Cancer progression models in evolutionary processes with frequency-dependent fitness

Mario Aguilar[†].

† Departamento de Bioquímica. Facultad de Medicina. Universidad Autónoma de Madrid.

< *mario.aguilarh@estudiante.uam.es* >

ABSTRACT: Cancer progression models (CPMs) use cross-sectional data, widely available in public databases, to understand the restrictions on the accumulation of mutations during tumour progression. CPMs can also be used to predict possible tumour progression pathways. Therefore, CPM predictions could be used to improve decisions when using adaptive therapies. The evolutionary model underlying CPMs is relatively simple if we assume that the restrictions on the order of mutations are concordant between genotype and phenotype. Yet, its function and its interpretation have not been examined when there is frequency-dependent fitness, which is common in cancer, for example when there is cooperation between different types of tumour cells. In general, the literature on these topics is straightforward and models used are not big enough to show further dynamics. Furthermore, there are not simulations in which new genotypes (new cell types) appear over time.

Introduction

Cancer evolves by the accumulation of mutations. We differentiate between driver mutation, that confers a growth advantage to the cancer cell in which it occurs, and passenger mutations, that do not impact fitness. Additionally, there are both cooperation and competition scenarios within cancerous environments, and that makes fitness a function of type frequency. The effects of a given mutation depend not only on the genetic background of the cell in which it appears, but also on the cells that are around it. Further-more, the concordance between the genotype of a clone and the presented phenotype (i.e., cell-autonomous effect) can be masked by other non-cell-autonomous effects, where the re-lease of extracellular signals may provide a way to recruit adjacent cells or communicate with them.[1,2] For instance, signalling between tumour cells was reported in glioblastomas, involving two related cytokines (IL6 and LIF).[3] These cytokines stimulate tumour growth, so one cell can benefit from what another secrets.

Today, several cancer progressions models (CPMs) have been developed to study the effect of mutations and to identify restrictions on the order of mutations during tumour progression,[4] but it is not clear what the real effect is when these data have intrinsic frequency-dependant fitness or not. Hence, our questions are: 1) Do subclone interactions drive cancer cell adaptation? Does this amplify intratumoral heterogeneity? 2) How CPMs behave when we have data with frequency-dependent fitness and when they are cell-autonomous? In this work, the current situation with CPMs and its performance in evolutionary processes with frequency-dependent fitness is reviewed. We start with a review of cancer, but also evolution theory and evolutionary game theory.

What is cancer?

Cancer is a generic term that defines a group of diseases that generically groups more than 200 pathologies. This disease starts when some of the body's cells grow uncontrollably and spread to other parts of the body. In most developed countries it is the second cause of mortality, preceded by cardiovascular diseases.[5,6]

Cancer can start almost anywhere in the human body. Normally, human cells grow and multiply (i.e., cell division) to form new cells as the body needs them. When cells grow old or become damaged, they die, and new cells take their place. However, this orderly process breaks down, and abnormal or damaged cells grow and multiply when they should not. These cells may form tumors, which are lumps of tissue. Tumors can be cancerous (malignant tumors) or not cancerous (benign). Cancerous tumors spread into nearby tissues and can travel to distant places in the body to form new tumors (i.e., metastasis). Benign tumors do not invade nearby tissues. When removed, benign tumors usually do not grow back, whereas cancerous tumors sometimes do. Furthermore, benign tumors can sometimes be quite large causing serious symptoms or be life threatening (e.g., benign tumors in the brain).[6,7]

Normally, when a tumor is detected, there is no longer just a single mutation, but several. However, the initiator mutation will appear in all cells that are part of the tumor, despite its heterogeneity. Commonly, there are between 6-8 different mutations, which directly or indirectly imply, in general, an advantage in cell growth and proliferation. Therefore, most cancers will be caused by between 6-8

mutations, and this development takes place over 20-30 years in humans.[7]

When mutations increase net cell growth under the specific micro-environmental conditions that exist in the cell in vivo, these mutations are called driver mutations. These are sometimes caused by epigenetic mechanisms, and, in addition, there are general alterations, common to all tumors, although many others are tumor specific. Not all tumor cells have the same mutations (heterogeneity), so this event has a significant impact on targeted therapies.[7,8]

