# A structural causal framework for interventions on evolutionary accumulation models

Ramon Diaz-Uriarte [1,2,*], Íñigo Ríos Arroyo[1,2], and Iain G. Johnston [3,4]

[1] Instituto de Investigaciones Biomédicas Sols-Morreale (IIBM), UAM-CSIC, Madrid, Spain
[2]Department of Biochemistry, Universidad Autónoma de Madrid, Madrid, Spain
[3]Department of Mathematics, University of Bergen, Bergen, Norway
[4]Computational Biology Unit, University of Bergen, Bergen, Norway
[*]Author for correspondence: r.diaz@uam.es

2026-10-03

**Abstract**

Evolutionary accumulation models (EvAMs), also known as cancer progression models (CPMs), infer dependencies in the order of accumulation of mutations during tumor progression from cross-sectional data. It has been suggested that EvAMs could be used to identify therapeutic targets, but there is no procedure in the literature for how to extract predictions under intervention from these models. A simple approach of conditioning on the absence of a mutation gives incorrect predictions. We address this gap by formalizing what "intervene" means for all currently available EvAM methods, using Pearl's *do* operator and conditional interventions. For each model, we show how to implement the intervention, identify equivalent procedures, and analyze whether the modularity assumption is justified. Drawing on individual-level causal DAGs that make fitness an explicit variable, we distinguish two types of intervention (killing and inactivating) that are conflated in standard EvAM representations. Since the goal is to prioritize intervention candidates, we recast the problem as one of ranking: we define three intervention objectives and provide a protocol for evaluating how well EvAMs rank targets. Our framework is not specific to cancer or EvAMs; it applies wherever fitted computational models can be interpreted as structural causal models.

## 1. Introduction

Cancer progression is driven largely by the sequential accumulation of somatic mutations during the life of an individual: cancer cells can accumulate driver mutations that increase cell-level fitness, increasing in frequency and spreading (and ultimately resulting in metastasis), at the expense of the whole organism (Hanahan and Weinberg, 2011).

Because of epistatic interactions, the accumulation of driver cancer mutations is often constrained to follow only some possible sequential orders, as the effect or viability of a mutation can depend on the presence of other mutations. Evolutionary accumulation models (EvAMs), also known as cancer progression models (CPMs), have been developed to infer these dependencies in the order of accumulation of mutations during tumor progression from cross-sectional data (Hainke *et al.*, 2012; Beerenwinkel *et al.*, 2015; Diaz-Uriarte and Johnston, 2025). From these dependencies, it is possible to derive the trajectories of tumor progression from an initial set of genotypes (e.g., wildtype) to the genotype with all driver mutations (Diaz-Uriarte and Vasallo, 2019; Hosseini *et al.*, 2019; Aga *et al.*, 2024).

EvAMs use cross-sectional data (data obtained from different samples which are taken at variable unknown time points —Ramazzotti *et al.*, 2015; Beerenwinkel *et al.*, 2015; Diaz-Uriarte and Johnston, 2025), a type of data which are very abundant, freely available for many cancer types, and easier to obtain than multiple samples from the same individual. These methods have also been used in very different domains (such as HIV mutation accumulation —Beerenwinkel *et al.*, 2006, 2007; Montazeri *et al.*, 2015; Posada-Céspedes *et al.*, 2021—, and malaria progression in children —Johnston *et al.*, 2019). This indicates their potential usefulness, and explains why new methods continue to be developed. Furthermore, EvAMs and related models could have a translational impact. Denoising data by inferring the most probable hidden genotype improves survival analysis of cancer data (Gerstung *et al.*, 2009), and it has been shown that these methods can be used to stratify cancer patients (Angaroni *et al.*, 2021; Fontana *et al.*, 2023) and could help triage malaria sufferers (Johnston *et al.*, 2019).

It has also been suggested (Desper *et al.*, 1999; Attolini *et al.*, 2010; Gerstung *et al.*, 2011) that EvAMs could be used to determine which genes (or gene products or molecular pathways) should be targeted or intervened upon (the types of interventions considered are explained in

section 2.2). The simplest procedure would be to use these methods to identify initiating events, which are often more promising targets than late mutation acquisitions (Ozawa *et al.*, 2014; Cheng *et al.*, 2012; Cristea *et al.*, 2017). But the inferences of EvAM methods are much richer than an ordering of early vs. late mutations, and it is not just order of accumulation, but particular consequences on downstream events that arguably matter most. For example, in so far as EvAMs can be used to predict evolutionary trajectories, we could use them to design interventions to drive the evolutionary process towards (or away from) chosen states (Hosseini *et al.*, 2019). Therefore, it should be possible to use EvAMs, and leverage their direct and indirect inferences about the evolutionary process, to identify therapeutic targets for specific objectives.

From that perspective, instead of focusing on the quality of the inferences about the patterns of dependencies, what we want to know is what could work from the point of view of interventions. Thus, we shift the focus from "correct" to "actionable" inferences. It is conceivable that some inferences about the true dependencies can be wrong whereas ranking of genes for the purpose of interventions is successful and useful. An EvAM will be useful to identify targets if it can provide a ranking of genes that is very similar (or highly correlated) to the true ranking of genes derived from the ground truth. But before this can be evaluated, we need a principled way to modify EvAMs to obtain predictions under intervention, a way to answer "what if" questions like "What does this fitted model say would be the evolutionary dynamics if we were to intervene on gene $A$?". As we show in section 2.3, treating intervention as conditioning on the absence of a mutation gives incorrect predictions.

Our focus is not the identification or estimation of causal effects from observational data —the central concern of much of the causal inference literature (Hernán and Robins, 2020; Morgan and Winship, 2015; Pearl, 2009). Rather, we address a logically prior question: given a fitted model, what is the correct procedure for extracting predictions under intervention? For models designed as regressions, this is straightforward: one substitutes the desired value into the fitted equation. But EvAMs were not designed for intervention, and the mapping from "target gene $A$" to a specific model manipulation is neither obvious nor uniform across model families.

We develop a rationale for modifying EvAMs that results from the application of Pearl's (2009) *do* operator and conditional interventions (Correa and Bareinboim, 2020; Eberhardt and Scheines, 2007; Korb *et al.*, 2004) (both are formalisms for representing what would happen under an intervention, as distinct from observational conditioning). To draw the distinction between two types of intervention, killing vs. gene inactivation, that are conflated in standard EvAM representations, we build on the individual-level causal DAGs of Frank (1997) and Otsuka (2016, 2019), which make fitness an explicit variable in the causal structure. We consider virtually all of the currently available EvAM methods, and for every intervention in each method, we include a detailed analysis of the role of modularity (Woodward, 2003; Pearl, 2009). We also discuss several ways of quantifying intervention effects and provide a protocol for evaluating EvAMs' ability to identify targets. Our notion of causal effects follows Hernán and Robins's (2020, p. v) "causal effects in populations, that is, numerical quantities that measure changes in the distribution of an outcome under different interventions." Since our goal is to prioritize intervention candidates rather than estimate individual effect sizes, we recast these quantities as a ranking problem (section 4.3).

The main contributions of our ms. are: (a) the conceptualization of what "intervene" means across a class of fitted models, using an interventionist causal framework, which provides a principled basis for manipulating the output of EvAMs to obtain predictions under intervention; (b) the differentiation between different types of intervention, which also sets the stage

for possibly more complex schemes (like stochastic interventions); (c) the criteria to measure gene importance from EvAMs.

## 2. Overview

To use EvAMs for interventions we need three things: (a) a clear definition of what an intervention is; (b) a principled way to modify fitted models to predict what happens under the intervention; (c) a way to quantify the effects of the intervention relative to an objective (the outcome whose distribution changes under intervention). We address (a) below in subsection 2.2, and (b), the key section of this ms., in section 3; (c) is discussed in section 4. In the rest of this section we list the EvAMs considered (section 2.1), and provide a simple example of why a naive approach to intervention (conditioning) would fail with EvAMs (section 2.3).

### 2.1. EvAMs considered

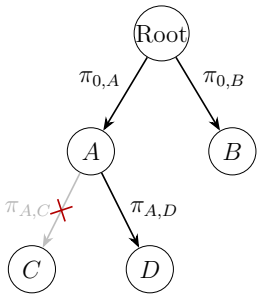
We consider here all EvAMs currently in use and with existing software implementations. These include Oncogenetic Trees (OT —Desper *et al.*, 1999; Szabo and Boucher, 2008), OncoBN (Nicol *et al.*, 2021), Conjunctive Bayesian Networks (CBN —Gerstung *et al.*, 2009; Montazeri *et al.*, 2016), Hidden Extended Suppes-Bayes Causal Networks (H-ESBCN —Angaroni *et al.*, 2021), Mutual Hazard Networks (MHN —Schill *et al.*, 2020), HyperHMM (Moen and Johnston, 2023), and HyperTraPS(-CT) (Johnston and Williams, 2016; Greenbury *et al.*, 2020; Aga *et al.*, 2024). For overviews see Diaz-Uriarte and Johnston (2025), Diaz-Uriarte and Herrera-Nieto (2022), Beerenwinkel *et al.* (2015).

### 2.2. What do we mean by "intervene on a gene"?

We consider two different meanings of "targeting a gene": "killing" and "inactivating" interventions. In a killing intervention targeting a gene makes all genotypes that bear a mutation in that gene (i.e., a particular allele) lethal; equivalently, a killing intervention turns a mutation in that gene into a lethal mutation. This is coherent with widespread uses of this term and some of the existing procedures in the field (Luo *et al.*, 2009; O'Neil *et al.*, 2017). A different type of intervention is the gene inactivating intervention (D'Antonio *et al.*, 2023; Zhan *et al.*, 2019) where, even if the gene is mutated, it has no downstream consequences, either because it is engineered back to the unmutated state or because its modified downstream products (e.g., mRNA, protein) are blocked; this is the type of intervention mentioned in Schill *et al.* (2024).

### 2.3. Intervention vs. conditioning

Pearl's (2009) *do* operator, written $do(X = x)$, denotes intervention —setting $X$ to $x$ by external action— as distinct from conditioning on the observation $X = x$. The two operations give different answers when $X$ shares causes with, or is caused by, other variables in the model. For example, in a "fork" structure, where $Z$ is a common cause of $X$ and $Y$ ($Z \to X$, $Z \to Y$), observing $X = x$ is informative about $Z$ and thus about $Y$, whereas intervening on $X$ to set $X = x$ makes $X$ independent of $Z$, and therefore independent of $Y$, so that $P(Y|X = x)$ is generally different from $P(Y|do(X = x))$. A separate issue arises with common effects (colliders): conditioning on a common effect of two variables induces a dependence between them, biasing inference further down. Section 3.1 defines interventions formally.




| Genotype | Predicted frequency | No_Int. | Int. | Cond. |
|---|---|---|---|---|
| WT | $(1-\pi_{0,A})\,(1-\pi_{0,B})$ | 0.350 | 0.350 | 0.538 |
| $A$ | $\pi_{0,A}\,(1-\pi_{0,B})\,(1-\pi_{A,D})\,\underline{(1-\pi_{A,C})}$ | 0.084 | 0.280 | 0.129 |
| $B$ | $\pi_{0,B}\,(1-\pi_{0,A})$ | 0.150 | 0.150 | 0.231 |
| $AB$ | $\pi_{0,A}\,\pi_{0,B}\,(1-\pi_{A,D})\,\underline{(1-\pi_{A,C})}$ | 0.036 | 0.120 | 0.055 |
| $AC$ | $\pi_{0,A}\,(1-\pi_{0,B})\,(1-\pi_{A,D})\,\underline{\pi_{A,C}}$ | 0.196 | 0 | 0 |
| $AD$ | $\pi_{0,A}\,(1-\pi_{0,B})\,\pi_{A,D}\,\underline{(1-\pi_{A,C})}$ | 0.021 | 0.070 | 0.032 |
| ⋮ | ⋮ | | | |
| $ABD$ | $\pi_{0,A}\,\pi_{0,B}\,\pi_{A,D}\,\underline{(1-\pi_{A,C})}$ | 0.009 | 0.030 | 0.014 |
| $ACD$ | $\pi_{0,A}\,(1-\pi_{0,B})\,\pi_{A,D}\,\underline{\pi_{A,C}}$ | 0.049 | 0 | 0 |
| ⋮ | | | | |

Figure 1: **Example intervention and predicted genotype frequencies.** On the left, an OT model. On the right, column "Predicted frequency" gives the predicted genotype frequencies of some of the genotypes (in the absence of errors: $\varepsilon_+ = \varepsilon_- = 0$; see Appendix, section D.7). Our intervention consists of making mutations in gene $C$ lethal (i.e., $C$ is the target). On the genotype frequency expressions, we underline the terms affected by the intervention, those that involve $\pi_{A,C}$. As explained in the text, removing the edge leading to $C$ (and any descendant edges from $C$, if there were any) is equivalent to setting $\pi_{A,C} = 0$. For example, the predicted frequency of genotypes {A,C} and {A,C,D} becomes 0. But the effect of the intervention is not the same as conditioning: setting to 0 the frequencies of all genotypes that have $C$, and rescaling all genotypes to a total frequency of 1. For example, the frequency of genotype {B} should not be affected by the intervention on $C$, but the frequency of genotype {B} would increase if we were to remove all genotypes with $C$ and rescale to have frequencies sum up to 1. The effect of intervening on $C$ (either inactivating $C$ or making all genotypes with $C$ mutated lethal —section 2.2) is that some {A} genotypes that would become {A, C} stay as {A} (the frequency of {A} increases as $1-\pi_{A,C}$ becomes 1), some {A, B}s that would become {A, B, C} remain as {A, B}, etc. As a numerical example, columns "No_Int.", "Int." and "Cond." correspond, respectively, to the genotype frequencies under no intervention, intervention on $C$, and setting to 0 the frequencies of all genotypes that have $C$, and rescaling, for a model with $\pi_{0,A} = 0.5, \pi_{0,B} = 0.3, \pi_{A,C} = 0.7, \pi_{A,D} = 0.2$. Similar examples can be constructed with all other models.

Now, let $M_m$ denote the fitted model from EvAM method $m$. Suppose we want to answer the question "what would $M_m$ predict if we were to intervene on gene $C$?". A naive approach to predicting the consequences of interventions, removing genotypes with the targeted gene and rescaling, shown in Fig. 1, is wrong. This operation would amount to conditioning, computing $P(.|C=0)$ ("among tumors without $C$ mutated, what are the genotype frequencies?"), whereas we want $P(.|do(C=0))$ ("if we prevent $C$ from mutating, what do genotype frequencies look like?"). The two distributions differ here because $C$ shares causes with other events in the model; the collider $A$ in Fig. 3(a.2) shows where conditioning would additionally introduce collider bias.

# 3. Modifying fitted EvAM models to predict the consequences of an intervention

We examine interventions on EvAMs at two levels. First, what we call individual-level interventions (section 3.3), which are at the level of the reproducing individual —the cell, in the cancer setting—, where parent and offspring are distinct entities and fitness is an explicit variable: what Otsuka (2016, p. 257) calls the causal structure over individual organisms (in contrast to the population changes it gives rise to); in particular, "individual-level" does not refer to an individual patient or tumor. Next, we examine what we call model-level interventions (section 3.4), which use the models as originally defined. We start with the individual-level interventions, because they allow us to differentiate between two types of interventions (killing and inactivating). Only the individual level requires the mapping to fitness (section 3.3).

Before addressing each model in turn, in section 3.1 we provide a summary, not specifically tied to EvAMs, of structural causal models, interventions, and the modularity assumption. To avoid repetition in figure legends and model-specific sections, in section 3.2 we describe and define the setup and the objects that are common to all models, in particular the modified transition rate matrix, $Q_{-g}$, that most of the model-specific sections end up at.

## 3.1. Structural causal models, interventions, and modularity

A structural causal model (SCM) is a 4-tuple $M = \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}) \rangle$, where $\mathbf{U}$ is a set of exogenous variables, determined by factors outside the model; $\mathbf{V} = \{V_1, \ldots, V_n\}$ is a set of endogenous variables, determined by variables in the model; $\mathcal{F} = \{f_1, \ldots, f_n\}$ is a set of functions, the structural equations, that assign to every variable in $\mathbf{V}$ a value based on other variables in the model: $v_i = f_i(pa_i, u_i)$, where $PA_i \subseteq \mathbf{V} \setminus \{V_i\}$ are the parents of $V_i$ and $U_i \subseteq \mathbf{U}$; and $P(\mathbf{U})$ is a probability distribution over the exogenous variables (Correa and Bareinboim, 2020, Definition 1; Pearl *et al.*, 2016, section 1.5.1). Every SCM is associated with a graphical causal model or causal diagram. The causal diagram of SCM $M$ has one node per endogenous variable (and possibly nodes for some or all exogenous variables) and an arrow from each member of $PA_i$ to $V_i$ (Pearl *et al.*, 2016, section 1.5.1; Pearl, 2009, Definition 7.1.1; Correa and Bareinboim, 2020, Definition 1). Structural causal models are a non-linear, non-parametric generalization of the structural equation models (SEM) used in economics, the social sciences, and biology (e.g., Kline, 2015; Shipley, 2016); one key difference is that in a structural causal model each equation is explicitly read as a (causal) mechanism that assigns a value to the variable on its left-hand side, not as an algebraic equality (Pearl, 2009, section 1.4.1, p. 27, and section 5.1).

An atomic intervention $do(X = x)$ replaces the structural equation of variable $X$, $f_X$, by the constant function $X = x$, and leaves all other structural equations unchanged (Pearl, 2009, Definitions 7.1.2 and 7.1.3; Correa and Bareinboim, 2020). In the causal diagram, the intervention removes all the arrows into $X$ (the diagram is "mutilated"), since $X$ no longer depends on its parents; this differs from conditioning on $X = x$, as explained in section 2.3. Following Pearl (2009, section 3.2.2) and Correa and Bareinboim (2020), we represent an intervention explicitly with a node $\sigma_X$ and an arrow $\sigma_X \to X$ (the dashed nodes in Figs. 2 and 3); an atomic intervention is then $\sigma_X = do(X = x)$.

But not all interventions need to set a variable to a constant. A conditional intervention replaces $f_X$ by another function, $X = g(pa_X^*)$, and a stochastic intervention makes $X$ follow a new distribution $P^*(X \mid pa_X^*)$ (Pearl, 2009, section 4.2; Correa and Bareinboim, 2020, Table

1), where $pa_X^*$ denotes that the parents of $X$ in the modified graph could differ from the original graph. If the new function that produces $X$ keeps the original parents of $X$ (i.e., $pa_X^* = pa_X$), the intervention neither removes nor introduces arrows: the causal structure (the edges in the graph) stays the same, and only the mechanism that generates $X$ from its parents changes. Eberhardt and Scheines (2007) call these interventions "parametric" (also called "soft"), in contrast to "structural" interventions, such as *do*, that remove the arrows into $X$; Korb *et al.* (2004) call them "dependent", because their effect depends on the state of the pre-existing parents of $X$ (also Korb and Nicholson, 2011, section 3.8.3). The killing intervention (section 3.3) is of this kind.

Predicting the effect of an intervention from the modified model requires that the intervention changes only the equation of the targeted variable, and that all other structural equations remain as they were. This is the modularity assumption: each structural equation represents a stable and autonomous mechanism, so that "(...) it is conceivable to change one such relationship *without* changing the others" (Pearl, 2009, section 1.3.1, p. 22; see also Woodward, 2003, p. 48 and section 7.4). In graph terms, a *do* intervention deletes the arrows into the targeted variable; modularity is what lets us read the surviving arrows as keeping their original structural equations (i.e., no change in form or parameters). Modularity is an assumption about the system being modeled, not something a fitted model guarantees. For each EvAM and intervention kind, sections 3.3 and 3.4 state its structural equations and discuss whether modularity is plausible.

