**Evolutionary fitness.** [What is Evolutionary Fitness? | Survival of the Fittest - YouTube](https://www.youtube.com/watch?v=oK0e3BcEkEE)

*An organism’s capacity to survive and reproduce in each environment.*

Fitness starts in the **genes** (genotype and phenotypes). Both determine the survival of a species. Evolutionary fitness is not necessarily the fastest or the strongest that survive, but those with the best **advantage** in that environment.

**Reproduction** is just as important as survival. Here, sexual selection comes in. Hence, fitness involves survival and reproduction. If you only have one, but not the other, genes will not likely be passed on to the next generation.

**Mutual exclusivity analysis identifies oncogenic network modules**



**Mutual exclusivity** of genomic alterations, indicating that genes belonging to the same functional pathway tend not to mutate simultaneously in the same patient, has been observed in various cancer types. <https://doi.org/10.1155/2019/4860367>

* **[@Diaz-Uriarte2015]**

**Cancer progression models (CPM)**: methods developed to try to **infer the possible restrictions in the order of accumulation of driver mutations from cross-sectional data**. Examples of methods:

* Oncogenetic tree (OT) model: in OTs progression starts from a common (non-altered) root, and branches out, so that there are several mutational pathways that can be observed simultaneously.
* Conjunctive Bayesian Networks (CBN): developed as a generalization of OTs. These are graphs where the occurrence of a mutation can depend on the occurrence of two or more parents (conjunction).

The disease progression models of OTs and CBNs assume that a mutation can only occur with non-negligible probability if the preceding parent mutation(s) in the graph have occurred, which has been called **monotonicity**. Thus, for driver genes, under strict OT and CBN models it would be impossible to observe a genotype that is not compatible with the relations specified in the graph.

Progression Networks have been proposed for learning models that include OTs, CBNs and can explicitly incorporate **deviations from monotonicity**. Retracing the Evolutionary Steps in Cancer (RESIC) differs from other methods because it attempts to find the order of events considering the evolutionary dynamics of mutation accumulation.

Having a **single graph means having a single set of restrictions that is common to all individuals**, but that does not mean that all cells follow the same path (so the actual genotypes and their paths can be quite diverse under one graph).

Problems derived from the used of the previous models:

1. First, **most of the mutations present in cancer cells are not driver mutations**, but passenger mutations not responsible for the development of cancer.
2. Our experimental data are from samples that aggregate over many cells and the joint and marginal frequencies of mutations of those aggregates can depend not only on the aggregation per se but also on when we sample (due, for example, to the clonal expansion episodes), and differ greatly from distributions obtained from the generative models.
3. Finally, development and evaluation of methods of reconstruction of order restrictions are conducted **without consideration for the evolutionary model of tumor progression**. This lack of consideration for the evolutionary model does not provide a clear mechanism for interpretation of deviations from the restrictions encoded in the graph. Of particular interest is **monotonicity** (a mutation in a driver gene can only be observed if the preceding parent mutations in the graph have occurred).

Data simulated from the generative models of OTs, CBNs and Progression Networks cannot be used to address any of those three problems (passengers, sampling, deviations from monotonicity). However, **it is possible to incorporate the order restrictions encoded in CBNs, OTs, and Progression Networks into plausible evolutionary models of tumor progression**.

Advantages for incorporation restrictions encoded in CPM into evolutionary models of cancer progression:

1. If we model together drivers (with possible restrictions) and passengers we can **address the consequences of having to filter drivers from passengers**.
2. Whether to use **single cell sampling** instead of tumor sampling and whether is better to collect samples at final stages or at **intermediates** **stages** also.
3. Using explicit evolutionary tumor growth models also allows us to examine the **consequences of deviations from monotonicity** and the genetic context dependence of driver status.

All studies of methods for inferring order restrictions are susceptible to this criticism since they simulate data directly from the generative (but non-evolutionary) OT/CBN/Progression Network model. This problem can be overcome by using plaus**ible evolutionary models that incorporate the restrictions in the order of mutations via a straightforward effect on the fitness of clones**. Moreover, deviations from monotonicity are not added to the model just as an unexplained error term but are an integral part of the evolutionary model that can be related, for instance, to the genetic context-dependence of the driver/passenger status.

**If different subjects have different sets of order restrictions, then no single OT will capture these patterns**, a limitation that is already recognized in the early literature on ongenetic trees, and that has prompted the development of mixtures of oncogenetic trees.

* **[@Sprouffske2011]**

A fundamental goal of research into carcinogenesis, with implications for **cancer prevention**, is to determine the **order of mutations that occur in a neoplasm** as it progresses from normal tissue to cancer. The order of mutations leading to colon cancer was constructed from these data by identifying the mutations whose frequency across tumors increased in conjunction with increases in the tumor size and grade.

A **path model** for a type of cancer is a **linear sequence of mutations that must occur in order**, beginning with wild type (non-altered). Because this is a path, each mutation has at most one following mutation (monotonicity), and a given mutation increases the chances of finding the next mutation. Oncogenetic tree models were developed to capture multiple independent mutational events, in effect enumerating several possible paths to cancer.

“**This is not an evolutionary model because the oncogenetic tree does not represent ancestral relationships within a neoplasm but rather a summary of the observed co-occurrences of mutations across independent neoplasms**.” The interpretation is that a given mutation increases the chances of obtaining the following mutation.

Here, we show that the temporal, or evolutionary, **order of mutations acquired in the clones that survive over the lifetime of a neoplasm are not consistent with the path order or the oncogenetic tree order reconstructed from cross-sectional data**.