Evolutionary theory applied to cancer

During the 20th-century, the established mechanisms in the theory of evolution proposed by Darwin and Wallace were combined with Mendel's inheritance laws, giving rise to Neo-Darwinism, which only recognizes natural selection and genetic mutations as factors that cause evolution.[9] In 1969, Kimura introduced Neutralism as a counterpoint to Neo-Darwinism, holding that most of the variation within and between species are due to random genetic drift of mutant alleles that are selectively neutral.[10] Nowadays, there is evidence of both points of view and they are recognized as complementary mechanisms, because both serve to explain part of the frequency variability that is inherent in evolution, defining evolution as a change in the allelic frequency of a population throughout generations by means of mechanisms such as natural selection, genetic drift, mutation and gene flow.[9]

Currently, thanks to the advancement of modern molecular biology and biochemistry, we understand the nature of genotypes and we have amassed much information concerning the structure and function of biological molecules which form molecular networks and determine all observable characteristics of organisms, i.e., their phenotype.[11]

Through different generations, a phenotype can be preserved while changing the genotype. This arises through the genetic code, which, by being degenerate, causes many of the changes that come about the nucleotide sequence to have no effect on the amino acid sequence and, therefore, on the protein. There are more ways to change the genotype but no the phenotype, for example, through several mutations, generating a set of metabolic networks quite different from the original, but all with the same metabolic flux (i.e., same phenotype).[12]

Nevertheless, although many variations of the genome do not change the phenotype, to produce evolutionary innovations, biological system must explore many phenotypic variants (through mutations) before finding one that may become an evolutionary advantage, but this must be done without losing selectively relevant characteristics previously found.[12]

If this idea is put in the context of the populations of individuals, intuitively, evolution is linked to large populations because the larger populations are, the larger is the phenotypic space that evolution can explore, but, however, evolution can occur even in small populations, as the theory of neutralism emphasizes. Indeed, Neutralism is more important the more restricted is the population size.[10] In summary, evolutions is the result of 1) appearance of variation 2) that is heritable and 3) that can spread in the population (so that there is a change in allelic frequency).

An evolving population traces a path in the space of the different possible genotypes. Regardless of the size of the resolution (i.e., the genotype described in terms of nucleotides, amino acids, or genes), it remains interesting how evolution navigates these evolutionary spaces. Assuming that evolution is restricted to a single gene or single nucleotide mutation, the fitness landscape is a good way to encode the mapping of genotypes to fitness and it provides a succinct representation of possible trajectories followed by an evolving population.[13]

Most mutations are neutral or deleterious, so for a beneficial mutation to be acquired by the population, it must not only appear, but be fixed. For example, a beneficial mutation occurs, spreads through the population due to selection, and soon fixes. Later, another such event may occur. This scenario is sometimes called the strong-selection weak-mutation (SSWM) regime. In this regime, mutations are rare (i.e., smaller than mutation rate times the population size) and selection is strong (i.e., much larger than 1/population size).[14,15] Events happen in selective sweeps and between sweeps, there is a single "ruling" population (Fig. 1).

On the way to understanding the different mechanics behind the different regimes that govern a continuously evolving population such as a tumour, evolutionary game theory (EGT) allowed us to link Darwinian fitness of the species and their evolution with game theory machinery.[16,17] This created a very useful methodology for simulation and analysis of dynamics of populations of players that belong to the real and biological world. Those individuals can be characterized by some strategies or phenotypes which can cooperate or compete with others to fulfill their evolutionary objectives.[18]

Evolutionary game theory.