### 3.2. Interventions on EvAMs: common setup

In all the causal graphs in this paper (Figs. 2 and 3), a node labeled with a gene is a variable: the state of that gene is 0 if it is not mutated and 1 if it is mutated; in the individual-level causal DAGs there is a node for each gene for both parent and offspring. In all cases, mutations are gained and not lost, so the state of a gene can only change from 0 to 1; we use "event" for that change, the gain of the mutation, which is the usual term in the EvAM literature. In the individual-level causal DAG (Fig. 2), fitness, $W$, is also an endogenous variable, and the node represents the value of fitness for a particular state of the genes, i.e., for a particular genotype. The $\varepsilon$ terms are exogenous variables; we only draw exogenous terms in the individual-level causal DAG, but they are implicit in the other models (and we refer to them explicitly when writing the structural equations). The $\sigma$ nodes denote interventions, not variables. In the model-level graphs (Fig. 3) Root is kept in the DAGs for coherence with most representations in the literature; the Root node is not a random variable but a constant: it is always in its single state (the initial, wildtype state), so that in the structural equations the condition on Root is trivially satisfied and events whose only parent is Root are unrestricted (e.g., $T_g = U_g$ in CBN, $X_g = U_g$ in OT). The Root node helps emphasize that its child events are those that have no restrictions (i.e., those that can occur in the wildtype genotype, the genotype without any mutations in the genes considered).

In the individual-level causal DAG, the arrows that run from the parent's loci to the offspring's have, as mechanism, transmission (inheritance): the state of a locus in the offspring is a function of its state in the parent and of the mutation term (section 3.3). In the model-level graphs the arrows denote restrictions or effects on rates. We give the structural equations of each model in its section, and Table 1 summarizes, for each model, the endogenous variables, structural equations, intervention, and modularity.

At the individual level, one of the two interventions considered (killing) is a conditional intervention on fitness, $W$. In all remaining cases (including inactivating), the intervention is

atomic and fixes the state of a gene (or locus): $\sigma_A = do(A = 0)$ means that gene $A$ is fixed as unmutated. Both the conditional and the atomic interventions considered here are persistent: they are applied at $t = 0$ and held there throughout the evolutionary process.

To obtain the model predictions under intervention we construct $M_{m,-g}$, the model for method $m$ when we target gene $g$. Although we are dealing with different models, we modify them following a similar procedure that takes advantage of interpreting these models as structural causal models and assumes modularity (section 3.1).

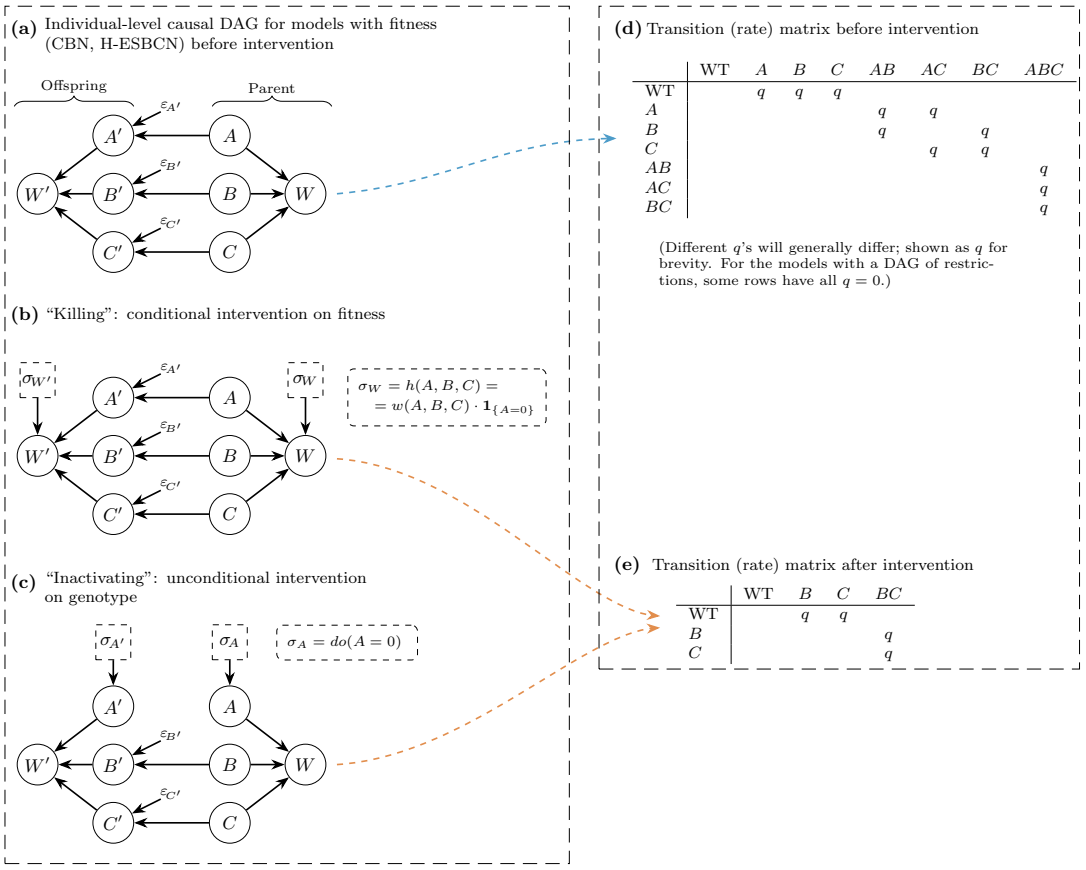
The interventions can, in most cases, be implemented in different equivalent ways, which we discuss in each section. But across the timed models, as well as HyperHMM and HyperTraPS, one of the equivalent ways of implementing the intervention yields the same kind of object: a modified transition (rate) matrix. For the three timed models (CBN, H-ESBCN, and MHN), the fitted model gives a continuous-time Markov chain over genotypes, with transition rate matrix $Q$. Under SSWM, this is also what the fitness landscape gives (section 3.3; Appendix, A.2). We write $Q$ with rows indexing the genotype of origin and columns indexing the genotype of destination: the entry in row $x$ and column $y$, $q_{x\to y}$, is the rate of transition from genotype $x$ to genotype $y$. The diagonal is implicit, as is standard: $q_{x\to x} = -\sum_{y\neq x} q_{x\to y}$, so each row sums to zero. Since events are gained one at a time, $q_{x\to y} = 0$ unless $y$ is $x$ with one more gene mutated. After intervening on gene $g$, no genotype with $g$ mutated can occur. We define $Q_{-g}$ as $Q$ restricted to the genotypes without $g$ mutated: we remove the row and the column of every genotype that contains $g$; the diagonal is again implicit, with the sum in $q_{x\to x} = -\sum_{y\neq x} q_{x\to y}$ now only over the genotypes retained in $Q_{-g}$. For HyperHMM and HyperTraPS, which return a conditional transition matrix instead of rates, the corresponding object is $R_{-g}$ (section 3.4.4). OT and OncoBN, being untimed, have no transition rate matrix (Appendix, *"Genotype predictions and hitting probabilities from EvAMs and error model" (section D.7)*). Fig. 2(e) and 3(f) show the $Q_{-A}$ that results from an intervention on $A$ (with panels Fig. 2(d) and 3(e) showing $Q$, the pre-intervention matrix).

For each case below, we state, in bold, the features introduced in section 3.1: the endogenous variables, the structural equations, the intervention and its implementation, and whether modularity is plausible. Table 1 presents a summary.

### 3.3. CBN and H-ESBCN: individual-level causal DAG interventions

Under the SSWM evolutionary regime, where the population is monomorphic except during brief clonal sweeps and evolution proceeds only uphill in fitness (Appendix, A.2), the CBN and H-ESBCN models implicitly specify a fitness landscape ("(...) the assignment of fitness values to genotypes that are connected by mutations" —Krug, 2021, p. 2) whose transition rate matrix is identical (up to a scaling constant) to that of the original EvAM model (Appendix, *"CBN and H-ESBCN: From DAGs of restrictions to fitness landscapes; equivalence of transition rate matrices", section A.3*).

This allows us to define a structural causal model, following Frank (1997), Otsuka (2016, 2019), Edelaar *et al.* (2023). The **endogenous variables** are the states of the loci of the parent, $A$, $B$, $C$, and of the offspring, $A'$, $B'$, $C'$, and the fitness of parent and offspring, $W$ and $W'$ (Fig. 2(a)). The **structural equations** pertain to fitness (of both parent and offspring) and offspring genotype (loci). For locus $A$ of the offspring, $A' = f_{A'}(A, \varepsilon'_A)$, where $\varepsilon'_A$ is the transmission error term, i.e., the mutation: the state of the offspring is that of the parent unless a mutation is gained. (As we code the unmutated state as 0, and since mutations are not reverted, $A' = \max(A, \varepsilon'_A)$, with $\varepsilon'_A \in \{0, 1\}$.) Analogous equations (with $A$ and $A'$ replaced by




| | WT | $A$ | $B$ | $C$ | $AB$ | $AC$ | $BC$ | $ABC$ |
|---|---|---|---|---|---|---|---|---|
| WT | | $q$ | $q$ | $q$ | | | | |
| $A$ | | | | | $q$ | $q$ | | |
| $B$ | | | | | $q$ | | $q$ | |
| $C$ | | | | | | $q$ | $q$ | |
| $AB$ | | | | | | | | $q$ |
| $AC$ | | | | | | | | $q$ |
| $BC$ | | | | | | | | $q$ |

| | WT | $B$ | $C$ | $BC$ |
|---|---|---|---|---|
| WT | | $q$ | $q$ | |
| $B$ | | | | $q$ |
| $C$ | | | | $q$ |

Figure 2: **Individual-level interventions on CBN and H-ESBCN (section 3.3).** Nodes are the states of the loci in the parent ($A$, $B$, $C$) and in the offspring ($A'$, $B'$, $C'$), and the fitness of parent and offspring, $W$ and $W'$; $\varepsilon_i'$ is the transmission error term (the mutation) for locus $i$. **(a)** Individual-level causal DAG before intervention (following Frank, 1997, Fig. 2; Otsuka, 2019, Fig. 5.1; Edelaar *et al.*, 2023). The DAG shows only genotype and fitness, as that is sufficient for these models; it could include phenotype or other mediating variables (e.g., Fig. 3 in Frank, 1997; Fig. 5.1 in Otsuka, 2019). **(b)** "Killing": conditional intervention on fitness, that replaces the original fitness landscape mapping ($w(A,B,C)$) by $\sigma_W = h(A,B,C)$, which sets to 0 the fitness of every genotype with $A$ mutated and leaves the causal structure intact. **(c)** "Inactivating": atomic *do* intervention on the genotype, $\sigma_A$ and $\sigma_{A'}$, which removes the arrows into $A'$. (With phenotype or other mediating variables, such as mRNA or protein, $\sigma$ would act on that variable.) **(d)** Transition rate matrix, $Q$, before intervention. **(e)** Transition rate matrix after intervening on $A$, $Q_{-A}$. $Q_{-A}$ is $M_{m,-g}$, with $g = A$, in step 3 of section 4.2. At $t = 0$ both interventions, (b) and (c), yield the same $Q_{-A}$.

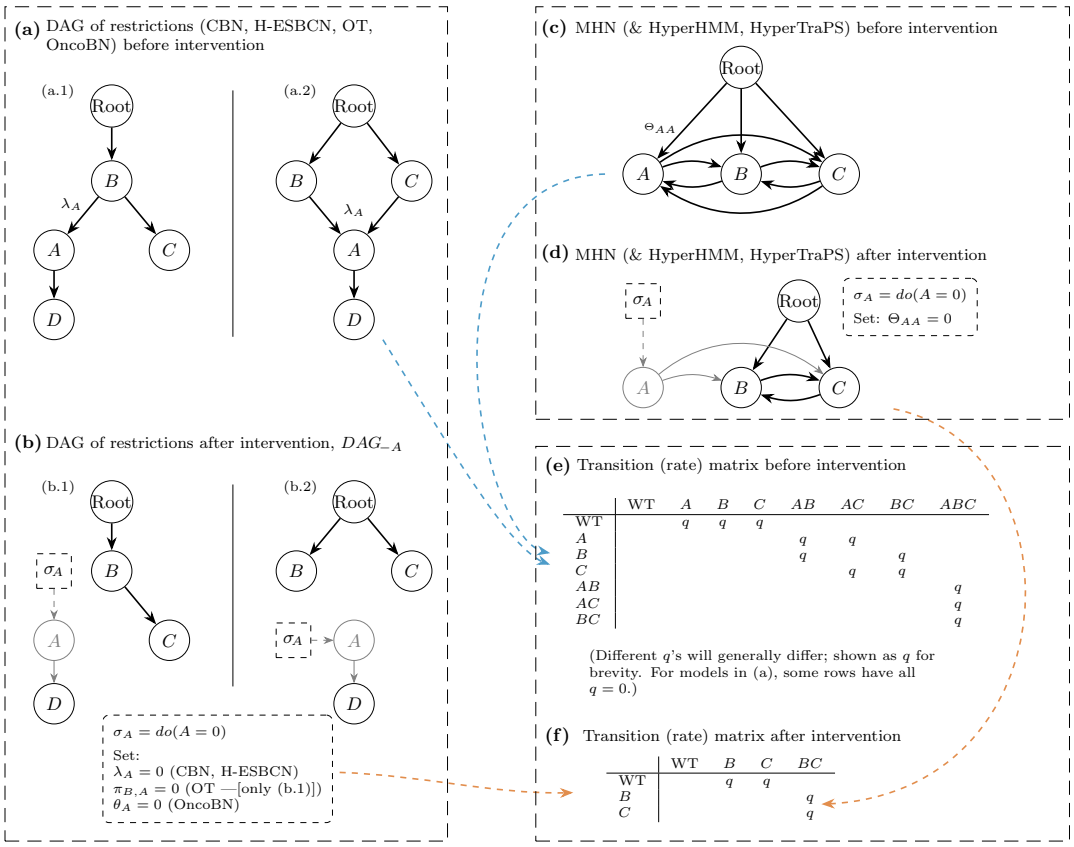



Figure 3: **Model-level interventions (section 3.4).** Nodes are the states of the genes; every intervention is the atomic $do(A = 0)$. **(a)** Before intervention: DAGs of restrictions for CBN, H-ESBCN without XOR, OT, OncoBN; for H-ESBCN with XOR the graph is not acyclic, since (a.2) has implicit arrows from $A$ to $B$ and $C$ (section 3.4.1). (a.2) cannot be represented by OT (two parents for $A$). Only $\lambda_A$ is shown annotating the edge into $A$; $\pi_{B,A}$ (OT) and $\theta_A$ (OncoBN) would annotate the same edge; this is the rate ($\lambda_A$) or the conditional probability ($\pi_{B,A}$, $\theta_A$) of a mutation in $A$ given that its parents are mutated, i.e., given that the restrictions specified by the DAG are satisfied. Four genes are shown, to include a descendant, $D$, of the intervened gene. **(b)** DAGs of restrictions after intervention, $DAG_{-A}$: only the arrows into $A$ are removed; $A \to D$ and the remaining parameters are unchanged, so $D$ can no longer arise. Implemented by setting the parameter of $A$ to 0 (sections 3.4.1 and 3.4.2). **(c)** MHN before intervention: $\Theta_{A,A}$ is the rate of $A$ when no other event is present. The same graph is used as a sketch for HyperHMM and HyperTraPS, where interactions can be of arbitrary order (sections 3.4.4 and 3.4.5). **(d)** MHN after intervention: the arrows into $A$ are removed; implemented by setting $\Theta_{A,A} = 0$ (section 3.4.3). For HyperHMM and HyperTraPS the intervention is implemented on the transition matrix, $R$, not on the graph. Both (c) and (d) have cycles but unfolded in time are acyclic (sections 3.4.3 and 3.4.4). **(e)** Transition rate matrix, $Q$, before intervention. **(f)** Transition rate matrix after intervening on $A$, $Q_{-A}$, the common endpoint of (b) and (d); this is $M_{m,-g}$, with $g = A$, in step 3 of section 4.2.

Table 1: **Interventions on EvAMs: summary of section 3.** In the model-level rows, killing and inactivating cannot be told apart, and the intervention is the atomic $do(A = 0)$ (section 3.4). Throughout the table, we use $A$ in lieu of the generic $g$ to facilitate comparison with Figs. 2 and 3. $Q_{-A}$ and $R_{-A}$ are defined in sections 3.2 and 3.4.4. What the intervention yields is the modified model, $M_{m,-A}$ (section 3.2).

| **Model (section; figure)** | **Endogenous variables** | **Structural equations** | **Intervention; implementation** | **Modularity, assumed on:** |
|---|---|---|---|---|
| Individual level: fitness from CBN and H-ESBCN under SSWM (3.3; Fig. 2) | States of the loci of parent ($A$, $B$, $C$) and offspring ($A'$, $B'$, $C'$); fitness of parent and offspring, $W$, $W'$. | Transmission: $A' = f_A(A, \varepsilon'_A)$; $\varepsilon'_A$ the mutation (same for $B'$, $C'$). Fitness: $W = w(A,B,C)$ (same for $W'$), with $w(.)$ the fitness landscape. | Killing: conditional, $\sigma_W = w(A,B,C) \cdot 1_{\{A=0\}}$: fitness when $A$ mutated set to 0, otherwise not modified. Inactivating: atomic, $do(A = 0)$, $do(A' = 0)$. | Killing: Fitness of the genotypes without $A$ mutated and transmission of every locus unchanged. Inactivating: fitness landscape unchanged, transmission of the other loci unchanged. |
| CBN and H-ESBCN (3.4.1; Fig. 3(a)–(b), (e)–(f)) | Waiting times $T_i$ until each event. Gene $i$ mutated at $t$ iff $T_i \leq t$. | AND: $T_A = \max(T_B, T_C) + U_A$, $U_A \sim \text{Exp}(\lambda_A)$, for $A$ with parents $B$, $C$. OR: $T_A = \min(T_B, T_C) + U_A$. XOR: $T_A = f_A(x, \varepsilon_A)$, where $x$ is genotype (XOR creates cycle), $\varepsilon_A$ uniform. | $do(A = 0)$, i.e., $T_A = \infty$. Implemented as $\lambda_A = 0$ ($DAG_{\lambda_A=0}$); equivalently $DAG_{-A}$ or $Q_{-A}$. | $\lambda_i$ as independent mechanisms. Holds provided $\lambda_i$ are rates. |
| OT and OncoBN (3.4.2; Fig. 3(a)–(b)) | States of the genes at the time of sampling, $X_A \in \{0,1\}$. | OT: $X_A = 1$ if $X_{parent(A)} = 1$ and $U_A = 1$. OncoBN: Conjunctive: $X_A = 1$ if all parents of $A$ are mutated and $U_A = 1$. Disjunctive: $X_A = 1$ if at least one parent is mutated and $U_A = 1$. Otherwise $X_A = 0$. $U_A \sim \text{Bernoulli}(\pi_{parent(A),A})$ (or $\theta_A$) | $do(A = 0)$. Implemented as $\pi_{parent(A),A} = 0$ ($DAG_{\pi_{parent(A),A}=0}$) —OT—, $\theta_A = 0$ ($DAG_{\theta_A=0}$) —OncoBN—. Equivalently $DAG_{-A}$. Untimed: no $Q$. | $\pi/\theta$ as independent mechanisms. Questionable: the $\pi/\theta$ conflate mechanism and sampling. |
| MHN (3.4.3; Fig. 3(c)–(f)) | States of the genes as a process in time, $X(t) \in \{0,1\}^n$ (the genotype). | Rate $q_{x \to x+A} = \Theta_{AA} \prod_{x_j=1} \Theta_{Aj}$; per event, $T_A = f_A(x, \varepsilon_A)$, $f_A$ inverse cdf of exponential with that rate, $\varepsilon_A$ uniform. Cyclic graph; acyclic when unfolded in time. | $do(A = 0)$. Implemented as $\Theta_{AA} = 0$; equivalently $\Theta_{-A}$ or $Q_{-A}$. | $\Theta_{ij}$ as independent mechanisms. More plausible than for OT and OncoBN, less well supported than for CBN and H-ESBCN. |
| HyperHMM (3.4.4; Fig. 3(c) as a sketch) | The genotype after $v$ steps, $X_v$: a process indexed by steps, over genotypes (the hypercube). | $X_{v+1} = f(X_v, \varepsilon_{v+1})$, with $\varepsilon_{v+1}$ uniform; $f$ is given by the per-step probabilities $r_{x \to x+j}$ (entries of $R$). No equation per event. | $do(A = 0)$: $r_{x \to x+A} = 0$, other per-step probabilities unchanged. Implemented as $R_{-A}$: the mass of the forbidden transition becomes a self-transition, $r_{x \to x} = r_{x \to x+A}$ (not renormalization). | The per-step transition probabilities. Heuristic argument: step as an attempt at mutation. Not sampling dependent: better grounded than for OT and OncoBN. |
| HyperTraPS (3.4.5; Fig. 3(c) as a sketch) | As HyperHMM. | As HyperHMM, with $R$ the HyperTraPS output transition matrix; the model structure fixes the form of the $r_{x \to x+j}$ as functions of the parameters. | As HyperHMM, on $R$, for every model structure. Zeroing parameters would renormalize the row and is not $do(A = 0)$. | As HyperHMM: on the transition probabilities, not on the parameters. |

$B, B'$, etc.) hold for the remaining loci. The structural equations for fitness are: $W = w(A,B,C)$ and $W' = w(A',B',C')$, with $w(.)$ the fitness landscape, that is, the function that gives the fitness of each genotype (and $w$ is the same function for parent and offspring).

