998–1013.

KNOWLES, D. A., GAEL, J. V. and GHAHRAMANI, Z. (2011). Message passing algorithms for Dirichlet diffusion trees. *International Conference on Machine Learning (ICML)*.

KNOWLES, D. A. and GHAHRAMANI, Z. (2015). Pitman-Yor diffusion trees for Bayesian hierarchical clustering. *IEEE Trans Pattern Anal Mach Intell* **37** 271–289.

KNUTH, D. E. (1976). Big omicron and big omega and big theta. *ACM Sigact News* **8** 18–24.

KOGA, Y. and OCHIAI, A. (2019). Systematic Review of Patient-Derived Xenograft Models for Preclinical Studies of Anti-Cancer Drugs in Solid Tumors. *Cells* **8**.

KONOPLEVA, M., MARTINELLI, G., DAVER, N., PAPAYANNIDIS, C., WEI, A., HIGGINS, B., OTT, M., MASCARENHAS, J. and ANDREEFF, M. (2020). MDM2 inhibition: an important step forward in cancer therapy. *Leukemia* **34** 2858–2874.

KRUMBACH, R., SCHÜLER, J., HOFMANN, M., GIESEMANN, T., FIEBIG, H. H. and BECKERS, T. (2011). Primary resistance to cetuximab in a panel of patient-derived tumour xenograft models: activation of MET as one mechanism for drug resistance. *Eur J Cancer* **47** 1231–1243.

KURTZEBORN, K., KWON, H. N. and KUURE, S. (2019). MAPK/ERK Signaling in regulation of renal differentiation. *Int J Mol Sci* **20**.

LAI, Y., WEI, X., LIN, S., QIN, L., CHENG, L. and LI, P. (2017). Current status and perspectives of patient-derived xenograft models in cancer research. *J Hematol Oncol* **10** 106.

LAPOINTE, F.-J. and LEGENDRE, P. (1991). The generation of random ultrametric matrices representing dendrograms. *Journal of Classification* **8** 177-200.

LIU, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* **89** 958–966.

MATHAI, A. M. (1980). Moments of the trace of a noncentral Wishart matrix. *Communications in Statistics - Theory and Methods* **9** 795-801.

MCCULLAGH, P. (2006). Structured covariance matrices in multivariate regression models Technical Report, Department of Statistics, University of Chicago.

MEZARD, M. and MONTANARI, A. (2009). *Information, Physics, and Computation*. Oxford University Press, Inc., USA.

NARAYAN, R. S., MOLENAAR, P., TENG, J., CORNELISSEN, F. M. G., ROELOFS, I., MENEZES, R., DIK, R., LAGERWEIJ, T., BROERSMA, Y., PETERSEN, N., MARIN SOTO, J. A., BRANDS, E., VAN KUIKEN, P., LECCA, M. C., LENOS, K. J., IN 'T VELD, S. G. J. G., VAN WIERINGEN, W., LANG, F. F., SULMAN, E., VERHAAK, R., BAUMERT, B. G., STALPERS, L. J. A., VERMEULEN, L., WATTS, C., BAILEY, D., SLOTMAN, B. J., VERSTEEG, R., NOSKE, D., SMINIA, P., TANNOUS, B. A., WURDINGER, T., KOSTER, J. and WESTERMAN, B. A. (2020). A cancer drug atlas enables synergistic targeting of independent drug vulnerabilities. *Nat Commun* **11** 2935.

NEAL, R. (2003). Density modeling and clustering using Dirichlet diffusion trees. *Bayesian Statistics* **7** 619-629.

NUNES, M., VRIGNAUD, P., VACHER, S., RICHON, S., LIÈVRE, A., CACHEUX, W., WEISWALD, L.-B., MASSONNET, G., CHATEAU-JOUBERT, S., NICOLAS, A., DIB, C., ZHANG, W., WATTERS, J., BERGSTROM, D., ROMAN-ROMAN, S., BIÈCHE, I. and DANGLES-MARIE, V. (2015). Evaluating Patient-Derived Colorectal Cancer Xenografts as Preclinical Models by Comparison with Patient Clinical Data. *Cancer Research* **75** 1560–1566.

OH, D. Y. and BANG, Y. J. (2020). HER2-targeted therapies - a role beyond breast cancer. *Nat Rev Clin Oncol* **17** 33–48.

RASHID, N. U., LUCKETT, D. J., CHEN, J., LAWSON, M. T., WANG, L., ZHANG, Y., LABER, E. B., LIU, Y., YEH, J. J., ZENG, D. and KOSOROK, M. R. (2020). High-Dimensional Precision Medicine From Patient-Derived Xenografts. *Journal of the American Statistical Association* **0** 1-15.

REPETTO, M. V., WINTERS, M. J., BUSH, A., REITER, W., HOLLENSTEIN, D. M., AMMERER, G., PRYCIAK, P. M. and COLMAN-LERNER, A. (2018). CDK and MAPK synergistically regulate signaling dynamics via a shared multi-site phosphorylation region on the scaffold protein Ste5. *Mol Cell* **69** 938–952.

ROBERT, C., GROB, J. J., STROYAKOVSKIY, D., KARASZEWSKA, B., HAUSCHILD, A., LEVCHENKO, E., CHIARION SILENI, V., SCHACHTER, J., GARBE, C., BONDARENKO, I., GOGAS, H., MANDALÁ, M., HAANEN, J. B. A. G., LEBBÉ, C., MACKIEWICZ, A., RUTKOWSKI, P., NATHAN, P. D., RIBAS, A., DAVIES, M. A., FLAHERTY, K. T., BURGESS, P., TAN, M., GASAL, E., VOI, M., SCHADENDORF, D. and LONG, G. V. (2019). Five-year outcomes with dabrafenib plus trametinib in metastatic melanoma. *N Engl J Med* **381** 626–636.