On the evolutionary game theory, players act without any rationality (i.e., players do not need to be rational as in the standard game theory). The reward and expected result should be an achievement of better access to some resources like food, females, living space, etc. The change of this evolutionary achievement is called the payoff. Those payoffs, represented in a matrix form (1), consist of all costs and benefits resulting from this change in the consequence of players interactions.

$$\begin{matrix} & \alpha & \beta \\ \alpha & A & B \\ \beta & C & D \end{matrix} \qquad (1)$$

When a player α encounters a player α, it receives pay-off A. A player α versus a player β receives pay-off B, whereas the other receives pay-off C. A player β versus a player β receives pay-off D. When we consider the absolute population of each type of player (α and β), then the population rate equations can be written as:

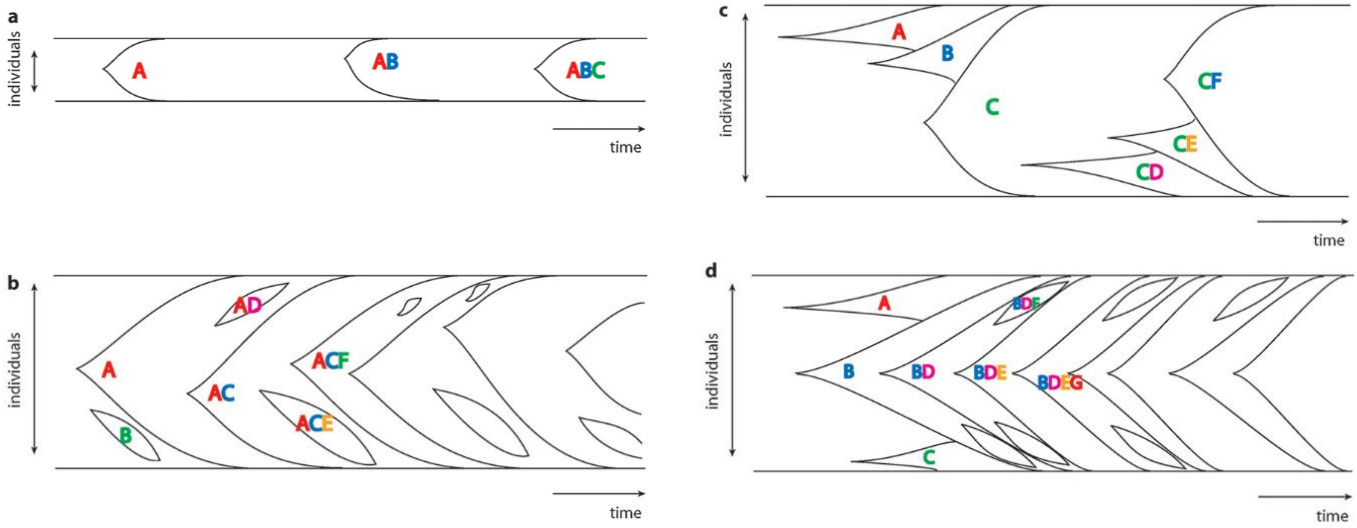

Figure 1. (a) *A small population in the SSWM regime. The beneficial A mutation emerges and fixes. Sequentially, other beneficial mutations (B and C) emerge and fix.* ***(b)*** *A larger population in the strong strong-selection strong-mutation regime. A mutation A occurs, but before it can fix another mutation B occurs and the two interfere. A second mutation, C, occurs in an individual with mutation A and these two begin fixing together, driving the single mutants to extinction. These dynamics continue with further mutations, E and F, occurring in the already-double-mutant population.* ***(c)*** *The clonal interference effect in large populations: a weak-effect beneficial mutation A occurs and begins to sweep but is outcompeted by a later but more-fit mutation B, which in turn is outcompeted by mutation C. C fixes before any larger mutations can occur; the process can then begin again.* ***(d)*** *The multiple mutation effect: several mutations, A, B, and C, of identical effect occur and begin to spread. Mutant lineage B happens to get a second beneficial mutation D, which helps it sweep, outcompeting A and C. Eventually this lineage gets a third beneficial mutation E. Mutations that occur in less-fit lineages, or those that do not happen to get additional mutations soon enough (such as BDF), are driven extinct.* Figure from Desai & Fisher, 2007, Figure 1, p. 1760.

$$\frac{d\alpha}{dt} = \alpha(Ap_\alpha + Bp_\beta) \qquad (2)$$

$$\frac{d\beta}{dt} = \beta(Cp_\alpha + Dp_\beta) \qquad (3)$$

where population fractions $p_\alpha = \frac{\alpha}{(\alpha+\beta)}$ and $p_\beta = \frac{\beta}{(\alpha+\beta)}$, and the fitness of each player is $f_\alpha = Ap_\alpha + Bp_\beta$ and $f_\beta = Cp_\alpha + Dp_\beta$. If A > C and B > D, then player α dominates player β and if $A < C$ and $B < D$, then player β dominates player α. This is a classic example of frequency-dependent selection.