The above causal model and its associated causal graph allow us to define interventions directly on the fitness landscape. **Intervention by killing** means setting the fitness of the affected genotypes (those bearing a mutation in the intervened-upon gene) to 0. This intervention by killing is not an atomic *do* intervention, but a conditional intervention (section 3.1): the original fitness mapping function, $w(.)$, is replaced by a new function, $h(.)$: $h(A,B,C) = w(A,B,C) \cdot 1_{\{A=0\}}$, where $1_{\{A=0\}}$ is the indicator function that takes value 1 when gene A is not mutated, and 0 otherwise. This function sets to 0 the fitness of every genotype with $A$ mutated and leaves all other fitness values unchanged; we denote it $\sigma_W = h(A,B,C)$ (notation as in Example 1 of Correa and Bareinboim, 2020, with $h$ instead of $g$, $W$ instead of $X$, and $A,B,C$ instead of $w$). This is shown in Fig. 2(b). The population-level dynamics then follow from the continuous-time Markov chain under SSWM, where the transition rates between genotypes are determined by the fitness landscape (Appendix, *"From fitness landscape to transition rate matrix" (section A.2)*).

In addition to "killing", it is worth considering a different intervention (section 2.2), shown in Fig. 2(c). This is the **inactivating intervention**; it involves inactivating a gene (or its products), and is the intervention mentioned in Schill *et al.* (2024, Fig. 3: "A drug treatment that suppresses event 2"). It is not a conditional intervention, but the usual atomic *do* intervention. In the notation of section 3.1, $\sigma_A = do(A = 0)$ and $\sigma_{A'} = do(A' = 0)$: the structural equations of $A$ and $A'$ are replaced by the constant functions $A = 0$ and $A' = 0$, which mutilates the DAG by removing the arrows into $A'$ from $A$ and from $\varepsilon'_A$. For example, inactivation of the mutant allele of gene $A$ makes the phenotype of any genotype with $A$ mutated identical to the phenotype of that same genotype without $A$ mutated. Mutations in $A$ are turned into neutral mutations (that could still fix by hitchhiking or drift —though not by drift under our strict SSWM model where we only move uphill in fitness) and no longer have any downstream consequences.

If the interventions are carried out at $t = 0$, before the evolutionary process starts, both killing and inactivating a gene are equivalent: compared to inactivating, killing leads to an increase in death rate because mutations that inactivation renders neutral lead to death under killing, but this difference is negligible for sensible values of mutation rate —recall we assume SSWM and thus $N\mu \ll 1$: Appendix, section A.2. Both inactivation and killing halve the size of the genotype space. Focusing on phenotypes, under both interventions the observable distinct states are the same, and the transition rates between phenotypes are the same (if the targeted gene is $A$, the population evolves in the sub-hypercube with $A = 0$). But killing and inactivating will not necessarily be identical if we intervene later. A simple example: when we intervene on $A$, populations made entirely of the $\{A,B\}$ genotype, under killing, will disappear; however, under inactivation of $A$ those populations do not disappear, but now look like a population of $\{B\}$s, that could then transition to $\{B,C\}$.

What about the implementation of the inactivating intervention? When this intervention is carried out at $t = 0$, we will never observe any genotype with $g$ mutated. Since these genotypes are now no longer part of the state space, this is the same as removing them from the transition rate matrix derived from the fitness landscape, i.e., $Q_{-g}$ (section 3.2). Under SSWM, this transition rate matrix is, therefore, identical to the one obtained by setting the fitness of the affected genotypes to 0. (If the intervention were carried out at a later $t$, a more complex algorithm would be required. Write $Q$ as in section 3.2, rows being the genotype of origin and columns the genotype of destination. We need to make two changes to $Q$, in this order: first, genotypes with $g$ mutated will have a 0 as destinations —i.e., all entries in columns

corresponding to genotypes with $g$ mutated would be 0; second, the row of every genotype $y$ with $g$ mutated mirrors the row of $x = y \setminus \{g\}$: the rate of $y$ gaining mutation $i \neq g$ is set to the original rate of $x$ gaining $i$. The diagonal is again implicit.)

What the intervention yields, in both cases, is a modified fitness landscape and, through it, a modified transition rate matrix. Under killing, the fitness landscape is $h(.)$ instead of $w(.)$; under inactivating the genotypes with $g$ mutated are no longer part of the state space. At $t = 0$, and under SSWM, the two give the same transition rate matrix, $Q_{-g}$ (section 3.2; Appendix, *"From fitness landscape to transition rate matrix" (section A.2)*). When the CBN or H-ESBCN model from which the fitness landscape was obtained is also available, $Q_{-g}$ is equivalent to the two implementations on the fitted model itself: setting the rate of the targeted gene to 0, $DAG_{\lambda_g=0}$, and removing the targeted gene and its strict descendants from the DAG of restrictions, $DAG_{-g}$ (section 3.4.1; Appendix, D.6).

**Modularity** requires, in both interventions, that the intervention changes only what it targets. Under killing, modularity requires that the intervention modifies only the fitness of the genotypes with $A$ mutated; fitness of genotypes not having $A$ mutated and the transmission of every locus must be unchanged. (Here, the fitness of each genotype is treated as a separate mechanism: $w(.)$ is a single function, but its values for the other genotypes must not change.) Under inactivating, the intervention replaces the equations of $A$ and $A'$; we thus require that the fitness landscape $w(.)$ be unchanged, and the transmission of the non-intervened-upon loci be unchanged.

## 3.4. Model-level interventions

Here, we intervene on the fitted models (not on an individual-level, fitness-based interpretation of the models, as in section 3.3). At this level, every intervention is $do(g = 0)$ and, as explained in section 3.4.1, killing and inactivation cannot be differentiated.

### 3.4.1. CBN and H-ESBCN: DAG of restrictions interventions

Independently of the individual-level causal DAG, CBN and H-ESBCN are structural causal models. The CBN model is also known as the noisy-AND causal model (Beerenwinkel *et al.*, 2007). For the (continuous-time) CBN and H-ESBCN, the **endogenous variables** are the waiting times $T_i$ until each event occurs; the exogenous terms are the $U_i$. The state of a gene follows from its waiting time, since mutations are not lost: gene $i$ is mutated at time $t$ if and only if $T_i \leq t$, so the genotype is a deterministic function of the waiting times and the sampling time. For example, in Fig. 3(a.2), the **structural equation** for the waiting time to event $A$ is given by $T_A = \max(T_B, T_C) + U_A$, where $U_A \sim \text{Exp}(\lambda_A)$ (e.g., eq. 3 of Gerstung *et al.*, 2009, or eq. 1 of Montazeri *et al.*, 2016). The same holds for H-ESBCN with AND; for H-ESBCN with OR, the structural equation replaces the max of the AND by min: $T_A = \min(T_B, T_C) + U_A$.

XOR relationships, however, do not admit a structural equation of this form because XORs, even if depicted as graphs that look like DAGs, create cycles. For the relationship in Fig. 3(a.2), having gained $A$ after gaining a mutation in one of the parents forbids gaining a mutation in the other parent (details in Appendix, section B). Thus, the XOR adds an implicit arrow from $A$ to $B$ and from $A$ to $C$. But, as for MHN (section 3.4.3), the cyclic graph unfolded in time (Peters *et al.*, 2017, sections 2.3.3 and 10.1; Pearl, 2009, sections 3.2.3 (Fig. 3.3) and 10.3.5) is acyclic. Therefore, for H-ESBCN with XOR, we state (as we will do for MHN, section 3.4.3) the structural equation for each event as a function of the current genotype, $x$, and not only of

its parents in the graph: $T_i = f_i(x, \varepsilon_i)$, with $\varepsilon_i$ uniform on $(0,1)$ and $f_i$ the inverse cumulative distribution function of an exponential whose rate is the rate of gaining $i$ from genotype $x$. For example, the rate of gaining $C$ is $\lambda_C$ if $A$ is not mutated in $x$; but if $A$ is mutated, the rate is 0 and, thus, $T_C = \infty$. In contrast, for AND and OR, $f_i$ depends on $x$ only through the parents in the graph and reduces to the max (AND)/min (OR) equations above. Strictly, then, the graph of an H-ESBCN with XORs is not a DAG. We nevertheless keep the established term "DAG of restrictions" (and the notation $DAG_{\lambda_g=0}$ and $DAG_{-g}$), with the understanding that, when XORs are present, the graph contains the implicit arrows described above.

For CBN and H-ESBCN, from the DAG of restrictions that specifies the model, Fig. 3(a), we use an atomic *do* **intervention** (which also involves DAG mutilation) fixing the gene as not-mutated. We can achieve this by setting $\lambda_A = 0$ (so $T_A = \infty$) thus preventing $A$ from ever gaining a mutation. This representation covers AND, OR, and XOR alike.

Setting $\lambda_g = 0$ in the model, $DAG_{\lambda_g=0}$, has the same consequences (in terms of genotypes that can exist and rates of transitions between genotypes) as eliminating, from the DAG that specifies the model, the node of the targeted gene and any other descendant nodes that strictly depend on the targeted gene (details and algorithm in Appendix, *"Interventions as modification of the DAGs: details and algorithm" (section D.1)*). We denote this modified DAG of restrictions as $DAG_{-g}$. In addition, a third equivalent procedure is removing, from the fitted transition rate matrix $Q$, all genotypes with gene $g$ mutated, yielding $Q_{-g}$, defined in section 3.2 (tests in Appendix, section D.3). In summary, $DAG_{\lambda_g=0}$, $DAG_{-g}$, and $Q_{-g}$ are equivalent ways of implementing the intervention and obtaining $M_{m,-g}$, which hold for all of AND, OR, and XOR. $DAG_{-g}$ is the default procedure we use in the code (Appendix, section D.1).

Those three procedures result in the same predictions as modifications of the fitness landscape (section 3.3; Appendix, *"Interventions on EvAMs as modifications of the fitness landscape" (section D.6)*), under SSWM, as we would expect from the structural causal model interpretation. This illustrates the expressive power of conceptualizing the interventions using individual-level causal DAGs, as well as the consistency of interventions guided by the modularity assumption across formalisms: the individual-level DAG (Fig. 2(a)-(c)) and the DAG of restrictions (Fig. 3(a)-(b)).

However, we cannot differentiate between "killing" and "inactivation" on the DAG of restrictions, in contrast to the individual-level causal DAGs. This is a consequence of the different granularity of the models in Figs. 2(a) and 3(a). Under Fig. 2(a) the model has the structure $Genotype \rightarrow Fitness \rightarrow Evolutionary\ Dynamics$; because both $Fitness$ and $Genotype$ are intermediate variables, we can intervene on each one separately (killing: conditionally change fitness; inactivation: unconditionally change genotype or phenotype) and they remain distinguishable. In contrast, the DAG of restrictions (and, as we will see, the MHN, HyperHMM, and HyperTraPS models) have the structure $Genotype \rightarrow Evolutionary\ Dynamics$. Fitness is, if at all, only implicit in the transition rate matrix, and the model structure is not expressive enough to differentiate the two interventions. There is no ambiguity if the intervention is performed before the process starts. But if the intervention is performed after the evolutionary process starts, we would need to specify which kind of intervention we are conducting. We would then need to modify not only the DAG, but also the transition rate of the existing populations: removing rows (killing) or transforming them (inactivating).

Because $\lambda_i$ are mechanism parameters that specify the rate at which $i$ happens (given its restrictions are met) **modularity** holds, provided the $\lambda$s describe independent mechanisms: we can intervene on one node, without affecting the remaining structural equations. For XOR, the same holds where the per-event mechanism depends on the current genotype rather than

only on the parents drawn in the graph. Thus, removing the transitions into genotypes with $g$ mutated leaves the other rates unchanged.

#### 3.4.2. OT and OncoBN: DAG of restrictions interventions

OT and OncoBN also specify restrictions as DAGs, but they are untimed models. OT and OncoBN give the probability of observing a particular genotype at the time of sampling under some sampling scheme. The **endogenous variables** are the states of the genes at the time of sampling: $X_g = 1$ if gene $g$ is mutated and $X_g = 0$ (unmutated) otherwise. The nodes of the DAG represent the variables $X_g$. For OT, "(...) each of the vertices of $V$ represent a certain mutation or other genetic alteration" and the alteration status of a tumor is the set of the vertices that are present (at the time of sampling) (Szabo and Boucher, 2008, Section 2.2, p. 3). For OncoBN, the observation of alterations is modeled as "(...) a binary random vector $(X_1, \ldots, X_p)$, where $X_j = 1$ if the $j$th alteration is detected in the sample" (Nicol *et al.*, 2021, Section 2, p. 4).

For both models, the **structural equations** give the state of each gene as a function of the states of its parents and of an independent Bernoulli draw; this is our restatement, in structural-equation form, of the original definitions (independent retention of each edge with probability $\pi$ in OT, Szabo and Boucher, 2008, Definition 1, p. 4; a conditional probability of $\theta_g$ given the parents in OncoBN, Nicol *et al.*, 2021, Section 2.1.1, p. 4). In OT, where every gene has a single parent, $X_g = 1$ if $X_{parent(g)} = 1$ and $U_g = 1$, and $X_g = 0$ otherwise; $U_g$ is a Bernoulli variable with parameter $\pi_{parent(g),g}$, the probability "of transition along the given edge by the time of observation" (Szabo and Boucher, 2008, p. 4). In OncoBN the equation depends on whether the model is conjunctive or disjunctive (Nicol *et al.*, 2021, define only the disjunctive rule in their paper, but their software implements both). In the conjunctive model, $X_g = 1$ if all the parents of $g$ are mutated and $U_g = 1$; in the disjunctive model $X_g = 1$ if at least one parent of $g$ is mutated and $U_g = 1$. In both, $X_g = 0$ otherwise, and $U_g$ is a Bernoulli variable with parameter $\theta_g$ (Nicol *et al.*, 2021, Section 2.1.1, p. 4). In both OT and OncoBN, predicted genotype frequencies are a function of both the conditional probabilities of the model and possible deviations from the models (denoted "spontaneous activation probability" in OncoBN); we set these to 0 in both models —details in Appendix, section D.7.

The **intervention** is the same atomic $do(g = 0)$ used for CBN and H-ESBCN, and it mutilates the DAG of restrictions in the same way (section 3.4.1). As with CBN and H-ESBCN, we also denote the modified DAG as $DAG_{-g}$. Also as for CBN and H-ESBCN, setting $\pi_{parent(g),g} = 0$, yielding $DAG_{\pi_{parent(g),g}=0}$ (OT), or $\theta_g = 0$, yielding $DAG_{\theta_g=0}$ (OncoBN), has the same consequences as eliminating, from the DAG that specifies the model, the node of the targeted gene, $g$, and any other descendant nodes that strictly depend on $g$ (see algorithm in Appendix, *"Interventions as modification of the DAGs: details and algorithm" (section D.1)* —as for CBN and H-ESBCN, and for implementation reasons, $DAG_{-g}$ is the default procedure in the code). In contrast to CBN and H-ESBCN, however, directly modifying fitness (as in Fig. 2(a)-(c)) is not possible with OT and OncoBN because of a lack of equivalence between these models and fitness landscapes (Diaz-Uriarte and Johnston, 2025). In summary, the results of the intervention, $\pi_{parent(g),g} = 0$ (OT) or $\theta_g = 0$ (OncoBN), can be expressed, equivalently, as $DAG_{\pi_{parent(g),g}=0}$ (OT) or $DAG_{\theta_g=0}$ (OncoBN), or $DAG_{-g}$. These are all equivalent ways of obtaining $M_{m,-g}$.

**Modularity** is harder to justify for OT/OncoBN than for CBN/H-ESBCN. Because OT/OncoBN give the probability of observing a particular genotype at the time of sampling under some sampling scheme, their parameters conflate the causal mechanism (e.g., the DAG + $\lambda$s

of CBN and H-ESBCN) with the sampling process. In CBN and H-ESBCN the $\lambda_i$s are invariant under changes in the sampling distribution: from the continuous-time Markov chain, we can obtain the probability of observing any genotype under arbitrary sampling distributions; therefore, with CBN and H-ESBCN modularity holds provided the $\lambda_i$s describe independent mechanisms, and mutilation of one part of the DAG should not affect the remaining structural equations. But OT/OncoBN are sampling-process dependent. If the sampling distribution were to change (say, from exponential to uniform), the $\pi$ parameters of OT would change even though the underlying causal mechanism has not. This interweaving of sampling and $\pi$ parameters makes it questionable that a DAG mutilation will not affect the remaining parameters. Still, this intervention where we hold the remaining parameters fixed is all we can do for the OT/OncoBN models without invoking additional assumptions.

#### 3.4.3. MHN: graph interventions

The **endogenous variables** are the states of the genes, taken as a process in time. MHN models tumor progression as "a continuous time Markov process $\{X(t), t \geq 0\}$ on all $2^n$ combinations of a predefined set of $n$ events", with state space $\{0,1\}^n$, where "$X(t)_i = 1$ means that event $i$ has occurred in the tumour by age $t$" (Schill *et al.*, 2020, Section 2.1, p. 242).

Fig. 3(c) shows an MHN model as a graph. This representation differs from Fig. 1(g) in Schill *et al.* (2020) by adding the Root node, similar to Fig. 3(a). We make the Root node explicit to highlight the similarity with the other models and emphasize that the $\Theta_{i,i}$ are analogous to the $Root \to i$ in the DAGs of restrictions models. This is also a structural causal model but the model for genes, in contrast to Fig. 3(a), includes cycles: gaining event $A$ has an effect on the probability of gaining $B$, but gaining $B$ also affects the probability of gaining $A$. A causal interpretation of this model is present in Schill *et al.* (2024), where they refer to these graphs as causal interaction networks between events (their Fig. 2), and ask what the model predicts under an intervention, "[a] drug treatment that suppresses event 2" (their Fig. 3).

Cycles can introduce several complications, including in the semantics of interventions (Bongers *et al.*, 2021). However, the MHN process is irreversible (mutations are gained, not lost) and sequential; unfolded in time (one copy of each variable per step, arrows only forward) it is acyclic (for how cyclic dependencies are unfolded in time into acyclic, time-indexed models see Peters *et al.*, 2017, sections 2.3.3 and 10.1; Pearl, 2009, sections 3.2.3 (Fig. 3.3) and 10.3.5); as a consequence, for a given genotype, the transition rates to other genotypes are well-defined and there is no simultaneity problem. We nevertheless show, and intervene on, the folded graph (Fig. 3(c) and (d)), for two reasons: it is the representation used by the MHN authors, and the intervention is persistent, so $do(A = 0)$ acts on every time-copy of $A$ in the same way; the mutilation is therefore the same at every step, and the folded graph shows it once.