SAWYERS, C. L. (2013). Perspective: combined forces. *Nature* **498** S7.

SISSON, S. A., FAN, Y. and BEAUMONT, M. (2018). *Handbook of approximate Bayesian computation*. CRC Press.

SOKAL, R. R. and ROHLF, F. J. (1962). The Comparison of Dendrograms by Objective Methods. *Taxon* **11** 33–40.

SUN, W., SANDERSON, P. E. and ZHENG, W. (2016). Drug combination therapy increases successful drug repositioning. *Drug Discov Today* **21** 1189–1195.

TENTLER, J. J., TAN, A. C., WEEKES, C. D., JIMENO, A., LEONG, S., PITTS, T. M., ARCAROLI, J. J., MESSERSMITH, W. A. and ECKHARDT, S. G. (2012). Patient-derived tumour xenografts as models for oncology drug development. *Nat Rev Clin Oncol* **9** 338–350.

TOPP, M. D., HARTLEY, L., COOK, M., HEONG, V., BOEHM, E., MCSHANE, L., PYMAN, J., MCNALLY, O., ANANDA, S., HARRELL, M., ETEMADMOGHADAM, D., GALLETTA, L., ALSOP, K., MITCHELL, G., FOX, S. B., KERR, J. B., HUTT, K. J., KAUFMANN, S. H., AUSTRALIAN OVARIAN CANCER STUDY, SWISHER, E. M., BOWTELL, D. D., WAKEFIELD, M. J. and SCOTT, C. L. (2014). Molecular correlates of platinum response in human high-grade serous ovarian cancer patient-derived xenografts. *Molecular Oncology* **8** 656-668.

TURNER, B. M., SEDERBERG, P. B., BROWN, S. D. and STEYVERS, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychol Methods* **18** 368–384.

VAN GEEL, R. M. J. M., TABERNERO, J., ELEZ, E., BENDELL, J. C., SPREAFICO, A., SCHULER, M., YOSHINO, T., DELORD, J. P., YAMADA, Y., LOLKEMA, M. P., FARIS, J. E., ESKENS, F. A. L. M., SHARMA, S., YAEGER, R., LENZ, H. J., WAINBERG, Z. A., AVSAR, E., CHATTERJEE, A., JAEGER, S., TAN, E., MAHARRY, K., DEMUTH, T. and SCHELLENS, J. H. M. (2017). A phase Ib dose-escalation study of encorafenib and cetuximab with or without alpelisib in metastatic BRAF-mutant colorectal cancer. *Cancer Discov* **7** 610–619.

VORA, S. R., JURIC, D., KIM, N., MINO-KENUDSON, M., HUYNH, T., COSTA, C., LOCKERMAN, E. L., POLLACK, S. F., LIU, M., LI, X., LEHAR, J., WIESMANN, M., WARTMANN, M., CHEN, Y., CAO, Z. A., PINZON-ORTIZ, M., KIM, S., SCHLEGEL, R., HUANG, A. and ENGELMAN, J. A. (2014). CDK 4/6 inhibitors sensitize PIK3CA mutant breast cancer to PI3K inhibitors. *Cancer Cell* **26** 136–149.

YOSHIDA, G. J. (2020). Applications of patient-derived tumor xenograft models and tumor organoids. *Journal of Hematology & Oncology* **13** 4.

YUAN, Y., WEN, W., YOST, S. E., XING, Q., YAN, J., HAN, E. S., MORTIMER, J. and YIM, J. H. (2019). Combination therapy with BYL719 and LEE011 is synergistic and causes a greater suppression of p-S6 in triple negative breast cancer. *Sci Rep* **9** 7509.

ZHANG, X., CLAERHOUT, S., PRAT, A., DOBROLECKI, L. E., PETROVIC, I., LAI, Q., LANDIS, M. D., WIECHMANN, L., SCHIFF, R., GIULIANO, M., WONG, H., FUQUA, S. W., CONTRERAS, A., GUTIERREZ, C., HUANG, J., MAO, S., PAVLICK, A. C., FROEHLICH, A. M., WU, M.-F., TSIMELZON, A., HILSENBECK, S. G., CHEN, E. S., ZULOAGA, P., SHAW, C. A., RIMAWI, M. F., PEROU, C. M., MILLS, G. B., CHANG, J. C. and LEWIS, M. T. (2013). A Renewable Tissue Resource of Phenotypically Stable, Biologically and Ethnically Diverse, Patient-Derived Human Breast Cancer Xenograft Models. *Cancer Research* **73** 4885-4897.

ZHAO, Y., SHUEN, T. W. H., TOH, T. B., CHAN, X. Y., LIU, M., TAN, S. Y., FAN, Y., YANG, H., LYER, S. G., BONNEY, G. K., LOH, E., CHANG, K. T. E., TAN, T. C., ZHAI, W., CHAN, J. K. Y., CHOW, E. K., CHEE, C. E., LEE, G. H., DAN, Y. Y., CHOW, P. K., TOH, H. C., LIM, S. G. and CHEN, Q. (2018). Development of a new patient-derived xenograft humanised mouse model to study human-specific tumour microenvironment and immunotherapy. *Gut* **67** 1845–1854.