A coordination game occurs if $A > C$ and $B < D$, then both α and β are stable. In a system with more α, it will keep being dominated by α. If currently there are more β, then there will be more β. More complex dynamics occur if $A < C$ and $B > D$, leading to a coexistence between the two players. In this case, a population with more α will then be dominated by β, and a population with more β will then be dominated by α.

In different contexts, it is certainly more reasonable to frequency-dependent fitness (i.e., the fitness of a phenotype depends on its frequency relative to the other phenotypes in the population) rather than a static fitness concept where only the type of individual (resident versus mutant) determines its reproductive success.

In a frequency dependent setting, the actual fitness of an individual is determined through interactions with other members of the population. The fitness of an individual now depends not only on its own type but also on the type of the interaction partners. This can be easily summarized in a payoff matrix as mentioned above. The mechanisms by which frequency-dependent fitness can occur are mainly diffusible factors, which will produce different effects in the cell (e.g., proliferation. growth) and in the environment (e.g., acidification, vascularization).[1,19]

Other mechanisms by which frequency-dependent fitness may appear are tight junctions (i.e., through connexons)[20] and trogocytosis (i.e., the process by which cells of the immune system extract surface molecules from other cells and present them on its own surface).[21] Although they have not been reported in cancer cells, they are known cellular mechanisms by which cells could communicate and modulate their behavior as a function of frequency.

When density dependence is considered, describing the environment becomes paramount. The environment is influenced by the presence and actions of individuals of the focal population. Measuring fitness under density dependence is based on the ability of mutant types to invade a resident population or, more precisely, the environment prescribed by the resident population. The fittest type is the one able to resist invasions by all the other types.[22] This is the evolutionary stable strategy (ESS) idiosyncrasy: due to interactions

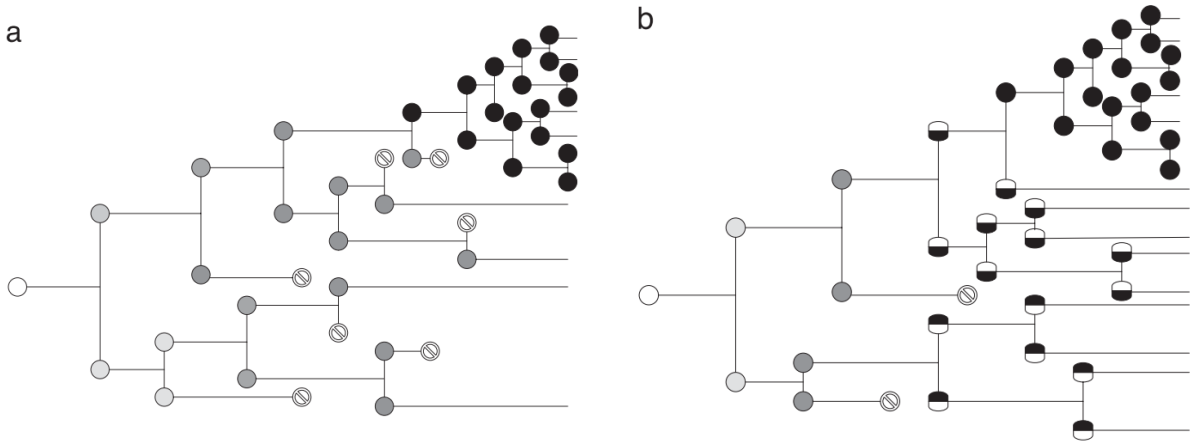

Figure 2. *Two different views of tumor progression.* ***(a)*** *Traditional view: Competition. Genetically unstable partially transformed cells (○) proliferate. The cells compete for limited resources; therefore, many die (⊘). Eventually, one cell accumulates sufficient mutations to express all the functions required to emerge as a fully malignant cell (●).* ***(b)*** *Hypothesis: Cooperation. Genetically unstable partially transformed cells (○) proliferate and yield different mutant cell types (◓, ◒). The different cell types cooperate with each other, enabling them to survive and proliferate. Eventually, one cell accumulates sufficient mutations to express all the functions required to emerge as a fully malignant cell (●).* Figure from Axelrod et *al*., 2006, Figure 1, p. 13475.