The MHN model specifies that the rate at which gene $A$ is acquired is a deterministic function of the current mutation state of all other genes ($q_{x \to x+A} = \Theta_{AA} \prod_{x_j=1} \Theta_{Aj}$ — Schill *et al.*, 2020, eq. 2). Thus, although Schill *et al.* (2020) specify the rates, not an equation per event, we can write the **structural equations**, expressing the time to event $A$, given the state of the other genes, as a deterministic function plus an exogenous noise term: $T_A = f_A(x, \varepsilon_A)$, where $x$ is the current state of the genotype, $\varepsilon_A$ a uniform $(0,1)$ noise term, and $f_A$ the inverse cumulative distribution function of an exponential with parameter $q_{x \to x+A}$.

Thus, as for the DAGs of restrictions for CBN and H-ESBCN (section 3.4.1; Fig. 3(a)-(b)), the **intervention** is an atomic $do(A = 0)$ on the $A$ variable; we implement this intervention

by setting $\Theta_{AA} = 0$ (or, equivalently, $\theta_{AA} = -\infty$). This amounts to eliminating $A$ from the $\Theta$ matrix, yielding $\Theta_{-A}$, as no genotype with that gene mutated is possible. Equivalently, we can also remove, from the fitted transition rate matrix $Q$, all genotypes with gene $A$ mutated, yielding $Q_{-A}$ (defined in section 3.2, tests in Appendix, section D.3), as we did for CBN and H-ESBCN (section 3.4.1). These are all equivalent ways of obtaining the modified model, $M_{m,-g}$.

For simplicity, Fig. 3(c) shows the original model of Schill *et al.* (2020). But the reasoning above applies, without modification, to the extended MHN model in Schill *et al.* (2024) that corrects for collider bias: preventing $A$ from gaining a mutation —the $do(A = 0)$ intervention— propagates downstream (to effects on the observation rate or on the rates of acquisition of other events) through the remaining structural equations, provided modularity keeps them unchanged.

The **modularity** assumption in this case plays a similar role as for CBN and H-ESBCN: in so far as the $\Theta_{ij}$ describe independent mechanisms, when we intervene by preventing $A$ from being gained, we do not affect the remaining structural equations. The parameters of MHN cannot be interpreted in terms of fitness landscapes (Diaz-Uriarte and Johnston, 2025); still, MHN's parameters are directly interpretable in terms of rates and effects and invariant under changes in the sampling distribution (in contrast to OT and OncoBN). Therefore, the modularity assumption seems more plausible for MHN than OT and OncoBN, though not as well supported as for CBN and H-ESBCN when these two can be mapped to fitness landscapes.

#### 3.4.4. HyperHMM: interventions via the conditional transition matrix

The **endogenous variables** are the states of the genes, but taken as a process indexed by steps rather than by time, and over whole genotypes rather than one variable per gene: "[r]ather than focussing on individual traits/characters as the elements of the system, these approaches consider every possible state of a system involving $L$ traits", and "[t]he transition graph linking possible states is then a directed hypercubic graph" (Moen and Johnston, 2023, p. 2).

HyperHMM, similar to MHN, defines, for each genotype $x$, the probability of gaining a new mutation. The (implicit) structural equation would be similar, but instead of events having pairwise interactions, interactions can be of arbitrary order (i.e., gaining $A$ might be affected not just by the independent effects of the state of $B$ on gaining $A$ and the effect of $C$ on gaining $A$, but by the effect of the $B,C$ state on gaining $A$). In contrast to MHN, which returns the parameters of the structural equation function, HyperHMM does not give an equation per event. It returns a conditional transition matrix, $R$, where each row gives the probability of the next genotype given that there is a transition. The entries of $R$, $r_{x \to x+j}$, are the **structural equations** of the model (in the standard structural form, $X_{v+1} = f(X_v, \varepsilon_{v+1})$, where $X_v$ is the genotype after $v$ steps, $\varepsilon_{v+1}$ is an exogenous uniform $(0,1)$ term, and $f$ gives $x + j$ when $X_v = x$ and $\varepsilon_{v+1}$ falls in an interval of size $r_{x \to x+j}$). Fig. 3(c), the folded (cyclic) graph that correctly depicts the pairwise interactions in MHN, is only a sketch for HyperHMM, because in HyperHMM interactions can be of arbitrary order and there is no single parameter per arrow.

Now, when considering interventions, from the HyperHMM conditional transition matrix, we want to predict population-level genotype frequencies when some genotypes are now made lethal, i.e., when transitions to lethal genotypes are forbidden. Each step in a discrete-time process that we index by the number of mutations can be seen as an "attempt at mutation": a mutation arises, and if the resulting genotype is viable the attempt succeeds (and the population genotype transitions to a different state). Intervention on $g$ involves forbidding transitioning

to genotypes having $g$; this is achieved by setting $r_{x\to x+g}$ to 0. But because the entries of a row are constrained to sum to 1, "set $r_{x\to x+g}$ to 0 and leave the others unchanged" cannot be used to implement the intervention: the row no longer sums to 1. We therefore need an interpretation of $R$ under which the intervention is well defined: we interpret the entries of $R$ as probabilities per step, that is, per attempt, and not as probabilities conditional on a transition occurring. With lethal genotypes, some attempts fail to generate a viable transition and so the population remains in its current state. Without intervention every genotype the model can reach is viable, so every attempt succeeds, and the two interpretations (conditional probabilities and probabilities per step) cannot be told apart.

Therefore, the **intervention** $do(g=0)$ sets $r_{x\to x+g}$ to 0 and leaves the per-step probabilities of all other transitions unchanged: the probability, per step, of moving from $x$ to $x+j$ ($j\neq g$) is still $r_{x\to x+j}$. Since these no longer sum to 1, the remainder, $r_{x\to x+g}$, can only be the probability that the attempt at mutation fails and the population stays at $x$. This gives $R_{-g}$: the post-intervention $r_{x\to x}$ is equal to the pre-intervention $r_{x\to x+g}$, the post-intervention $r_{x\to x+g}$ is 0, and every other entry is unchanged. In terms of the structural equation, the values of $\varepsilon_{v+1}$ that led to $x+g$ now lead to $x$, and all the others lead where they did without intervention. In $R_{-g}$, unlike in $R$, a step may leave the genotype unchanged, and the number of steps is no longer the number of mutations. One could consider a different implementation: set $r_{x\to x+g}$ to 0 and divide the row by $1-r_{x\to x+g}$; this would give the correct destinations on leaving $x$, but would raise the per-step probability of every remaining transition, that is, it would speed up mechanisms that the intervention should not affect. That the per-step probabilities of the remaining transitions are unaffected is the modularity assumption, which we discuss below; it holds, for instance, under SSWM, where making $x+g$ lethal does not affect the mutation supply to the other genotypes (as the limiting factor for population-level transitions between genotypes is mutation rate).

We get exactly the same $R_{-g}$ if we think of an underlying process in continuous time. If we view the entries of $R$ as arising from underlying rates, $R$ is the embedded chain of that process: it gives where the process moves when it moves, but the clock, how often it moves, is lost. The intervention on the underlying process removes one rate, that of the transition to $x+g$, and keeps the others fixed: mutations to $x+g$ still arise at the same pace but now fail, and mutations to the other genotypes arise as before. Thus, the pace of mutation attempts does not change, and $R_{-g}$ is the transition matrix of the intervened process when each step is an attempt: attempts arise at the pace of the unintervened process, and all attempts to move to $x+g$ leave the population at $x$.

Based on that argument, we use the following procedure. Let $R$ be the original conditional transition matrix. Suppose we intervene on gene $g$. For a genotype $x$ that does not contain $g$ mutated, $r_{x\to x+g}$ is the probability of a transition from genotype $x$ to the now-lethal combination, $x+g$.

1. For every genotype $x$ (that does not contain $g$ mutated), set $r_{x\to x}$ to $r_{x\to x+g}$, and set $r_{x\to x+g}$ to 0 (i.e., genotypes that contain gene $g$ cannot be transitioned to). This yields $R_{-g}$ (the HyperHMM implementation of $M_{m,-g}$).
2. With $R_{-g}$ obtain $\hat{\mathbf{h}}_{-g}$, the predicted hitting probabilities after intervening in gene $g$.
3. With $R_{-g}$ obtain $\hat{\mathbf{f}}^{v}_{-g}$, the predicted distribution of genotypes after $v$ steps ($v$ taking all integer values from 0 to the number of loci), as per the usual $\hat{\mathbf{f}}^{v}_{-g}=\mathbf{f}^{0}\,R^{v}_{-g}$, where $\mathbf{f}^{0}$ is the (row) vector $(1,0,0,....)$.

4. Obtain the predicted population composition via the weighted sum $\hat{\mathbf{f}}_{-g} = \sum_{v=0}^{v=\text{number of loci}} \hat{\mathbf{f}}^{v}_{-g}\ P(V=v)$. $P(V=v)$ is the empirical (observed) frequency, in the training sample, of observations with $v$ mutations. Equivalently (since in the original, unintervened data, each step produces exactly one mutation), $P(V=v)$ is the frequency of samples that have undergone $v$ steps. Also because of this equivalence, the summation goes from 0 to the number of loci, $L$, which is the maximum number of steps any sample in the original data can have undergone. We keep $P(V=v)$ because the step is the only clock of the model, and we assume that the intervention does not change when samples are observed (similar to what we assume for the sampling time of the timed models).

   Post-intervention, $v$ no longer equals the number of mutations (some steps are self-transitions), but we keep the summation bounded by $L$: $P(V=v)$ is empirically defined from the unintervened training data, which has $v \leq L$; extending the sum beyond $L$ would require extrapolating $P(V=v)$ for $v > L$, for which there is no empirical grounding.

With HyperHMM, **modularity** is assumed for the per-step probabilities: forbidding one transition leaves the per-step probabilities of all the other transitions unchanged. Given the interpretation of a step as an attempt at mutation, the self-transition is not a further heuristic laid on top of modularity: it is necessary, because the row must sum to 1. What is a heuristic argument is the interpretation of a step, which the fitted $R$ cannot confirm, because without intervention every attempt succeeds. The mechanistic interpretability of this intervention, and the appeal to modularity, however, is arguably better grounded for HyperHMM than for OT and OncoBN: HyperHMM's transition probabilities are not sampling-regime dependent in the way that OT and OncoBN's parameters are. Compared to MHN, however, modularity is less well supported: MHN's $\Theta$s are rates, whereas HyperHMM's per-step probabilities are mechanisms only under the heuristic interpretation of a step as an attempt at mutation.

### 3.4.5. HyperTraPS

HyperTraPS, when used with cross-sectional data, returns as output a transition matrix on the hypercube structurally identical to HyperHMM's. This is the case regardless of the model structure (Aga *et al.*, 2024) used: zero-parameter, $L$-parameter (independent, non-interacting features), $L^2$ or pairwise interactions (similar to MHN), $L^3$ (three-way interactions), $L^4$ (four-way interactions), and independent rates on each edge of the hypercube (like HyperHMM). The **endogenous variables** and the **structural equations** are therefore those of HyperHMM (section 3.4.4): the genotype as a process indexed by steps, and the per-step transition probabilities, $X_{v+1} = f(X_v, \varepsilon_{v+1})$. Model structures differ in the form of these probabilities as a function of the parameters: from independent effects of each gene to interactions of arbitrary order (Aga *et al.*, 2024, p. 8). Therefore, to intervene on gene $g$, and regardless of model structure or whether regularisation has been used during fitting, we implement the **intervention** $do(g=0)$ using the exact same procedure as for HyperHMM, section 3.4.4, with $R$ replaced by HyperTraPS's output transition matrix. Even in models where transitions are expressed in terms of individual parameters —from independent effects to four-way interactions—, zeroing a parameter, as we do for MHN, would not correctly implement $do(g=0)$. HyperTraPS returns transition probabilities, not rates, so if we zeroed parameters, the row of the intervened genotype would be normalized to sum to 1 when the matrix were recomputed from the modified parameters. As a consequence, the probability of the forbidden transition would be spread over the remaining transitions; this would be the renormalization that we explained in

section 3.4.4 was inappropriate —and, as detailed there, the probability of the now forbidden transition has to be assigned to a self-transition.

The **modularity** assumption is the same as for HyperHMM: forbidding one transition leaves the per-step probabilities of the other transitions unchanged (and the heuristic interpretation of a step as an attempt at mutation is what makes the intervention well defined). The above holds too when HyperTraPS is fitted with a restricted model structure. In this case, the transition probabilities are functions of a small set of parameters shared across genotypes; but (as explained above) we intervene on the transition probabilities themselves and hold the other transition probabilities fixed: modularity is assumed of the transition probabilities, as for HyperHMM, and not of the parameters, as for MHN (and no particular change in the parameters might be able to produce the intervened matrix, $R_{-g}$, of HyperTraPS).

## 4. Quantifying intervention effects: intervention objectives

Section 3 discusses how to modify the models to obtain predictions under interventions. To properly assess if EvAMs can be used to identify therapeutic targets, we also need to specify what "identify" means, as different objectives can lead to different rankings of the same genes as targets.

Here we propose three intervention objectives that reflect three different ways of quantifying intervention effects and can be used in studies that evaluate the usefulness of EvAMs. The ground truth of these intervention objectives can be measured from empirical data (e.g., from clinical experimental studies) or from simulated data (e.g., by rerunning evolution on fitness landscapes where we modify the fitness of specified genotypes). The predicted values come from the EvAM methods themselves, independently of any fitness landscape interpretation: an EvAM yields hitting probabilities or genotype frequencies after intervention via the method-specific manipulations described in section 3, and these are what we compare against the ground truth. We can then assess the usefulness of these approaches with respect to a given objective: an EvAM will be useful to identify targets if it can provide a ranking of genes, with respect to the specified objective, that is very similar (or highly correlated) to the true ranking of genes defined from the ground truth. We first describe the objectives, and next explain how to evaluate the performance of EvAMs in simulation studies; a protocol is provided in the Appendix *"Assessing intervention objectives: a protocol" (section C)*.

**Increase/decrease the probability of going through (hitting) a given genotype (O_genotype):**
Minimize or maximize the probability that evolution, starting from the WT genotype, will ever go through, or hit, a specific genotype; for example, minimize for genotypes related to metastasis, and maximize for genotypes that can be easily targeted by other means. The simplest and most effective way to avoid a given genotype would be to make lethal any of the genes that it has mutated, but we do not consider this a suitable intervention, because it is trivial. Instead, we can focus on interventions on genes, different from the ones mutated in the genotype of interest, that can cause changes in the hitting probability of the genotype of interest by operating via the causal dependencies in the acquisition of genes.

**Decrease overall frequency of mutations or alterations (O_mut):**
Make the average number of mutations as small as possible.

**Increase frequency of the non-mutated, starting genotype, WT (O_WT):**
Delay the accumulation of altered genotypes, regardless of the number of mutations.

The first objective, O_genotype, involves steering evolution in specific directions (to either avoid, or go through, particular stages). O_genotype is independent of sampling time (it is the hitting, or first passage, probability of the underlying Markov chain). O_WT and O_mut are unavoidably dependent on sampling time, since both are obtained from the genotype frequencies at a particular time (section 4.4). Details on how to assess these three intervention objectives are provided in the Appendix, *"Assessing intervention objectives: a protocol" (section C)*. We are not considering measures related to intra-tumor heterogeneity or diversity. The EvAMs considered here cannot be used to provide these predictions, since they are either models of between-subject bulk samples or between-subject single cell samples (Diaz-Uriarte and Johnston, 2025); furthermore, if we assume SSWM, within-tumor heterogeneity is 0 (except for brief periods of fast clonal sweep).

## 4.1. Obtaining the ground truth of intervention effects in simulation studies

If we assume SSWM, for any fitness landscape we can obtain the transition rate matrix $\mathcal{Q}$ under SSWM (Appendix, *"From fitness landscape to transition rate matrix", section A.2*). From $\mathcal{Q}$, and via competing exponentials, we can obtain the embedded discrete-time chain and from its transition matrix the hitting (or first passage) probabilities (Norris, 1997; Privault, 2018), $\mathbf{h}$, starting from the WT state. From $\mathcal{Q}$ we can also obtain the distribution of genotypes, $\mathbf{f}$, at any specified time. Assuming SSWM does not require us to focus on additional complicating factors such as bulk sequencing (Diaz-Uriarte and Johnston, 2025), mutation rates, or population size. This simplification is arguably appropriate to start gaining an understanding of the potential use of these methods to identify therapeutic targets and guide interventions (see also Nichol *et al.*, 2015).

Then, we can obtain the hitting probabilities and distribution of genotypes after intervening on each gene in turn. For each gene $g$, this is done by first setting the fitness of any genotype that carries a mutation in the target gene $g$ to 0, and then obtaining the transition rate matrix (and transition matrix of the embedded discrete-chain) that corresponds to this new fitness landscape. We denote by $\mathcal{Q}_{-g}$ the true transition rate matrix after intervening on gene $g$, i.e., after making all genotypes with a mutation in gene $g$ lethal. Using the transition matrix of the embedded discrete-time chain we can obtain the hitting probabilities under intervention ($\mathbf{h}_{-g}$); using the transition rate matrix, we can obtain the distribution of genotypes ($\mathbf{f}_{-g}$), according to a prespecified sampling scheme; these constitute the ground truths of the hitting probabilities and distribution of genotypes that correspond to targeting gene $g$; see Fig. 4.

Alternatively, if we are not willing to assume SSWM, we can use any forward genetic simulator that operates on fitness landscapes and includes flexible sampling mechanisms, such as OncoSimulR (Diaz-Uriarte, 2017) or TumorSim.jl (Fontaneda and Diaz-Uriarte, 2023). Using a sensibly large number of simulated runs we can obtain estimates of $\mathbf{h}, \mathbf{f}$, as well as $\mathbf{h}_{-g}, \mathbf{f}_{-g}$ for every gene.

## 4.2. Obtaining the predictions of intervention effects from EvAMs

The predictions from each EvAM method can be obtained as follows:

1. From the true $\mathcal{Q}$ (section 4.1) (or, if $\mathcal{Q}$ is not available for the evolutionary regime considered, using simulations), generate a sample.

Genotype frequencies

Truth

| Targets | | WT | A | B | ⋯ | ABCD | ⋯ |
|---|---|---|---|---|---|---|---|
| None | $\mathbf{f}$ | $f(\text{WT})$ | $f(A)$ | $f(B)$ | | $f(\text{ABCD})$ | |
| A | $\mathbf{f}_{-A}$ | $f_{-A}(\text{WT})$ | 0 | $f_{-A}(B)$ | | 0 | |
| B | $\mathbf{f}_{-B}$ | $f_{-B}(\text{WT})$ | $f_{-B}(A)$ | 0 | ⋯ | 0 | ⋯ |
| ⋮ | | | | | | | |
| E | $\mathbf{f}_{-E}$ | | | ⋯ | | $f_{-E}(\text{ABCD})$ | ⋯ |
| ⋮ | | | | | | | |

$\text{CPM}_1$

| Targets | | WT | A | B | ⋯ | ABCD | ⋯ |
|---|---|---|---|---|---|---|---|
| None | $\hat{\mathbf{f}}_1$ | $\hat{f}_1(\text{WT})$ | $\hat{f}_1(A)$ | $\hat{f}_1(B)$ | | $\hat{f}_1(\text{ABCD})$ | |
| A | $\hat{\mathbf{f}}_{1,-A}$ | $\hat{f}_{1,-A}(\text{WT})$ | 0 | $\hat{f}_{1,-A}(B)$ | | 0 | |
| B | $\hat{\mathbf{f}}_{1,-B}$ | $\hat{f}_{1,-B}(\text{WT})$ | $\hat{f}_{1,-B}(A)$ | 0 | ⋯ | 0 | ⋯ |
| ⋮ | | | | | | | |
| E | $\hat{\mathbf{f}}_{1,-E}$ | | | ⋯ | | $\hat{f}_{1,-E}(\text{ABCD})$ | ⋯ |
| ⋮ | | | | | | | |

Figure 4: **True and predicted hitting probabilities and genotype frequencies under interventions.** As explained in the text, from the transition rate matrix, $\mathcal{Q}$, of the true fitness landscape we can obtain the true distribution of genotypes under no intervention (row "None"), which yields vector $\mathbf{f}$, with individual entries for each genotype $f(WT), f(A), \ldots$; we can also obtain the corresponding genotype frequency vectors when intervening on each possible gene ($\mathbf{f}_{-A}, \mathbf{f}_{-B}, \ldots$). Similarly, we can obtain the hitting probabilities under no intervention, $\mathbf{h}$, and under each possible intervention ($\mathbf{h}_{-A}, \mathbf{h}_{-B}, \ldots$). (For simplicity, this table only shows genotype frequencies, but a similar construction exists for hitting probabilities.) If using $\mathcal{Q}$ is not possible, we can use forward genetic simulators to estimate $\mathbf{f}$ and $\mathbf{h}$. For each EvAM $m$ (for simplicity, only one shown), after fitting the corresponding model, we can obtain the predicted genotype frequencies $\hat{\mathbf{f}}_m$ and hitting probabilities $\hat{\mathbf{h}}_m$; after modifying the fitted models to predict the consequences of an intervention (section 3), we obtain the predicted genotype frequencies ($\hat{\mathbf{f}}_{m,-A}, \hat{\mathbf{f}}_{m,-B}, \ldots$) and hitting probabilities ($\hat{\mathbf{h}}_{m,-A}, \hat{\mathbf{h}}_{m,-B}, \ldots$) under each intervention. In the table, some entries are necessarily 0, as shown (e.g., when intervening on $g = A$, all genotypes with A mutated have frequency, and hitting probability, of 0).