It is currently unknown which of these clonal evolution dynamics are found in real tumors, though there is evidence for high mutation rates and small fitness effects in colon cancer.

We have shown that an **evolutionary analysis applied to intratumor samples can overcome the limitations of cross-sectional analyses**. This can also resolve the conflicting results arising from analyses using cross-sectional data.

* **[@Diaz-Uriarte2017]**

**Cancer progression models (CPMs)** use genotype frequency data from **cross-sectional samples to identify these constraints**, and return **Directed Acyclic Graphs** (DAGs) of restrictions where arrows indicate dependencies or constraints. On the other hand, fi**tness landscapes, which map genotypes to fitness, contain all possible paths of tumor progression**. Thus, we expect a correspondence between DAGs from CPMs and the fitness landscapes where evolution happened.

Finding these constraints can single out therapeutic targets and disease markers and has lead to the development of cancer progression models (CPMs), that try to identify these constraints using genotype frequency data from cross-sectional samples. CPMs return directed acyclic graphs (DAGs) of restrictions where arrows between genes indicate direct dependencies or constraints in the order of accumulation of mutations. Under the CPM model, only genotypes that fulfil the restrictions encoded by the arrows in the DAG can exist.

**Whereas DAGs of restrictions from CPMs do not contain information about the fitness of individual genotypes, fitness landscapes (or genotype-fitness maps) associate to each genotype its fitness value**. Thus, similarly to DAGs of restrictions, a fitness landscape, if we assume that populations will only move uphill in fitness, specifies what genotypes can be observed. If we had detailed knowledge about the fitness landscape, we could predict the possible paths of tumor progression and identify genes that would block those paths.

DAGs of restrictions should provide accurate predictions about what genotypes can and cannot exist during tumor progression, and the **same landscape should not lead to inferring widely different DAGs.**

The DAGs and the landscapes make the same predictions about what genotypes we should observe. The landscapes are representable because the DAGs of restrictions capture the epistatic interactions that determine what genotypes are accessible. The constraints reflected in the DAGs of restrictions imply **sign epistasis, an interaction between genes where a mutation is beneficial or deleterious** (i.e. can have different sign) **depending on the genetic background or what other genes are mutated**.

Although DAGs of restrictions represent sign epistasis, they cannot represent **reciprocal sign epistasis**, a genetic interaction where two mutations that individually increase fitness reduce it when combined, or vice versa. The DAGs of restrictions can only say what mutations need to be present before another mutation is viable. Nor any other DAG of restrictions, could represent the fitness landscape that would result if reciprocal sign epistasis between, say, genes A and C turned genotype AC into a low fitness or non-viable genotype. This is a potentially serious limitation of CPMs because reciprocal sign epistasis is probably common in cancer.

**Synthetic lethality** is an **epistatic interaction** where the combination of two mutations is lethal when each individual mutation is not; synthetic lethality between mutations that individually increase fitness constitutes reciprocal sign epistasis. Moreover, reciprocal sign epistasis is a key structural feature of fitness landscapes: it can lead to **multiple peaks and affects ruggedness and predictability of the evolutionary process**, which itself affects our opportunities to block tumor progression.

CPMs assume that **acquiring a mutation in one gene does not decrease the probability of acquiring a mutation in another gene.** We also saw that the same fitness landscape produced genotype frequency data that lead to inferring widely different DAGs.

**pathTiMEx: Joint Inference of Mutually Exclusive Cancer Pathways and Their Progression Dynamics**

Virtually all cancer progression models are designed to infer tumorigenesis from cross-sectional data, that is, **single-time snapshots from multiple patients**.

**Conjunctive Bayesian Networks (CBNs)**

**directed acyclic graphs (DAG).**

Dudas:

* If we had detailed knowledge about the fitness landscape, we could predict the possible paths of tumor progression and identify genes that would block those paths.

**Weinreich, D. M., Watson, R. A., & Chao, L. (2005). *PERSPECTIVE: SIGN EPISTASIS AND GENETIC COSTRAINT ON EVOLUTIONARY TRAJECTORIES. Evolution, 59(6), 1165–1174.* doi:10.1111/j.0014-3820.2005.tb01768.x**

Under sign epistasis, mutations are beneficial on some genetic backgrounds and deleterious on others; in other words, the sign of the fitness effect of such a mutation is conditional on genetic background.

**Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M., & Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. Nature, 445(7126), 383–386. doi:10.1038/nature05451**

Box 1. Different class of epistasis.

**Reciprocal sign epistasis**. Synthetic lethality is an epistatic interaction where the combination of two mutations is lethal when each individual mutation is not; synthetic lethality between mutations that individually increase fitness constitutes reciprocal sign epistasis. Moreover, reciprocal sign epistasis is a key structural feature of fitness landscapes: it can lead to multiple peaks and affects ruggedness and predictability of the evolutionary process, which itself affects our opportunities to block tumour progression.

For other fitness landscapes, however, the correspondence cannot hold. Although DAGs of restrictions represent sign epistasis, they cannot represent reciprocal sign epistasis, a genetic interaction where two mutations that individually increase fitness reduce it when combined.

The DAGs of restrictions can only say what mutations need to be present before another mutation is viable. Thus, neither the DAG in Figure 1e, nor any other DAG of restrictions, could represent the fitness landscape that would result if reciprocal sign epistasis between, say, genes A and C turned genotype AC into a low fitness or non-viable genotype.

But the assessment of CPMs has used data simulated from generative models that are encoded by DAGs of restrictions therefore assuming very restricted fitness landscapes.

**same fitness landscape produced genotype frequency data that lead to inferring widely different DAGs**

(Cristea et al., 2017)