between players in time, the population composition may stabilize and achieve some equilibrium state (either mono- or polymorphic). Such a state is called evolutionary stable. Phenotypes that cannot be replaced by any other when implemented by the most of population is called evolutionary stable strategy.[23]

Evolutionary game theory (EGT) methodology allows us to predict in general the behaviour of the population if it tends to become homo- or heterogeneous. The former means that only one strategy will survive and will dominate the population. Additional information can give us the usage of so-called replicator dynamics where one can observe the change in population composition in time starting from an initial state.[24]

Cancer meets game theory

Although cancer is composed of a spectrum of diseases that involves many genotypes and phenotypes, it is genetically understood as the accumulation of mutations in a cell clone. This is a common leitmotif for all types of cancer.[6,7]

There are some important questions for basic cancer research and oncology: What is the role of different kinds of intervention (exposures or treatments) in the evolution of cancer? Can accelerated proliferation make carcinogenesis more likely? Can cell killing enable the evolution of heterogeneity and increase the viability of cancerous cells?

Classical models of cancer development assume that mutations, which promote the growth of cancer cells affect only the units in which they occur.[25] Moreover, mutations can also affect neighbouring cells,[2] and combined with cooperative behaviours, competition for resources such as space, oxygen and nutrition, occurs between different subspecies within the same tumour.[26] Hence, internal communication between tumour cells, and between tumour and normal cells, their competition for resources, hierarchical subordination, and collaboration, play an important role in cancer development and differentiation or disease transmission and reaction to stress factors including therapy.[27] Here is where game theory comes in. For example, Kaznatcheev et. *al*., developed a game assay to measure effective evolutionary games in co-cultures of tumour cancer cells that are sensitive and resistant to alectinib. These results showed that when working with evolutionary game theory in oncology, "we can treat not only the player, but also the game".[28]

On the one hand, the aim of these theoretic models is to study the possibility of coexistence or even domination of newly formed tumour cells, which have acquired new strategies (phenotypes) by mutations. The capability to detect not only cancer cells, but their majority malignant phenotype in a sufficiently early stage of the disease, would allow direct surgical interventions that, otherwise, would not arrive in time.

On the other hand, we have the results of previous works[29] in which evolutionary games are applied to describe the influence of external resources such as stimulant treatments and the cell-killing treatments. They show that, depending on the nature of the intervention, external interventions influence tumour heterogeneity by changing the baseline of the payoff matrix (i.e., changing the baseline cost of the meeting) or by introducing new phenotypes in the population. This can be interpreted indicating that the treatment can generate new phenotypes of cancer cells of the cancer subtype. This discussion is almost absent in the literature.

Furthermore, an additional key factor for game theory applications in oncology is the impact of the ecosystem or the interactions between tumour cells and their environment. This has been studied[30] indicating that elimination of as many cancer cells as possible may not be the best strategy. Indeed, they found that destroying only

some fraction of the cancer cells (with a particular phenotype) may be far more efficient. This is because in EGT the result depends on the interactions between the players, not on the size of the population.

Therefore, a treatment based exclusively on indiscriminately removing most cancer cells may only have a temporary effect, since the rest of the cells would eventually re-establish the tumour cell population. A more effective alternative would be based on changing the way cells interact with each other and with their environment. This could affect fitness and direct the evolution of cancer to aggressive cell types or at least to a stable coexistence that would be less harmful to the patient.