2. Use this sample as input for each EvAM method.

3. From the output of each EvAM, obtain a modified EvAM after intervening on (making lethal the mutated allele of) gene $g$. Let $M_m$ denote the fitted model from method $m$; $M_{m,-g}$ is the modification of $M_m$ that results from the killing intervention on $g$ (i.e., after the intervention that makes a mutant in gene $g$ a lethal mutation). How we obtain $M_{m,-g}$ is explained in section *"Modifying fitted EvAM models to predict the consequences of an intervention" (section 3)*.

4. From $M_{m,-g}$, use the standard procedures for each method (Appendix, *"Genotype predictions and hitting probabilities from EvAMs and error model", section D.7*) to obtain

the predicted hitting probabilities, $\hat{\mathbf{h}}_{m,-g}$, and the predicted distribution of genotypes, $\hat{\mathbf{f}}_{m,-g}$, after targeting gene $g$. For example, in Fig. 4 this yields each of the rows after "None" in panel "$\text{CPM}_1$".

Hitting probabilities do not depend on sampling time; they are obtained from the transition matrix of the embedded discrete-time chain for CBN, H-ESBCN, and MHN, and from the transition matrix of HyperHMM; for OT and OncoBN we use a heuristic procedure similar to Diaz-Uriarte and Vasallo (2019) —Appendix, *"Genotype predictions and hitting probabilities from EvAMs and error model" (section D.7)*. Appendix, *"OT, OncoBN, CBN, H-ESBCN, MHN: transition matrix from intervening on the model versus restricting and renormalizing" (section D.4)*, shows that intervening on the model and restricting the transition matrix of the model without intervention to the genotypes without $g$ (and renormalizing rows) give the same result.

### 4.3. Ranking of interventions

For each of these three objectives we could focus on the causal effect defined as the difference in one statistic (e.g., the hitting probability of some genotype), between the non-intervened regime and the intervention in each gene —a framing in the spirit of Hernán and Robins's "causal effects in populations, that is, numerical quantities that measure changes in the distribution of an outcome under different interventions" (2020, p. v). But, if the objective is to prioritize intervention candidates, it is more useful to reframe the problem as one of ranking interventions.

For O_mut and O_WT the most promising EvAMs would be those that rank the interventions, with respect to those two measures, in the same way as the truth. For O_genotype, the focus is not to compute the rank correlation between $\mathbf{h}_{-g}$ and $\hat{\mathbf{h}}_{m,-g}$; this would emphasize the agreement between the hitting probabilities of all the genotypes for each particular $g$. Instead, for each genotype whose frequency we might want to increase or decrease, compute the rank correlation between the true and the predicted hitting probability of that genotype under all possible interventions. This is the procedure that directly answers "what should I target if I want to increase/decrease the hitting probability of this genotype". Details in Appendix, section C.1.

### 4.4. Predicted genotype frequencies: at what time?

True genotype frequencies will likely depend on time for both clinical trial data (when patients are measured) and simulated data (when samples are obtained from the evolutionary process in section 4.1). For EvAMs, predicted genotype frequencies can depend on time for some models.

The simplest case is when samples under intervention are obtained under a sampling regime (e.g., exponential with some unknown rate, uniform, etc) that is similar to the sampling regime of the non-intervened processes that generated the training data for the EvAMs. In this case, for CBN, H-ESBCN, and MHN, the predictions of the intervened-upon models can use the standard "exponential with rate 1" model: these models are fitted assuming that the true sampling distribution is an exponential of rate 1, regardless of the true sampling distribution (which would not be known to the method anyway). Therefore, for these models we can obtain the predicted genotype frequencies as per equation (4) in Schill *et al.* (2020), assuming sampling

time is distributed as an exponential of rate 1. For OncoBN and OT, being untimed models, we can obtain the distribution of genotypes as per their standard expressions (which do not include time), setting the model error to 0. It is not possible for these models to match the time at which genotype predictions are made from EvAMs and the true sampling time. For HyperHMM and HyperTraPS, the scheme in *"HyperHMM: interventions via the conditional transition matrix" (section 3.4.4)* already incorporates obtaining samples weighted by $P(V = v)$, the number of steps.

If the ground truth of interventions is obtained under a sampling scheme that differs substantially from the sampling scheme of the training data, the above procedure is not adequate. For CBN, H-ESBCN, and MHN, we can obtain predicted genotype frequencies at arbitrary times from $Q$ (as $\pi^t = \pi^0 \exp\{Qt\}$, where exp is the matrix exponential, $\pi^0$ is the row vector of the population composition at time 0, with 1 for WT and 0 for the remaining elements), but it would be necessary to express the $t$ of the true sampling time (or distribution of times) in the scale of the $t$ of the EvAMs model (recall that these were fitted assuming sampling with exponential rate 1). For HyperHMM and HyperTraPS an analogous adjustment of the $P(V = v)$ would be needed to match the "true steps" that correspond to the ground truth. No such adjustments seem directly possible for OT and OncoBN. Further work to address adjusting the time-dependent predictions of EvAMs is warranted, but this is independent of the intervention framework in section 3.

## 5. Discussion

After differentiating between two types of intervention (killing and inactivating) and explaining why a naive approach to intervention leads to incorrect predictions for EvAMs, we have presented a conceptualization of interventions on each of the currently available EvAM methods (OT, OncoBN, CBN, H-ESBCN, MHN, HyperHMM, HyperTraPS). We show both what an intervention represents, and the equivalent ways of implementing the intervention. We then provide a protocol for assessing how useful EvAMs can be for ranking interventions, i.e., choosing therapeutic targets. The only prior work that has considered intervention in this context is Schill *et al.* (2024, Fig. 3), who considered suppressing an event (equivalent to our inactivating intervention) in MHN; our framework generalizes this to all EvAM families and additionally distinguishes it from the killing intervention.

Our framework is therefore relevant for using EvAMs to design interventions (Hosseini *et al.*, 2019) and, more generally, for making evolutionary predictions with the aim of driving evolution towards, or away from, specific directions (Wortel *et al.*, 2023). In fact, our work focuses on EvAMs, but the general procedure is not specific to cancer or EvAMs: we take a class of fitted computational models, interpret them as structural causal models, and derive principled intervention predictions using Pearl's (Pearl, 2009) framework, with a careful model-by-model analysis of where modularity is and isn't justified. This generalizes beyond EvAMs, and is a template for any domain where fitted models can encode causal structure.

As well, the entities that the model applies to, as emphasized in Diaz-Uriarte and Johnston (2025), are essential to the conceptualization of some of the interventions. The individual-level causal DAGs of Frank (1997) and Otsuka and collaborators (Otsuka, 2016, 2019; Edelaar *et al.*, 2023), in those cases where we can apply them to EvAMs, make fitness an explicit node in the causal graph. And this is what enables the distinction between killing interventions and inactivating interventions —a distinction that cannot be expressed in the DAG-of-restrictions or matrix-of-effects (MHN) or transition matrix (HyperHMM/HyperTraPS) formalisms. This

is a consequence of these models being, at their most basic, models of accumulation: they describe what events occur and in what order, not necessarily the fitness mechanism that drives selection. Other interpretations of these models are possible (such as frequency-dependent fitness; see discussion in Diaz-Uriarte and Johnston, 2025), but none of them introduce fitness as an explicit variable on which one can intervene separately. When certain conditions are satisfied (SSWM —Appendix, section A.3), some of these models (CBN, H-ESBCN) can be mapped to a fitness landscape, enabling the use of individual-level causal DAGs where fitness is an explicit variable and the killing/inactivating distinction becomes expressible.

The assessment of modularity, a crucial assumption in the interventionist account of causation (Woodward, 2003; Pearl, 2009), is also closely related to the entities being modeled. The individual-level causal DAGs allow us to think of interventions as surgical modifications of the fitness landscape, where we make lethal all genotypes that carry a given mutation, leaving the fitness of the rest of the genotypes untouched. This makes the feasibility of the intervention (to follow the dictum "no causation without manipulation" —Holland, 1986; Hernán and Robins, 2020) and the plausibility of the modularity assumption easy to scrutinize. For instance, under other evolutionary scenarios, like frequency-dependent fitness, this same conceptualization and modularity assumption would be questionable. That said, the individual-level causal DAG and the killing/inactivating distinction it supports do not depend on the SSWM assumption: SSWM is required only for the equivalence between fitness-landscape modifications and EvAM parameter manipulations in CBN and H-ESBCN.

This framework can be applied without modification or extended to deal with several other scenarios. Interventions that affect multiple genes are expressible in our framework; for DAG-of-restrictions, matrix of effects of MHN, and transition matrices we can only use the *do* operator, but if individual-level causal DAGs are possible, then we can differentiate between killing and inactivating. This opens the door to searching for optimal combinations of targets.

In those scenarios where the individual-level causal DAG holds, we can extend the conditional intervention to any gene with a more general intervention function, $\sigma_W$ that takes an extra argument, say $L \in \{0, A, B, \ldots\}$, so $L = 0$ denotes no intervention and therefore when $L = 0$, $h = w(A,B,C)$; when $L = A$, $h = w(A,B,C) \cdot 1_{\{A=0\}}$; when $L = B$, $h = w(A,B,C) \cdot 1_{\{B=0\}}$; etc. $L$ is not a parent of the original $W$, but conditional intervention functions can have as inputs variables other than the original parents (Correa and Bareinboim, 2020). The framework of Correa and Bareinboim (2020) (see also Eberhardt and Scheines, 2007; Korb *et al.*, 2004; Korb and Nicholson, 2011) would also allow us to consider stochastic interventions, resulting in partial inactivation or fitness modification of a fraction of the population. Similarly, we can incorporate (possibly conditional) activation interventions, $do(A = 1)$, which would enable modeling events detrimental to the progression; we do not develop this here.

As well, interventions that are expressible in the DAG-of-restrictions, matrix of effects of MHN, and transition matrices of HyperHMM/HyperTraPS can be applied without modifications to EvAM models used with cohorts of intra-tumor phylogenetic trees, such as TreeMHN (Luo *et al.*, 2023) and HyperTraPS (Aga *et al.*, 2024), and only with minor modifications to models with reversible transitions (Johnston and Diaz-Uriarte, 2025; Johnston, 2026).

The minor modifications needed for reversible models concern the implementation of interventions, not the causal interpretation. Reversibility introduces transitions in both directions between states, which produces cycles in the influence-diagram summary. But in the underlying structural causal model —where events are indexed by time— the structural equations remain acyclic: each transition at time $t$ depends only on the state at times $s < t$. Reversibility allows the population to revisit states; it does not allow events to cause themselves. Thus the

concerns raised in Bongers *et al.* (2021) do not apply here: reversibility is a property of the state-transition graph, not of the underlying causal structure.

A more substantive change concerns which outcome measures remain meaningful under reversibility. In irreversible accumulation models, hitting probabilities are informative because they reflect which paths the chain commits to. In a finite reversible chain on a connected state space, by contrast, hitting probabilities are not useful since they are 1 for any accessible genotype given enough time. The natural outcome measures shift to stationary distributions and time-indexed quantities at biologically relevant timescales. Adapting the present ranking-based evaluation protocol to these alternative outcome measures is straightforward in principle (one ranks targets by their effect on the chosen measure rather than on hitting probability), but the choice of outcome measure itself is no longer obvious and may need to be problem-specific.

Finally, the current framework presents several limitations, worth further theoretical work. It is unclear how, or even if, stochastic interventions —which are readily considered in the individual-level causal DAGs— are expressible in the DAG-of-restrictions, matrix of effects of MHN, and transition matrices of HyperHMM/HyperTraPS. Conditional and time-varying treatment strategies that depend on repeatedly monitoring and possibly correcting the course of action based on the state of the system, such as adaptive therapy (Hansen and Read, 2020; West *et al.*, 2023), would require a more complex causal framework (Hernán and Robins, 2020; Tsiatis *et al.*, 2020; Chakraborty and Moodie, 2013), even if we can use individual-level causal DAGs. Of course, interventions that do not map cleanly to the conditional or the $do(g = 0)$ operation cannot be represented in the current framework. Furthermore, (severe) lack of modularity, such as an intervention on gene $g$ affecting the parameters for gene $h$, would invalidate the framework; assessing the plausibility of this assumption is context-specific.

## 6. Acknowledgments

Álvaro San Martín and Eric Macías Fasio, for literature searches, and initial discussion. Members of the Stochastic Biology group at the University of Bergen for discussion. We thank Jürgen Degenfellner, ZHAW Zurich University of Applied Sciences, Winterthur, for independently running the test suite, identifying a flaw in the test gate and a missing dependency, and verifying the Figure 1 computations by hand. We thank an anonymous reviewer for comments that have substantially improved the manuscript. Supported by grants PID2024-156888OB-I00 funded by MICIU /AEI/10.13039/501100011033 / FEDER, EU and PID2019-111256RB-I00 funded by MCIN /AEI/10.13039/501100011033 to RDU. This work was supported by the Trond Mohn Foundation [project HyperEvol under grant agreement No. TMS2021TMT09 to IGJ], through the Centre for Antimicrobial Resistance in Western Norway (CAMRIA) [TMS2020TMT11].

### AI use

AI software was used in preparing this work. Claude (Sonnet 4.6; Opus 4.6, 4.7, 4.8; and Fable 5.1) was used for text editing (grammar, spelling), and for generating LaTeX/TikZ code from the authors' hand-drawn figure sketches. Claude Code (using the same models) was used in software development, including writing portions of the test suite and some functions, verifying computational outputs, text editing (grammar, spelling, improving readability), initial preparation of a LaTeX table, cross-checking the revised text against the reviewer's comments,

and cross-referencing direct quotes against primary sources. All AI-generated text and code were reviewed and verified by the authors, who are solely responsible for the work.

## 7. Code and data availability

This article is theoretical and has no associated primary dataset. The source code implementing the intervention methods is archived in Zenodo, DOI: https://doi.org/10.5281/zenodo.23102402 (Diaz-Uriarte *et al.*, 2026), and is also available from https://github.com/rdiaz02/scm-interv-evams.

## 8. Bibliography


Aga, O. N. L., Brun, M., Dauda, K. A., Diaz-Uriarte, R., Giannakis, K., and Johnston, I. G. 2024. HyperTraPS-CT: Inference and prediction for accumulation pathways with flexible data and model structures. *PLOS Comput. Biol.*, **20**(9), e1012393. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012393, DOI: https://doi.org/10.1371/journal.pcbi.1012393.

Angaroni, F., Chen, K., Damiani, C., Caravagna, G., Graudenzi, A., and Ramazzotti, D. 2021. PMCE: Efficient inference of expressive models of cancer evolution with high prognostic power. *Bioinformatics*, **38**(3), 754–762. DOI: https://doi.org/10.1093/bioinformatics/btab717.

Attolini, C. S.-O., Cheng, Y.-K., Beroukhim, R., Getz, G., Abdel-Wahab, O., Levine, R. L., Mellinghoff, I. K., and Michor, F. 2010. A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. *PNAS*, **107**(41), 17604–17609. URL: http://www.pnas.org/content/107/41/17604.short, DOI: https://doi.org/10.1073/pnas.1009117107.

Beerenwinkel, N., Eriksson, N., and Sturmfels, B. 2006. Evolution on distributive lattices. *J. Theor. Biol.*, **242**(2), 409–420. URL: http://www.sciencedirect.com/science/article/pii/S0022519306001159, DOI: https://doi.org/10.1016/j.jtbi.2006.03.013.

Beerenwinkel, N., Eriksson, N., and Sturmfels, B. 2007. Conjunctive Bayesian networks. *Bernoulli*, **13**(4), 893–909. URL: http://projecteuclid.org/euclid.bj/1194625594, DOI: https://doi.org/10.3150/07-BEJ6133.

Beerenwinkel, N., Schwarz, R. F., Gerstung, M., and Markowetz, F. 2015. Cancer evolution: Mathematical models and computational inference. *Syst. Biol.*, **64**(1), e1–e25. URL: http://sysbio.oxfordjournals.org/content/64/1/e1, DOI: https://doi.org/10.1093/sysbio/syu081.

Bongers, S., Forré, P., Peters, J., and Mooij, J. M. 2021. Foundations of structural causal models with cycles and latent variables. *Ann. Stat.*, **49**(5). URL: https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-5/Foundations-of-structural-causal-models-with-cycles-and-latent-variables/10.1214/21-AOS2064.full, DOI: https://doi.org/10.1214/21-AOS2064.

Chakraborty, B. and Moodie, E. E. M. 2013. *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. Springer, 2013 edition.

Charlesworth, B. and Charlesworth, D. 2010. *Elements of Evolutionary Genetics*. Roberts and Co. Publishers, Greenwood Village, Colo. URL: http://books.google.com/books?id=dgNFAQAAIAAJ.

Cheng, Y.-K., Beroukhim, R., Levine, R. L., Mellinghoff, I. K., Holland, E. C., and Michor, F. 2012. A mathematical methodology for determining the temporal order of pathway alterations arising during gliomagenesis. *PLoS Comput. Biol.*, **8**(1), e1002337. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3252265&tool=pmcentrez&rendertype=abstract, DOI: https://doi.org/10.1371/journal.pcbi.1002337.

Correa, J. and Bareinboim, E. 2020. A Calculus for Stochastic Interventions: Causal Effect Identification and Surrogate Experiments. *Proc. AAAI Conf. Artif. Intell.*, **34**(06), 10093–10100. URL: https://ojs.aaai.org/index.php/AAAI/article/view/6567, DOI: https://doi.org/10.1609/aaai.v34i06.6567.

Cristea, S., Kuipers, J., and Beerenwinkel, N. 2017. pathTiMEx: Joint Inference of Mutually Exclusive Cancer Pathways and Their Progression Dynamics. *J. Comput. Biol.*, **24**(6). URL: http://online.liebertpub.com/doi/abs/10.1089/cmb.2016.0171, DOI: https://doi.org/10.1089/cmb.2016.0171.

D'Antonio, L., Fieni, C., Ciummo, S. L., Vespa, S., Lotti, L., Sorrentino, C., and Carlo, E. D. 2023. Inactivation of interleukin-30 in colon cancer stem cells via CRISPR/Cas9 genome editing inhibits their oncogenicity and improves host survival. *J. Immunother. Cancer*, **11**(3), e006056. URL: https://jitc.bmj.com/content/11/3/e006056, DOI: https://doi.org/10.1136/jitc-2022-006056.

Desper, R., Jiang, F., Kallioniemi, O. P., Moch, H., Papadimitriou, C. H., and Schäffer, A. A. 1999. Inferring tree models for oncogenesis from comparative genome hybridization data. *J Comput Biol*, **6**(1), 37–51. URL: http://view.ncbi.nlm.nih.gov/pubmed/10223663.

Diaz-Uriarte, R. 2017. OncoSimulR: Genetic simulation with arbitrary epistasis and mutator genes in asexual populations. *Bioinformatics*, **33**(12), 1898–1899. URL: https://academic.oup.com/bioinformatics/article/33/12/1898/2982052/OncoSimulR-genetic-simulation-with-arbitrary, DOI: https://doi.org/10.1093/bioinformatics/btx077.

Diaz-Uriarte, R. 2018. Cancer progression models and fitness landscapes: A many-to-many relationship. *Bioinformatics*, **34**(5), 836–844. URL: https://academic.oup.com/bioinformatics/article/34/5/836/4557185, DOI: https://doi.org/10.1093/bioinformatics/btx663.

Diaz-Uriarte, R. and Herrera-Nieto, P. 2022. EvAM-Tools: Tools for evolutionary accumulation and cancer progression models. *Bioinformatics*, **38**(24), 5457–5459. DOI: https://doi.org/10.1093/bioinformatics/btac710.

Diaz-Uriarte, R. and Johnston, I. G. 2025. A Picture Guide to Cancer Progression and Evolutionary Accumulation Models: Systematic Critique, Plausible Interpretations, and Alternative Uses. *IEEE Access*, **13**, 62306–62340. URL: https://ieeexplore.ieee.org/document/10950363/, DOI: https://doi.org/10.1109/ACCESS.2025.3558392.

Diaz-Uriarte, R. and Vasallo, C. 2019. Every which way? On predicting tumor evolution using cancer progression models. *PLOS Comput. Biol.*, **15**(8), e1007246. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007246, DOI: https://doi.org/10.1371/journal.pcbi.1007246.

Diaz-Uriarte, R., Ríos Arroyo, Í., and Johnston, I. G. 2026. Code for "A structural causal framework for interventions on evolutionary accumulation models". Zenodo. URL: https://zenodo.org/records/23102402, DOI: https://doi.org/10.5281/zenodo.23102402.

Eberhardt, F. and Scheines, R. 2007. Interventions and Causal Inference. *Philos. Sci.*, **74**(5), 981–995. URL: https://www.cambridge.org/core/product/identifier/S0031824800005651/type/journal_article, DOI: https://doi.org/10.1086/525638.

Edelaar, P., Otsuka, J., and Luque, V. J. 2023. A generalised approach to the study and understanding of adaptive evolution. *Biol. Rev.*, **98**(1), 352–375. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12910, DOI: https://doi.org/10.1111/brv.12910.

Fontana, D., Crespiatico, I., Crippa, V., Malighetti, F., Villa, M., Angaroni, F., De Sano, L., Aroldi, A., Antoniotti, M., Caravagna, G., Piazza, R., Graudenzi, A., Mologni, L., and Ramazzotti, D. 2023. Evolutionary signatures of human cancers revealed via genomic analysis of over 35,000 patients. *Nat. Commun.*, **14**(1), 5982. URL: https://www.nature.com/articles/s41467-023-41670-3, DOI: https://doi.org/10.1038/s41467-023-41670-3.

Fontaneda, D. and Diaz-Uriarte, R. 2023. Adaptive therapy in cancer: The role of restrictions in the accumulation of mutations. URL: https://www.biorxiv.org/content/10.1101/2023.05.18.541330, DOI: https://doi.org/10.1101/2023.05.18.541330.

Frank, S. A. 1997. The Price Equation, Fisher's Fundamental Theorem, Kin Selection, and Causal Analysis. *Evolution*, **51**(6), 1712–1729. DOI: https://doi.org/10.2307/2410995.

Gerstung, M., Baudis, M., Moch, H., and Beerenwinkel, N. 2009. Quantifying cancer progression with conjunctive Bayesian networks. *Bioinformatics*, **25**(21), 2809–2815. URL: http://dx.doi.org/10.1093/bioinformatics/btp505, DOI: https://doi.org/10.1093/bioinformatics/btp505.

Gerstung, M., Eriksson, N., Lin, J., Vogelstein, B., and Beerenwinkel, N. 2011. The Temporal Order of Genetic and Pathway Alterations in Tumorigenesis. *PLoS ONE*, **6**(11), e27136. URL: http://dx.plos.org/10.1371/journal.pone.0027136, DOI: https://doi.org/10.1371/journal.pone.0027136.

Gillespie, J. H. 1983. A simple stochastic gene substitution model. *Theor. Popul. Biol.*, **23**(2), 202–215. DOI: https://doi.org/10.1016/0040-5809(83)90014-x.

Gillespie, J. H. 1984. Molecular Evolution Over the Mutational Landscape. *Evolution*, **38**(5), 1116–1129. DOI: https://doi.org/10.2307/2408444.

Gillespie, J. H. 1994. *The Causes of Molecular Evolution*. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford, New York.

Greenbury, S. F., Barahona, M., and Johnston, I. G. 2020. HyperTraPS: Inferring Probabilistic Patterns of Trait Acquisition in Evolutionary and Disease Progression Pathways. *Cell Syst.*, **10**(1), 39–51.e10. URL: http://www.sciencedirect.com/science/article/pii/S2405471219303850, DOI: https://doi.org/10.1016/j.cels.2019.10.009.

Hainke, K., Rahnenführer, J., and Fried, R. 2012. Cumulative disease progression models for cross-sectional data: A review and comparison. *Biom. J. Biom. Z.*, **54**(5), 617–40. URL: https://pubmed.ncbi.nlm.nih.gov/22886685/, DOI: https://doi.org/10.1002/bimj.201100186.

Hamilton, M. B. 2009. *Population Genetics*. Wiley-Blackwell, West Sussex. URL: http://www.blackwellpublishing.com/hamiltongenetics/.

Hanahan, D. and Weinberg, R. A. 2011. Hallmarks of Cancer: The Next Generation. *Cell*, **144**(5), 646–674. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867411001279, DOI: https://doi.org/10.1016/j.cell.2011.02.013.

Hansen, E. and Read, A. F. 2020. Modifying Adaptive Therapy to Enhance Competitive Suppression. *Cancers*, **12**(12), 3556. URL: https://www.mdpi.com/2072-6694/12/12/3556, DOI: https://doi.org/10.3390/cancers12123556.

Hernán, M. A. and Robins, J. M. 2020. *Causal Inference: What If*. CRC Press. URL: https://miguelhernan.org/whatifbook.

Holland, P. W. 1986. Statistics and Causal Inference. *J. Am. Stat. Assoc.*, **81**(396), 945–960. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354, DOI: https://doi.org/10.1080/01621459.1986.10478354.

Hosseini, S.-R., Diaz-Uriarte, R., Markowetz, F., and Beerenwinkel, N. 2019. Estimating the predictability of cancer evolution. *Bioinformatics*, **35**(14), i389–i397. URL: https://academic.oup.com/bioinformatics/article/35/14/i389/5529151, DOI: https://doi.org/10.1093/bioinformatics/btz332.

Johnston, I. G. 2026. Extracting useful information about reversible evolutionary processes from irreversible evolutionary accumulation models. URL: http://arxiv.org/abs/2601.13010, arXiv:2601.13010, DOI: https://doi.org/10.48550/arXiv.2601.13010.

Johnston, I. G. and Diaz-Uriarte, R. 2025. A hypercubic Mk model framework for capturing reversibility in disease, cancer, and evolutionary accumulation modelling. *Bioinformatics*, **41**(1), btae737. DOI: https://doi.org/10.1093/bioinformatics/btae737.

Johnston, I. G. and Williams, B. P. 2016. Evolutionary Inference across Eukaryotes Identifies Specific Pressures Favoring Mitochondrial Gene Retention. *Cell Syst.*, **2**(2), 101–111. URL: https://www.sciencedirect.com/science/article/pii/S2405471216300291, DOI: https://doi.org/10.1016/j.cels.2016.01.013.

Johnston, I. G., Hoffmann, T., Greenbury, S. F., Cominetti, O., Jallow, M., Kwiatkowski, D., Barahona, M., Jones, N. S., and Casals-Pascual, C. 2019. Precision identification of high-risk phenotypes and progression pathways in severe malaria without requiring longitudinal data. *Npj Digit. Med.*, **2**(1), 1–9. URL: https://www.nature.com/articles/s41746-019-0140-y, DOI: https://doi.org/10.1038/s41746-019-0140-y.

Kline, R. B. 2015. *Principles and Practice of Structural Equation Modeling, Fourth Edition*. The Guilford Press, New York.

Korb, K. B. and Nicholson, A. E. 2011. *Bayesian Artificial Intelligence*. CRC Press, Boca Raton, FL.

Korb, K. B., Hope, L. R., Nicholson, A. E., and Axnick, K. 2004. Varieties of Causal Intervention. In C. Zhang, H. W. Guesgen, and W.-K. Yeap, editors, *PRICAI 2004 Trends Artif. Intell.*, pages 322–331, Berlin, Heidelberg. Springer. DOI: https://doi.org/10.1007/978-3-540-28633-2_35.

Krug, J. 2021. Accessibility percolation in random fitness landscapes. In E. Baake and A. Wakolbinger, editors, *Probabilistic Structures in Evolution*, pages 1–22. EMS Press. URL: https://ems.press/books/ecr/186/3493, DOI: https://doi.org/10.4171/ECR/17-1/1.

Luo, J., Solimini, N. L., and Elledge, S. J. 2009. Principles of cancer therapy: Oncogene and non-oncogene addiction. *Cell*, **136**(5), 823–37. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2894612&tool=pmcentrez&rendertype=abstract, DOI: https://doi.org/10.1016/j.cell.2009.02.024.

Luo, X. G., Kuipers, J., and Beerenwinkel, N. 2023. Joint inference of exclusivity patterns and recurrent trajectories from tumor mutation trees. *Nat. Commun.*, **14**(1), 3676. DOI: https://doi.org/10.1038/s41467-023-39400-w.

Moen, M. T. and Johnston, I. G. 2023. HyperHMM: Efficient inference of evolutionary and progressive dynamics on hypercubic transition graphs. *Bioinformatics*, **39**(1), btac803. DOI: https://doi.org/10.1093/bioinformatics/btac803.

Montazeri, H., Günthard, H. F., Yang, W.-L., Kouyos, R., Beerenwinkel, N., and The Swiss HIV Cohort Study 2015. Estimating the dynamics and dependencies of accumulating mutations with applications to HIV drug resistance. *Biostatistics*, **16**(4), 713–726. URL: http://biostatistics.oxfordjournals.org/content/16/4/713, DOI: https://doi.org/10.1093/biostatistics/kxv019.

Montazeri, H., Kuipers, J., Kouyos, R., Böni, J., Yerly, S., Klimkait, T., Aubert, V., Günthard, H. F., Beerenwinkel, N., and The Swiss HIV Cohort Study 2016. Large-scale inference of conjunctive Bayesian networks. *Bioinformatics*, **32**(17), i727–i735. URL: http://bioinformatics.oxfordjournals.org/content/32/17/i727, DOI: https://doi.org/10.1093/bioinformatics/btw459.

Morgan, S. L. and Winship, C. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research, 2nd Ed*. Cambridge University Press, New York, NY. URL: http://dx.doi.org/10.1017/CBO9781107587991.

Nichol, D., Jeavons, P., Fletcher, A. G., Bonomo, R. A., Maini, P. K., Paul, J. L., Gatenby, R. A., Anderson, A. R. A., and Scott, J. G. 2015. Steering Evolution with Sequential Therapy to Prevent the Emergence of Bacterial Antibiotic Resistance. *PLOS Comput. Biol.*, **11**(9), e1004493. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004493, DOI: https://doi.org/10.1371/journal.pcbi.1004493.

Nicol, P. B., Coombes, K. R., Deaver, C., Chkrebtii, O., Paul, S., Toland, A. E., and Asiaee, A. 2021. Oncogenetic network estimation with disjunctive Bayesian networks. *Comput. Syst. Oncol.*, **1**(2), e1027. URL: https://onlinelibrary.wiley.com/doi/10.1002/cso2.1027, DOI: https://doi.org/10.1002/cso2.1027.

Norris, J. R. 1997. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. URL: https://www.cambridge.org/core/books/markov-chains/A3F966B10633A32C8F06F37158031739, DOI: https://doi.org/10.1017/CBO9780511810633.

O'Neil, N. J., Bailey, M. L., and Hieter, P. 2017. Synthetic lethality and cancer. *Nat. Rev. Genet.*, **18**(10), 613–623. URL: http://www.nature.com/nrg/journal/v18/n10/abs/nrg.2017.47.html, DOI: https://doi.org/10.1038/nrg.2017.47.

Orr, H. A. 2002. The population genetics of adaptation: The adaptation of DNA sequences. *Evolution*, **56**(7), 1317–1330. URL: http://www.bioone.org/doi/abs/10.1554/0014-3820%282002%29056%5B1317%3ATPGOAT%5D2.0.CO%3B2, DOI: https://doi.org/10.1554/0014-3820(2002)056[1317:TPGOAT]2.0.CO;2.

Otsuka, J. 2016. Causal Foundations of Evolutionary Genetics. *Br. J. Philos. Sci.*, **67**(1), 247–269. URL: http://bjps.oxfordjournals.org/content/67/1/247, DOI: https://doi.org/10.1093/bjps/axu039.

Otsuka, J. 2019. *The Role of Mathematics in Evolutionary Theory*. Elements in the Philosophy of Biology. Cambridge University Press, Cambridge.

Otto, S. P. and Day, T. 2007. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press, Princeton.

Ozawa, T., Riester, M., Cheng, Y.-K., Huse, J. T., Squatrito, M., Helmy, K., Charles, N., Michor, F., and Holland, E. C. 2014. Most Human Non-GCIMP Glioblastoma Subtypes Evolve from a Common Proneural-like Precursor Glioma. *Cancer Cell*, **26**(2), 288–300. URL: http://www.sciencedirect.com/science/article/pii/S1535610814002657, DOI: https://doi.org/10.1016/j.ccr.2014.06.005.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference, 2nd Ed*. Cambridge Univ. Press, New York, NY.

Pearl, J., Glymour, M., and Jewell, N. P. 2016. *Causal Inference in Statistics: A Primer*. John Wiley & Sons, Chichester, UK.

Peters, J., Janzing, D., and Schölkopf, B. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, Cambridge, MA, USA. URL: https://mitpress.mit.edu/books/elements-causal-inference.

Posada-Céspedes, S., Zyl, G. V., Montazeri, H., Kuipers, J., Rhee, S.-Y., Kouyos, R., Günthard, H. F., and Beerenwinkel, N. 2021. Comparing mutational pathways to lopinavir resistance in HIV-1 subtypes B versus C. *PLOS Comput. Biol.*, **17**(9), e1008363. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008363, DOI: https://doi.org/10.1371/journal.pcbi.1008363.

Privault, N. 2018. *Understanding Markov Chains, 2nd Ed*. Springer Undergraduate Mathematics Series. Springer Singapore, Singapore. URL: http://link.springer.com/10.1007/978-981-4451-51-2, DOI: https://doi.org/10.1007/978-981-4451-51-2.

Ramazzotti, D., Caravagna, G., Olde Loohuis, L., Graudenzi, A., Korsunsky, I., Mauri, G., Antoniotti, M., and Mishra, B. 2015. CAPRI: Efficient inference of cancer progression models from cross-sectional data. *Bioinforma. Oxf. Engl.*, **31**(18), 3016–3026. URL: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv296, DOI: https://doi.org/10.1093/bioinformatics/btv296.

Schill, R., Solbrig, S., Wettig, T., and Spang, R. 2020. Modelling cancer progression using Mutual Hazard Networks. *Bioinformatics*, **36**(1), 241–249. URL: https://academic.oup.com/bioinformatics/article/36/1/241/5524604, DOI: https://doi.org/10.1093/bioinformatics/btz513.

Schill, R., Klever, M., Lösch, A., Hu, Y. L., Vocht, S., Rupp, K., Grasedyck, L., Spang, R., and Beerenwinkel, N. 2024. Correcting for Observation Bias in Cancer Progression Modeling. *J. Comput. Biol. J. Comput. Mol. Cell Biol.*, **31**(10), 927–945. DOI: https://doi.org/10.1089/cmb.2024.0666.

Shipley, B. 2016. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R*. Cambridge University Press, Cambridge, 2 edition edition. DOI: https://doi.org/10.1017/CBO9781139979573.

Spedicato, G. A. 2017. Discrete Time Markov Chains with R. *R J.*, **9**(2), 84–104. URL: https://doi.org/10.32614/RJ-2017-036/, DOI: https://doi.org/10.32614/RJ-2017-036.

Szabo, A. and Boucher, K. 2002. Estimating an oncogenetic tree when false negatives and positives are present. *Math. Biosci.*, **176**(2), 219–236. URL: http://dx.doi.org/10.1016/S0025-5564(02)00086-X, DOI: https://doi.org/10.1016/S0025-5564(02)00086-X.

Szabo, A. and Boucher, K. M. 2008. Oncogenetic trees. In W.-Y. Tan and L. Hanin, editors, *Handbook of Cancer Models with Applications*, pages 1–24. World Scientific. URL: https://www.worldscientific.com/doi/abs/10.1142/9789812779489_0001, DOI: https://doi.org/10.1142/9789812779489_0001.

Tsiatis, A. A., Davidian, M., Holloway, S. T., and Laber, E. B. 2020. *Dynamic Treatment Regimes Statistical Methods for Precision Medicine*. CRC Press, Boca Raton, FL.

Weinreich, D. M. and Chao, L. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution*, **59**(6), 1175–1182. DOI: https://doi.org/10.1111/j.0014-3820.2005.tb01769.x.

Weinreich, D. M., Delaney, N. F., DePristo, M. A., and Hartl, D. L. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, **312**(5770), 111–114. URL: http://science.sciencemag.org/content/312/5770/111, DOI: https://doi.org/10.1126/science.1123539.

Weissman, D. B., Desai, M. M., Fisher, D. S., and Feldman, M. W. 2009. The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.*, **75**(4), 286–300. URL: http://dx.doi.org/10.1016/j.tpb.2009.02.006, DOI: https://doi.org/10.1016/j.tpb.2009.02.006.

West, J., Adler, F., Gallaher, J., Strobl, M., Brady-Nicholls, R., Brown, J., Roberson-Tessi, M., Kim, E., Noble, R., Viossat, Y., Basanta, D., and Anderson, A. R. 2023. A survey of open questions in adaptive therapy: Bridging mathematics and clinical translation. *eLife*, **12**, e84263. DOI: https://doi.org/10.7554/eLife.84263.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford; New York. DOI: https://doi.org/10.1093/0195155270.001.0001.

Wortel, M. T., Agashe, D., Bailey, S. F., Bank, C., Bisschop, K., Blankers, T., Cairns, J., Colizzi, E. S., Cusseddu, D., Desai, M. M., van Dijk, B., Egas, M., Ellers, J., Groot, A. T., Heckel, D. G., Johnson, M. L., Kraaijeveld, K., Krug, J., Laan, L., Lässig, M., Lind, P. A., Meijer, J., Noble, L. M., Okasha, S., Rainey, P. B., Rozen, D. E., Shitut, S., Tans, S. J., Tenaillon, O., Teotónio, H., de Visser, J. A. G. M., Visser, M. E., Vroomans, R. M. A., Werner, G. D. A., Wertheim, B., and Pennings, P. S. 2023. Towards evolutionary predictions: Current promises and challenges. *Evol. Appl.*, **16**(1), 3–21. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.13513, DOI: https://doi.org/10.1111/eva.13513.

Zhan, T., Rindtorff, N., Betge, J., Ebert, M. P., and Boutros, M. 2019. CRISPR/Cas9 for cancer research and therapy. *Semin. Cancer Biol.*, **55**, 106–119. DOI: https://doi.org/10.1016/j.semcancer.2018.04.001.

# Appendices

## A. Transition rate matrix from fitness landscapes under SSWM and fitness landscapes from CBN and H-ESBCN

Here we detail how we obtain fitness landscapes from CBN and H-ESBCN such that, under strong selection weak mutation (SSWM), the transition rate matrix from the fitness landscape is identical to the one from the CBN/H-ESBCN models. This transition rate matrix is the $Q$ of section 3.2, shown in Figs. 2(d) and 3(e).