These ideas are in the same lane of Axelrod's paradigm,[31] in which work we were based in order to acknowledge a more in-depth understanding

Cooperation theory

Cooperation occurs between tumours and stromal cells, and it is hypothesized that there is also cooperation between partially transformed tumour cells.[32] Cooperation comes through mutualism between tumour cells partially transformed by sharing by-products. For example, one subclone could produce paracrine GF (growth factor) that another subclone requires and vice versa. Cooperation can also appear in the form of commensalism, where only one subclone benefits. For example, a subclone that induces vascular GF could provide oxygen and nutrients to other non-angiogenic subclones

The traditional view is that the subclones that exist within a tumour accumulate all the hallmarks of cancer developing while avoiding the defence mechanisms of the host (Fig. 2a). The minimum set of genotypes or phenotypes that a cancer cell must acquire to become malignant is called "the hallmarks of cancer". These stamps include: 1) self-sufficiency of growth signals; 2) insensitivity to growth inhibition signals; 3) evasion of programmed cell death; 4) unlimited replicative potential; 5) sustained angiogenesis; and 6) tissue invasion and metastasis. However, cooperation theory suggests that these subclones do not need to accumulate all the hallmarks to proliferate in a tumour niche (Fig. 2b). Cells that have acquired any of these hallmarks, or other changes necessary to become malignant, are referred to as partially transformed. Some hallmarks could cooperate with other partially transformed tumour cells to form a community, creating a favourable environment for manifesting a tumour phenotype and developing as a malignant tumour.

It is important to consider that cooperation could account for the frequency of cancer. Eventually, a clone develops an entire deck of mutations. When a subclone develops the full deck of mutations, it will outnumber those that depend on cooperation. The hypothesis raised of cooperation between tumours implies that a tumour with all the mutations will evolve from a normal cell much faster (therefore, its fitness will be greater) with the possibility of cooperation since, during the intermediate stage, when no cell has accumulated the full set of mutations, the tumour may show rapid (malignant) growth (Fig. 3). This rapid growth means that cells lacking only one compatible resource to have the complete deck, could be proliferating at a high rate, instead of proliferating little or die, as expected in a traditional vision. This hypothesis is consistent with the facts that are already known since at least 3 hallmarks involve the sharing of resources: angiogenesis, self-sufficiency of growth signals and tissue invasion and metastasis.

Considering all the above-mentioned, one way to show that this hypothesis is plausible is to simulate verisimilar scenarios with frequency-dependent fitness which are common in cancer when there is cooperation between different types of tumour cells.

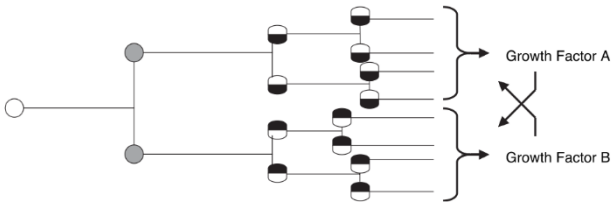

Figure 3. *Intratumor cooperation can occur among partially transformed mutant cells with complementary needs (e.g., different GFs). Some cells ( ) produce only GF A, and other cells ( ) produce only GF B, but both GFs are present in the environment. Crossfeeding is a form of cooperation that enables each cell type to survive and proliferate.* Figure from Axelrod et *al*., 2006, Figure 1, p. 13475.

Cancer progression models (CPMs)

Prediction of possible pathways of tumour progression can be done with cancer progression models (CPMs). CPMs were originally developed to identify restrictions features of tumour progression that are common to a homogeneous type of cancer, where each individual is undergoing an independent evolutionary process with the same restrictions for the entire tumor.[33] CPMs also encode all possible mutational pathways or trajectories of tumour progression, from the initial genotype to the genotype with all mutated driver genes using cross-sectional data.[4]

Hence, CPMs could improve our ability to predict disease progression by taking advantage of the growing and available number of cross-sectional data sets. Figure 4 shows are collected some of the CPMs considered state-of-the-art algorithms:

- OncoTree (OT): Oncogenetic trees are directed tree structures that model the process of occurrence of genetic alterations during carcinogenesis. OncoTree uses cross-sectional data with information on mutations within tumour samples. The accumulation of mutations is described with order constraints, which makes it possible to represent them as trees. OT can only set to represent the alterations as a tree when the mutations of a gene are directly dependent on the mutation of a gene. OT weights along edges can be directly represented as the transition probability at the time of observation.[4,34]