### A.1. Selection coefficients: definition and notation

As is customary (e.g., Gillespie, 1984; Orr, 2002; Weinreich *et al.*, 2006), we use $s$ for the selection coefficient. Let $s_{x\to x+i}$ be the selection coefficient for the mutation of gene $i$ on a genotype $x$ without gene $i$ mutated. Thus, the fitness of a genotype $x$ that gains mutation $i$, $W_{x+i}$, is $(1+s_{x\to x+i})W_x$. In other words, $s_{x\to x+i}=(W_{x+i}-W_x)/W_x$, where $W_{x+i}$ is the fitness of genotype $x$ with the addition of gene $i$ mutated.

We use the notation $s_{x\to x+i}$ to make it explicit that the fitness effect of gaining mutation $i$ can depend on background (i.e., there might not be a single, unique $s_i$). For EvAMs that specify deterministic restrictions using DAGs, we can simplify the notation and write just $s_i$ for the selection coefficient for mutation $i$ when the restrictions for the acquisition of mutation $i$ are satisfied. When the restrictions are not satisfied, any genotype with mutation $i$ has fitness zero, so we need not consider this transition; if restrictions are satisfied, the selection coefficient is the same regardless of genotype background. In some descriptions below, we will use $s_i$ when it is clear, from the context, whether we mean $s_i$ as in EvAMs that use DAGs, or the more general $s_{x\to x+i}$ that allows the effect of a mutation $i$ to depend on the genotype.

### A.2. From fitness landscape to transition rate matrix

We assume SSWM (Gillespie, 1984; Orr, 2002; see also Gillespie, 1994, 1983; Weinreich *et al.*, 2006; Nichol *et al.*, 2015; Krug, 2021): $Ns_i \gg 1$ and $N\mu \ll 1$, where $\mu$ is mutation rate and $N$ is population size. The $s_i$ are sufficiently small that their probability of fixation is approximately $2s_i$ (Gillespie, 1994, p. 232; Orr, 2002, p. 1319). Then, in a haploid population the time to fixation is exponentially distributed with mean $1/(2N\mu s_i)$, so the transition rate is $2N\mu s_i$ (we can use results from haploids, because we are dealing with clonal, or asexual, reproduction —e.g., Charlesworth and Charlesworth, 2010, p. 50; Hamilton, 2009, pp. 185–189; Otto and Day, 2007, section 3.3).

In addition, we make the stronger assumption that evolution only proceeds uphill in the fitness landscape, i.e., it only evolves by fixing beneficial mutations (Krug, 2021, section 6.2). Thus, we exclude tunneling and escaping from local fitness maxima (Weinreich and Chao, 2005; Weinreich *et al.*, 2006; Weissman *et al.*, 2009), which can occur even under SSWM. Therefore, under our model, the probability of transition to a genotype of smaller fitness is 0 (see also, for a similar approach, Nichol *et al.*, 2015, eq. 3). We make this simplifying assumption so that we do not need to consider population size and mutation rates (which affect escape from local fitness maxima and tunneling — Weinreich and Chao, 2005; Weissman *et al.*, 2009) nor the role of bulk sequencing (Diaz-Uriarte and Johnston, 2025).

### A.2.1. Scaling the transition rate matrix

We can scale the transition rate so that the process is faster or slower, or expressed in different units of time (for example, Gillespie 1984, shows expressions for the scaling of the transition rates so that time is in units of $N$ generations). We refer to this as the "scaled transition rate matrix".

## A.3. CBN and H-ESBCN: From DAGs of restrictions to fitness landscapes; equivalence of transition rate matrices

We can generate a fitness landscape from CBN and H-ESBCN such that the fitness landscape is not only representable (Diaz-Uriarte, 2018; Diaz-Uriarte and Johnston, 2025), but also yields a transition rate matrix identical (up to a scaling constant) to that of the original CBN/H-ESBCN model. When the transition rate matrices match this way, then in the limit of very large samples from repeated evolution experiments under SSWM on this fitness landscape, fitting the corresponding CBN/H-ESBCN, and assuming the correct parameters are inferred, the transition rates between genotypes from the parameters of the CBN/H-ESBCN model would match the transition rates between genotypes from the fitness landscape. Furthermore, obtaining identical transition rate matrices from both the CBN/H-ESBCN model and its implied fitness landscape means that interventions performed by directly manipulating the fitness landscape or the CBN/H-ESBCN model should yield identical results to interventions via $DAG_{\lambda_g=0}$, $DAG_{-g}$, and $Q_{-g}$ (section 3.4.1), as confirmed by our tests (file `kill-gene-equivalences-TESTS.R`).

For CBN and H-ESBCN, the transition rate $x \to x+i$ is equal to $\lambda_i$ when the restrictions for the acquisition of $i$ are satisfied. And, under our assumed evolutionary regime, that transition rate is $\propto s_i$ (section A.2), where $s_i (= s_{x \to x+i}) = (W_{x+i} - W_x)/W_x$ (section A.1). Thus, starting from the WT genotype (e.g., with fitness 1), we can obtain the fitness of all accessible genotypes (those that can exist under the model) by setting $s_i = a\lambda_i$, and iteratively computing $W_{x+i} = W_x(1+s_i)$.

This will produce a fitness landscape so that the transition rate matrices from CBN and H-ESBCN and the fitness landscape are identical (up to a scaling constant). If we then scale the fitness-landscape derived transition rate matrix (section A.2.1) by multiplying each transition rate by $1/a$, the transition rate matrix that we obtain directly from the EvAM models (for CBN and H-ESBCN) and the scaled transition rate matrix from the fitness landscape, will be identical (and the individual entries equal to $\lambda_i$): Fig. 5.

What is an appropriate value of $a$ depends on the distribution of the $\lambda_i$s. For example, in MC-CBN (https://github.com/cbg-ethz/MC-CBN) and EvAM-Tools (Diaz-Uriarte and Herrera-Nieto, 2022), randomly generated CBN models have, by default, $\lambda \sim \mathcal{U}(1/3, 3)$ so the average $\lambda$ will be close to 1.667. Thus, a value of $a = 0.006 (\approx 0.01/1.667)$ will generally be adequate to ensure that the $s_i = a\lambda_i$ are small enough ($\approx 0.01$) for SSWM so that the entries of the transition rate matrix are proportional to the $s_i$.

This correspondence between EvAM models and fitness landscapes is only available for CBN and H-ESBCN. For EvAMs with stochastic dependencies (MHN, HyperHMM, HyperTraPS), there is no clear mapping between these models and fitness landscapes (Diaz-Uriarte and Johnston, 2025). For OT and OncoBN, it is not possible to directly map an OT or OncoBN model to a fitness landscape as OT and OncoBN are untimed models; however, the deterministic dependencies that OT and OncoBN encode are a subset of those captured by

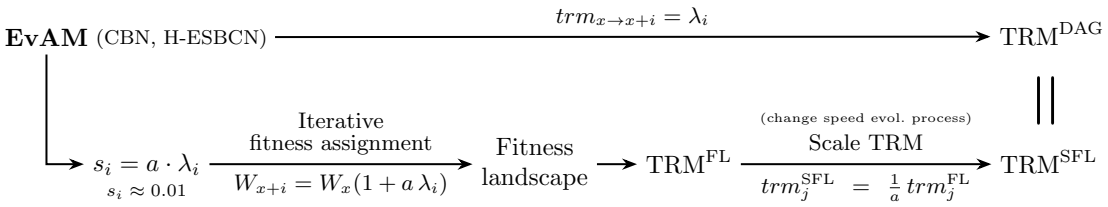



Figure 5: **DAGs of restrictions to fitness landscapes: equivalence of transition rate matrices: $\mathrm{TRM}^{\mathrm{DAG}} = \mathrm{TRM}^{\mathrm{SFL}}$.**
$\mathrm{TRM}^{\mathrm{DAG}}$: transition rate matrix from the DAG of restrictions for CBN and H-ESBCN (A.3); this is $Q$ (section 3.2), shown in Figs. 2(d) and 3(e).
$\mathrm{TRM}^{\mathrm{FL}}$: transition rate matrix from the fitness landscape (A.2).
$\mathrm{TRM}^{\mathrm{SFL}}$: scaled transition rate matrix (A.2.1).
$trm_j^{\mathrm{SFL}}, trm_j^{\mathrm{FL}}$: individual entries of $\mathrm{TRM}^{\mathrm{SFL}}$ and $\mathrm{TRM}^{\mathrm{FL}}$, respectively.
$s_i, \lambda_i, W_x, W_{x+i}$: sections A.1 and A.3.
$a = 0.006$ (section A.3 for rationale).

CBN and H-ESBCN, so these types of dependencies are mapped to fitness landscapes via CBN/H-ESBCN.

## B. H-ESBCN with XOR: why it induces a cycle

XOR forbids any genotype that contains the child and its parents: with XOR in Fig. 3(a.2), genotype $\{A,B,C\}$ (or any genotype containing genes $A,B,C$ all mutated) is not possible. Gaining a mutation in gene $A$ is possible *only* when exactly one of the parents is already mutated: not with none, and not once $B$ and $C$ are both present. These cases allow following the graph down, from parent to descendant, to decide what mutations can occur (what events can be gained). But, and here lies the issue, neither can the second parent be gained once $A$ (and, therefore, one of the parents) is present —the child node has an effect back on its parents: an implicit arrow from $A$ to $B$ and from $A$ to $C$. This precludes writing the equations in terms of just events and their parents in the graph.

If the waiting times of $B$ and $C$ depended only on their own parents (here, Root), both would eventually occur with probability 1 (as the time to their occurrence would be exponentially distributed with rates $\lambda_B$ and $\lambda_C$, as specified by H-ESBCN for events whose only parent is Root). But then, any $\lambda_A > 0$ would give positive probability to $\{A,B,C\}$, which is forbidden. Therefore, enforcing the XOR exclusion at the level of events requires the waiting time of the second parent to depend on whether the child (here, $A$) is present, but that means the mechanisms of parents and child cannot be autonomous. Put differently, in contrast to AND and OR, we cannot write the structural equations moving down the graph; we cannot write the equation for a gene solely as a function of its parents, because, with an XOR, the occurrence of a parent can be forbidden by the acquisition of a mutation in the child. The XOR amounts to a cycle: $A$ is connected back to its parents.

In fact, the XOR relationship in Fig. 3(a.2) is a limit case of HyperTraPS with three-way ($L^3$) interactions. It is equivalent to, in HyperTraPS, the combination of: (a) each of $B$ and $C$ having an extreme positive effect in the acquisition of $A$, extreme because without them $A$ cannot be gained (and they have to cancel $A$'s baseline rate, which is effectively zero); (b) $A$

having an extreme negative effect on each one of $B$ and $C$ (preventing their occurrence); (c) $\{B,C\}$ together having an extreme negative effect on $A$ —larger than the combined individual positive effect of each of $B$ and $C$ on $A$. (MHN cannot represent this scenario, even in the limit, because it can only model pairwise, not higher-order interactions.)

As explained in the main text (section 3.4.1), we address the cycles induced by XORs by expressing the structural equation as a function of the current genotype, not only the ancestors in the graph. Equivalently, the model can be stated over genotypes, Fig. 3(e): the next genotype is a function of the current genotype and of independent exponential waiting times, one per possible transition. Since mutations are only gained, this graph over genotypes is acyclic by construction; this representation applies to AND, OR, XOR, as well as MHN. It is also the representation we use for HyperHMM and HyperTraPS (section 3.4.4), with steps in place of continuous time and the per-step transition probabilities in place of the rates: the next genotype is a function of the current genotype and a single exogenous term per step, $X_{v+1} = f(X_v, \varepsilon_{v+1})$.

## C. Assessing intervention objectives: a protocol

Here we provide details on how to assess whether EvAMs can identify promising genes with respect to the three objectives considered in section 4.

For all three measures, O_genotype, O_mut, and O_WT, the ground truth is obtained from a fitness landscape in which the targeted gene's manipulation —setting the fitness of affected genotypes to 0— has been carried out (this could possibly be done experimentally, though we focus here on the computational manipulation of setting fitness to 0). The predictions of the interventions are then obtained from an EvAM method. Finally, predictions and ground truth are compared. None of these measures require that the EvAM method have a fitness landscape interpretation: they require only that we can obtain hitting probabilities (O_genotype) or genotype frequencies (O_mut, O_WT) from the EvAM. The fitness landscape is needed only for the ground truth, not for the predictions.

### C.1. Increase/decrease the probability of going through (hitting) a given genotype (O_genotype)

We assess whether EvAMs can correctly predict which genes, when targeted, most decrease (or increase) the probability of hitting (going through) a given genotype, the "focal genotype". When considering a particular focal genotype, $z$, the set of possible target genes cannot include any of the genes mutated in $z$ since the trivially best approach to have a hitting probability of 0 would be to target any of the mutated genes in $z$.

We might want to consider, as focal genotypes, all accessible genotypes in the true fitness landscape that have a certain number of mutations. The rationale is to evaluate performance over a large number of genotypes that can capture complex dependencies and are assessed under a range of interventions. For example, in a simulation study involving 9 genes, we might want to focus on genotypes in the true fitness landscape that have between 3 and 5 mutations. This removes from consideration genotypes with only 1 or 2 mutations (too few mutations to reflect complex dependencies), and at the same time all genotypes assessed are evaluated under at least 5 different interventions, so that the correlation coefficient between truth and prediction is computed with at least 5 points (as we will probably want to use the "no intervention" as an additional point).

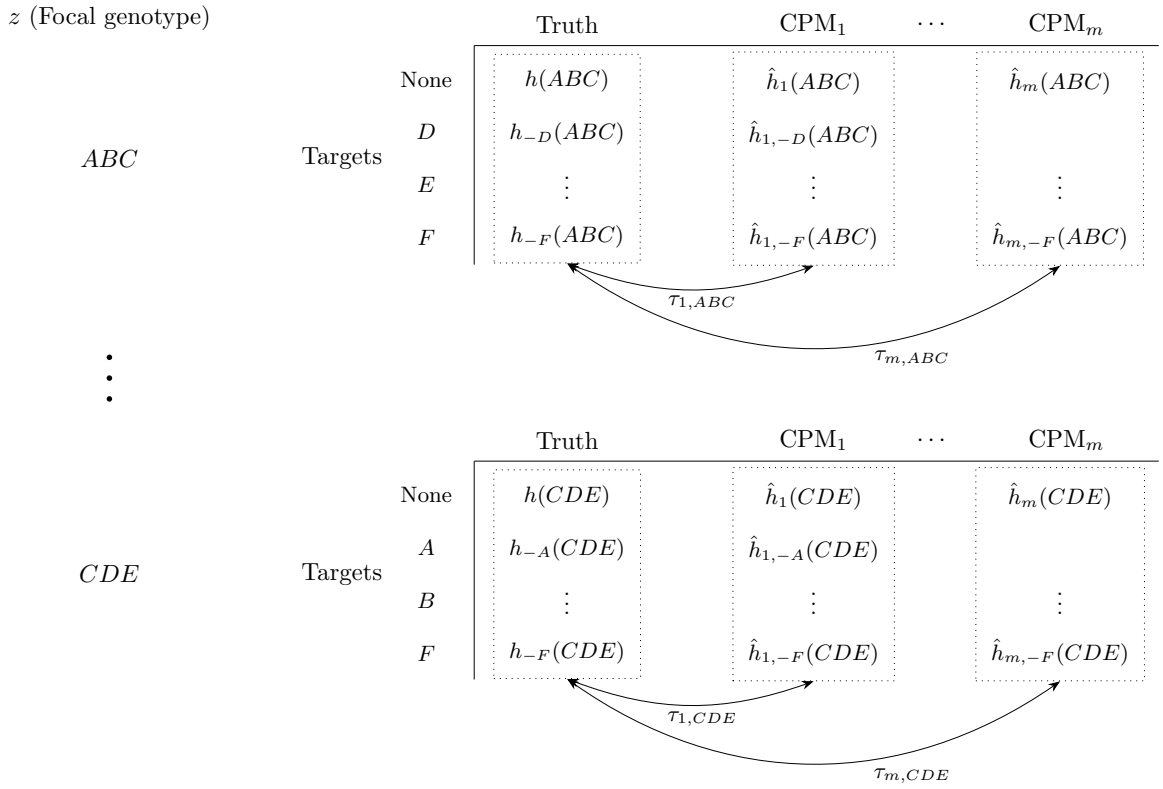



Figure 6: **Assessing O_genotype: Increase/decrease the probability of going through (hitting) a given genotype.** For each method, we can compute Kendall's $\tau$ correlation between the truth and the predictions of the hitting probability under all possible interventions for every focal genotype (the figure shows two focal genotypes, $\{A,B,C\}$ and $\{C,D,E\}$). For example, in the bottom row, $h_{-F}(CDE)$ is the true hitting probability of genotype $\{C,D,E\}$ when we intervene on gene $F$, and $\hat{h}_{1,-F}(CDE)$ the predicted hitting probability of genotype $\{C,D,E\}$ when we intervene on gene $F$ from $\text{CPM}_1$. For a given method, the value of the statistic is the average rank correlation over all the focal genotypes (so for method $\text{CPM}_1$ we would average $\tau_{1,ABC},\ldots,\tau_{1,CDE}$ ).

For a given fitness landscape we can assess this objective as:

1. Focal genotypes are all accessible genotypes in the true fitness landscape with a certain number of mutations (e.g., 3 to 5 mutations); we use $z$ to denote any one of the focal genotypes.

2. If under SSWM, from $\mathcal{Q}$, the true transition rate matrix of the fitness landscape, obtain the transition matrix corresponding to the embedded discrete-time chain; from it, obtain the hitting (or first passage) probabilities (Norris, 1997; Privault, 2018), **h**, starting from the WT state (hitting probabilities can be readily obtained using, for example, the `markovchain` R package — Spedicato, 2017). Thus, $h(z)$, the entry of **h** corresponding to a focal genotype $z$, is the hitting probability of genotype $z$ (and we make no difference between transient and absorbing genotypes, so the hitting probability of an absorbing genotype is also its absorption probability). Alternatively, if $\mathcal{Q}$ is not available for the evolutionary regime considered, use simulations to estimate **h**.

3. For every gene $g$, obtain $\mathcal{Q}_{-g}$, the transition rate matrix when intervening on gene $g$ and, from it, the transition matrix corresponding to the embedded discrete-time chain. From it, obtain the hitting probability of genotypes under the intervention, $\mathbf{h}_{-g}$. $h_{-g}(z)$ is the hitting probability of genotype $z$ when intervening on gene $g$. (If $\mathcal{Q}_{-g}$ is not available for the evolutionary regime considered, use simulations to estimate $\mathbf{h}_{-g}$.)

4. For each EvAM, $M_m$:

   4.1. From $M_m$, obtain the predicted hitting probability of genotypes under no intervention $\hat{\mathbf{h}}_m$. $\hat{h}_m(z)$ is the hitting probability, for method $m$, of genotype $z$ under no intervention.

   4.2. Modify the model results under each intervention to obtain $M_{m,-g}$ and, from it, the predicted hitting probability of genotypes $\hat{\mathbf{h}}_{m,-g}$. $\hat{h}_{m,-g}(z)$ is the predicted hitting probability, under model $m$, of genotype $z$ when intervening on gene $g$.

   4.3. For each focal genotype, $z$:

      4.3.1. For each EvAM, compute the rank correlation between the vectors $(h(z), h_{-A}(z), h_{-B}(z), \ldots)$ and $(\hat{h}_m(z), \hat{h}_{m,-A}(z), \hat{h}_{m,-B}(z), \ldots)$.

      In other words, compute the rank correlation between the true (from steps 1. and 3.) and the predicted hitting probability of $z$ (from steps 4.1. and 4.2.) under all possible interventions (i.e., those that target a gene $g$, where $g$ is not one of the genes mutated in $z$); include also the "no intervention" as a data point (it is conceivable that some intervention could be worse than doing nothing).

      Here, and elsewhere, we suggest measuring correlation using Kendall's $\tau$ (not Spearman's rank correlation), to emphasize choosing one intervention vs. another (i.e., pair agreement) rather than overall ordering agreement. (Differences in results from different correlation coefficients are, however, unlikely to be relevant.)

      Note that what we would compute is different from the rank correlation between $\mathbf{h}_{-g}$ and $\hat{\mathbf{h}}_{m,-g}$; this would emphasize the agreement between the hitting probabilities of all the genotypes for each particular $g$. This is not as useful as the measure proposed here, which directly answers "what should I target if I want to increase/decrease the hitting probability of genotype $z$".

   4.4. Return the average rank correlation coefficient over all $z$.

### C.2. Decrease overall frequency of mutations or alterations: O_mut

For a given fitness landscape we can assess this objective as:

1. From $\mathcal{Q}$, the true transition rate matrix of the fitness landscape, obtain $\mathbf{f}$, the distribution of genotypes, using the same time distribution as used when generating the samples (section 4.1). (As above, if $\mathcal{Q}$ is not available, we can instead use forward genetic simulations to estimate $\mathbf{f}$.)

   From $\mathbf{f}$, obtain the true population mean number of mutations as $D = \sum_z n(z) f(z)$, where $n(z)$ is the number of mutations of genotype $z$ and $f(z)$ is as defined above.

2. For every $g$, from $\mathcal{Q}$ (or using forward genetic simulations on the modified fitness landscapes), obtain $\mathbf{f}_{-g}$, and from it the true mean number of mutations under intervention on $g$: $D_{-g} = \sum_z n(z) f_{-g}(z)$.

3. For each EvAM, $M_m$:

   3.1. From $M_m$, obtain the predicted distribution of genotypes under no intervention $\hat{\mathbf{f}}_m$. $\hat{f}_m(z)$ is the predicted frequency, for method $m$, of genotype $z$ under no intervention. From $\hat{\mathbf{f}}_m$ obtain the predicted mean number of driver mutations, $\hat{D}_m = \sum_z n(z) \hat{f}_m(z)$.

   3.2. From $M_{m,-g}$ obtain $\hat{\mathbf{f}}_{m,-g}$ and from this, the predicted mean number of driver mutations, $\hat{D}_{m,-g} = \sum_z n(z) \hat{f}_{m,-g}(z)$.

   3.3. Compute the rank correlation between the vectors $(D, D_{-A}, D_{-B}, \ldots)$ and $(\hat{D}_m, \hat{D}_{m,-A}, \hat{D}_{m,-B}, \ldots)$.

   In other words, compute the rank correlation between the true and predicted mean number of mutations under all possible interventions; include also the "no intervention" as a data point.

As we are measuring the correlation between the true and predicted effects of changes, EvAMs that perform well here could also be used to choose interventions that increase the overall frequency of mutations.

### C.3. Increase frequency of the non-mutated, starting genotype, WT: O_WT

Very similar to C.2. For a given fitness landscape we can assess this objective as:

1. Obtain $f(WT)$, i.e., $f(z)$ when $z$ is the WT genotype.

2. For every $g$, obtain $f_{-g}(WT)$, the frequency of the WT genotype when intervening on gene $g$.

3. For each EvAM, $M_m$:

   3.1. Obtain $\hat{f}_m(WT)$ (predicted frequency of WT for model $m$).

   3.2. For every $g$, obtain $\hat{f}_{m,-g}(WT)$ (predicted frequency of WT for model $m$ under intervention $g$).

   3.3. For each EvAM, compute the rank correlation between the vectors $(f(WT), f_{-A}(WT), f_{-B}(WT), \ldots)$ and $(\hat{f}(WT), \hat{f}_{-A}(WT), \hat{f}_{-B}(WT), \ldots)$.

## D. Interventions: code and algorithmic details

File `intervention.R` is the main entry point into the intervention code. It provides additional details about the functions used and how they map onto the notation used here.

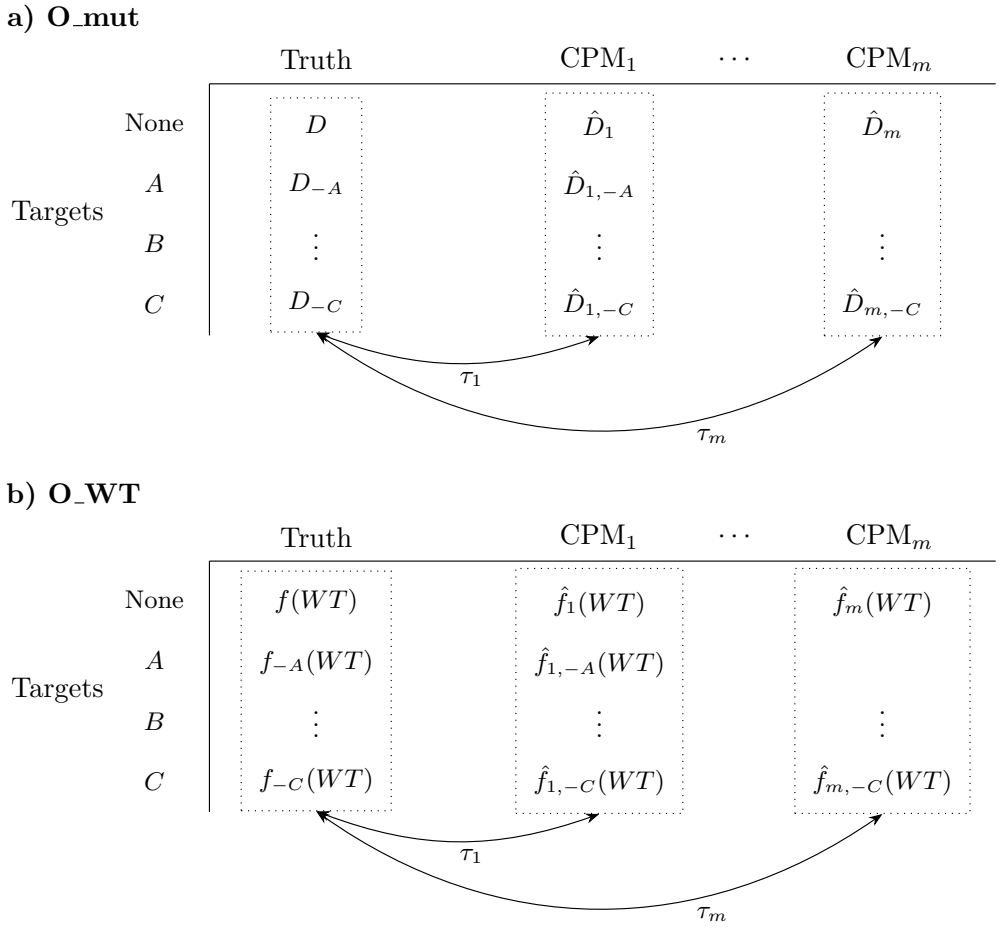



Figure 7: **Assessing O_mut (decreasing overall frequency of mutations) and O_WT (increasing the frequency of the WT genotype)**. See text for meaning of symbols. For each CPM, the value of the statistic is the rank correlation between predictions and truth.

### D.1. Interventions as modification of the DAGs: details and algorithm

OT, OncoBN, CBN, and H-ESBCN specify the EvAM model using DAGs of restrictions. Our default procedure for targeting gene $g$ involves eliminating, from the DAG that specifies the model, the node of the targeted gene and any other descendant nodes that strictly depend on the targeted gene. We denote this as $DAG_{-g}$. Specifically:

- OT: the models are trees, and thus a node depends on only one ancestor. If gene $h$ depends on targeted gene $g$, making $g$ lethal precludes $h$ and all of its descendants from accumulating. Thus, for OT, we remove from the DAG the targeted gene $g$ and all of its descendants.
- CBN: the models are DAGs, with AND conjunction. If a gene $h$ depends on the targeted $g$ and some other genes $o_1, o_2, \ldots$, making $g$ lethal means $h$ and all of its descendants must be removed from the DAG too. Thus, for CBN, we remove from the DAG the targeted gene $g$ and all of its descendants.

- For OncoBN, under the disjunctive model, a gene $h$ can depend, with an OR relationship, on $g$, the targeted gene, and other genes $o_1, o_2, \ldots$; in other words, $h$ should be removed if and only if it depends only on $g$. Thus, for OncoBN under the disjunctive model we remove from the DAG the targeted gene $g$ and all of its descendants that depend only on $g$. For OncoBN under the conjunctive model we operate as for CBN.
- For H-ESBCN, a gene $h$ can depend on $g$, the targeted gene, and other genes $o_1, o_2, \ldots$. The dependency can be conjunctive (AND) as for CBN, disjunctive (OR) as for OncoBN, and exclusive OR (XOR); the type of dependency of a gene on its immediate ancestors is the same for all ancestors. Thus, for H-ESBCN, we remove the targeted gene $g$ and its descendants as dictated by whether the dependency is AND (similar to what we do with CBN) or OR/XOR (similar to what we do with OncoBN).

In all the cases above, when we target gene $g$, a gene $h$ should be removed from the DAG if and only if $h$ appears in accessible genotypes only with gene $g$. This leads to a simple algorithm for removing genes from the DAG that does not require traversing the DAG to assess exclusive dependencies, and that is easy to implement if the DAG is stored as a collection of pairs of edges $From, To$:

1. Identify the genotypes that become non-viable after the intervention: all those that have $g$.
2. Find all genes, $h_1, h_2, \ldots$ that appear in formerly viable genotypes only with $g$.
3. Remove, from the DAG, any edges that involve $g, h_1, h_2, \ldots$ (i.e., any edges that involve $g, h_1, h_2, \ldots$ as origin or destination).

As a simple example, Fig. 3(b) shows that, under $do(A = 0)$, only the arrows into $A$ are removed: the $A \to D$ relationship and its parameter are unchanged; but $D$, which requires $A$, can no longer arise. This is why $DAG_{-g}$ can drop $D$ without loss, and that is what the algorithm above implements.

***Code:*** The above algorithm is in function `kill_gene_DAG` (called from `kill_gene`) in file `kill-gene-and-output-from-cpm.R`.

***Code:*** Function `intervene_cpm_every_gene`, in file `intervention.R`, is the main intervention function, which calls `kill_gene` and obtains predicted genotype frequencies and hitting probabilities (function `get_genotype_freqs_cpm`).

***Code:*** Function `kill_gene_DAG_param_0` called by `kill_gene_by_params_to_0` kills by setting the corresponding parameter ($\lambda_i, \theta_i, \pi_{parent(i),i}$, for CBN/H-ESBCN, OncoBN, and OT, respectively) to 0; tests showing equivalence of setting the parameter and removing from DAG in file `kill-gene-equivalences-TESTS.R`.

## D.2. MHN: Interventions as modifications of the $\Theta$ matrix and the transition rate matrix

As explained in *"MHN: graph interventions" (section 3.4.3)*, we intervene on a gene $g$, $do(g = 0)$, by setting $\Theta_{g,g} = 0$. The net effect of this is setting $Q_{x \to x+g}$, the transition rate from genotype $x$ to genotype $x$ plus a mutation in gene $g$, to 0. To prevent having non-zero transition rates from genotypes that are not accessible, we also set $\Theta_{i,g} = 0$, where $i$ are each of the remaining genes, to ensure that all transitions out of a genotype with $g$ mutated are 0. (Setting $\Theta_{g,i} = 0$ can also be done, but is unnecessary if we have set $\Theta_{g,g} = 0$).

Setting $\Theta_{g,g} = 0$ amounts to eliminating $g$ from the $\Theta$ matrix. This procedure, $\Theta_{-g}$, is what we use by default. ***Code:*** Function `kill_gene_MHN` (called from `kill_gene`) in file `kill-gene-and-output-from-cpm.R`.

***Code:*** Function `kill_gene_MHN_theta_minus_Inf` called by `kill_gene_by_params_to_0`. The code sets $\theta = -\infty$, as $\theta$, not $\Theta$, is the parameter directly used in our code.

***Code:*** Tests showing equivalence of the two procedures in file `kill-gene-equivalences-TESTS.R`.

### D.3. CBN, H-ESBCN, MHN: equivalence of setting parameters to 0 and removing entries from the transition rate matrix

Setting $\lambda_g = 0$ (CBN, H-ESBCN) or $\Theta_{gg} = 0$ (MHN) gives rate 0 to every transition into a genotype with $g$ mutated and leaves all other rates unchanged. The genotypes with $g$ mutated are therefore unreachable, and the chain on the remaining genotypes is $Q_{-g}$, as defined in section 3.2.

***Code:*** Tests showing the equivalence of: a) setting the corresponding parameter ($\Theta_{g,g}, \lambda_g$) to 0; b) removing, from the transition rate matrices, all genotypes with gene $g$ mutated (and then computing the distributions of genotypes under sampling with the time distributed as an exponential with rate 1), in file `kill-gene-equivalences-TESTS.R`.

### D.4. OT, OncoBN, CBN, H-ESBCN, MHN: transition matrix from intervening on the model versus restricting and renormalizing

For OT, OncoBN, CBN, H-ESBCN and MHN, the transition matrix after an intervention on gene $g$ can be obtained in two equivalent ways: by building it from the intervened model, or by taking the transition matrix of the model without intervention, keeping only the genotypes without $g$, and dividing each row by its sum. (Section D.3 gives the corresponding result for the transition rate matrix; OT and OncoBN have no transition rate matrix, and the matrix used to obtain hitting probabilities is the heuristic one of section D.7; for them, this section is the only place where the equivalence is shown.) The reason is that, for OT, OncoBN, CBN and H-ESBCN, the weight of the transition from genotype $x$ to genotype $x$ plus gene $i$ depends only on $i$ ($\pi_{parent(i),i}$ for OT, $\theta_i$ for OncoBN, $\lambda_i$ for CBN and H-ESBCN), and for MHN it depends on the genes already present in $x$, which never include $g$; intervening on $g$ therefore removes the genotypes that contain $g$ and leaves the weights of the remaining transitions unchanged. Recall that for CBN, H-ESBCN, and MHN the transition matrix is that of the embedded discrete-time chain, obtained from the transition rate matrix via competing exponentials: each row is the rates divided by their sum. For OT and OncoBN the matrix is built in the same way, with the conditional probabilities in place of the rates (section D.7). Thus, with the weights unchanged, restricting and renormalizing gives the same matrix as building it from the intervened model. But, and this is crucial, this row renormalization is not the rescaling of genotype frequencies shown to be wrong in section 2.3: there, the frequencies of the genotypes containing $g$ are set to 0 and the rest rescaled, which is conditioning; here, for each genotype, we redistribute, over the remaining possible transitions, the probability that the intervention has forbidden.

This equivalence does not extend to HyperHMM and HyperTraPS. Restricting and renormalizing $R$ would give the embedded chain of the intervened process, but that is not the matrix we use: as we argue in section 3.4.4, the entries of $R$ are used as per-step probabilities, with a

step being an attempt at mutation, and renormalizing would change the per-step probabilities of the remaining transitions. (Note, though, that whether or not $R$ is renormalized, the hitting probabilities are the same, since a self-transition only delays the chain; only the predicted genotype frequencies after $v$ steps differ.)

***Code:*** Tests in file `trans-mat-restrict-renormalize-TESTS.R`, for OT, OncoBN (conjunctive and disjunctive), CBN, H-ESBCN and MHN.

### D.5. HyperHMM: Interventions via creation of a self-transition in the transition probability matrix

The algorithm has been explained in full in *"HyperHMM: interventions via the conditional transition matrix" (section 3.4.4)*. In the code, right after the fitting, transition probabilities below $10^{-12}$ are set to 0, and rows renormalized to sum to 1, before computing any predictions (genotype frequencies and hitting probabilities); these small transition probabilities are numerical noise from the fitting. For computational reasons, the unreachable genotypes, those with $g$ mutated, are dropped from the matrix after killing; this leaves predictions unchanged, since those genotypes have predicted frequency and hitting probability 0.

***Code:*** Function `kill_gene_HyperHMM_drop_unreachable` in file `kill-gene-and-output-from-cpm.R`. Tests in file `tests/kill-HyperHMM-TESTS.R`.

### D.6. Interventions on EvAMs as modifications of the fitness landscape

***Code:*** Function `intervene_fitness_landscape_every_gene`, in file `intervention.R` returns the results of intervening on every gene via modification of the fitness landscape. This function calls `kill_gene_fitness_landscape` which is the one that does the actual modification of the fitness landscape for all genotypes affected by killing gene $g$.

***Code:*** Equivalence of killing via fitness landscapes modification and DAG of restrictions intervention is shown in file `kill-gene-equivalences-TESTS.R`, section "CBN and H-ESBCN: intervene by modifying the fitness landscape identical to DAG intervention".

### D.7. Genotype predictions and hitting probabilities from EvAMs and error model

For CBN, H-ESBCN, MHN we can obtain the genotype frequency predictions from the transition rate matrix. This transition rate matrix is obtained directly from the expressions of MHN (Schill *et al.*, 2020), and can be obtained from the DAG of restrictions and the values of the set of $\lambda$ parameters for CBN and H-ESBCN (Montazeri *et al.*, 2016; Angaroni *et al.*, 2021; see also Diaz-Uriarte and Johnston, 2025). From the transition rate matrix, the predicted genotype frequency is obtained as per equation (4) in Schill *et al.* (2020), assuming sampling time is distributed as an exponential of rate 1.

For OT and OncoBN, predicted genotype frequencies are a function of both the conditional probabilities of the model and deviations from the models. For OT the error model includes possibly different false positive and false negative observation errors and deviations from the model; the observation error is reflected in both the estimated false positive $\varepsilon_+$ and the estimated false negative $\varepsilon_-$ errors, but the true deviations from the model are reflected only in

$\varepsilon_+$ (see Szabo and Boucher, 2002, 2008, and Supplementary Material to Diaz-Uriarte and Herrera-Nieto, 2022). OncoBN has a different error model from OT, with a "spontaneous activation model" that allows child mutations to accumulate even if the parents in the DAG have not been mutated (Nicol *et al.*, 2021; see also Supplementary Material to Diaz-Uriarte and Herrera-Nieto, 2022, section 5.2). Both OT (Szabo and Boucher, 2002, 2008) and OncoBN (Nicol *et al.*, 2021), by default, allow for deviations from the models (both for estimation and for predicting genotype frequencies). This matters for the intervention. For example, in OncoBN, with $\varepsilon > 0$ a gene can be gained even when its parents are not mutated, so setting $\theta_g = 0$ would not prevent it, and only with $\varepsilon = 0$ is $\theta_g = 0$ the same as $do(g = 0)$. Thus, to use a similar procedure when generating predicted genotype frequencies as the one for CBN/H-ESBCN/MHN, we set the error terms ($\varepsilon_+$ —`epos`— and $\varepsilon_-$ —`eneg`— for OT and $\varepsilon$ for OncoBN) to 0 (see also discussion in sections 3.2 and 5 in the Supplementary Material to Diaz-Uriarte and Herrera-Nieto, 2022). This ensures that $do(g = 0)$ effectively prevents the gaining of dependent events.

***Code:*** In function `get_full_output`, `epos` is set to 0 by default; see file `kill-gene-and-output-from-cpm.R`. Function `get_full_output` is called from `get_genotype_freqs_cpm`.

For CBN, H-ESBCN, and MHN hitting probabilities can be obtained from the transition matrix of the embedded discrete-time chain. For HyperHMM/HyperTraPS from the transition matrix of HyperHMM/HyperTraPS. For OT and OncoBN we can use the same procedure as used in Diaz-Uriarte and Vasallo (2019) (section 3.1 in Supplementary file S4); thus, we construct a matrix where the "genotype transition rate" entries (with 0 in the diagonal) are the conditional probabilities ($\pi_{parent(i),i}$, $\theta_i$, for OT and OncoBN, respectively) and then row-scale that matrix so that entries sum to 1 across rows. This is the "transition matrix" used to obtain the hitting probabilities. This is a heuristic procedure, because OT and OncoBN are untimed models that provide no transition matrices (nor transition rate matrices). Section D.4 shows what an intervention does to this matrix.

***Code:*** Tests in files `tests/hitting-probs-direct-and-drop-unreachable-TESTS.R` and `tests/trm-TESTS.R`.

### D.8. Interventions: ground truth from fitness landscapes.

***Code:*** Function `intervene_fitness_landscape_every_gene`, from `intervention.R`, which calls `kill_gene_fitness_landscape` from file `kill-gene-and-output-from-cpm.R`