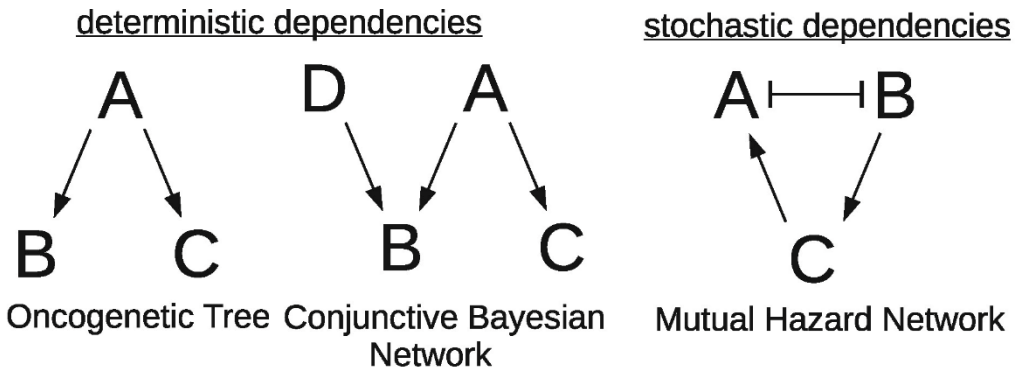

Figure 4. *Overview of several types of cancer progression models. For models with deterministic dependencies, A→B denotes that A is necessary for B, and A⊣B denotes that A prevents B. For models with stochastic dependencies, A→B denotes that A makes B more likely, and A⊣B denotes that A makes B less likely*. Figure from Schill et *al*., 2020, Figure 1, p. 242.

- Conjunctive Bayesian Networks (CBN): CBN is an extension of oncogenetic trees that describes mutation accumulation with order constraints. CBN allows modeling the dependency of mutations on more than one event, so the model outputs are directed acyclic graph (DAGs). In DAGs, some nodes have multiple parents in-stead of just one like in trees. CBN also uses cross-sectional data with information on genomic aberrations as OncoTrees.
  The CBN input is an array of driver alteration events, where each element of the array encodes the presence (or absence) of a mutation using binary. It is from this matrix that a DAG is returned. This CPM models the accumulated mutations in a stochastic way in each patient. It is the fixing of these mutations with respect to a set of order restrictions that we try to infer from the DAG obtained.
  CBN considers the possibility of accommodating errors in the data (e.g., genotyping errors), assuming that each genotype contains all the alterations that appear in its parenteral genotype, that is, there is no back-mutation.
  Finally, CBN uses simulated annealing with a nested expectation-maximization algorithm to estimate the model parameters (i.e., fixation rate, error rate and DAG restrictions).[33,35]

- Mutational Hazard Networks (MHN): In MHN there are no deterministic dependencies but all transitions (to genotypes with an additional mutation) are possible. However, the transitions will be affected by inhibitory stochastic dependencies (with low probability) and promoting stochastic dependencies (with high probability). Hence, in a MHN model there is no strict dependency between events: an event A can be mod agree (or less likely) if another event B occurred previously, but B is not required. Also, there is no need for acyclic dependencies: event A can make event B more or less likely, and vice versa too.[36,37]
- CAncer PRogression Inference (CAPRI): CAPRI is also an extension of oncogenetic trees that describes the accumulation of mutation with order constraints. Like CBN, it allows modeling the dependence of an event on several previous events (i.e., outputs are DAGs). CAPRI uses the same data type (cross-sectional data) with information about genomic aberrations as OT and CBN. The input to CAPRI is a binary transition matrix as in CBN, returning a constraint DAG.
  CAPRI tries to identify alterations that constitute some kind of selective advantage, where an alteration in one gene predispose to undergo a subsequent alteration in another gene.
  Regarding errors in observational data, CAPRI is also able to deal with them, and also assumes that each genotype contains all the alterations that appear in its previous genotype (as in CBN).
  CAPRI uses Mann-Whitney U-tests (from bootstrap resamples of the input data) to examine temporal priority and probability rising and then optimization (e.g., using hill climbing) of the penalized (AIC or BIC) maximum likelihood fit to find the final DAG.[33,38]

- CAncer PRogression Extraction with Single Edges (CAPRESE): CAPRESE is very similar to OT in that it describes the accumulation of mutations with order constraints that can be represented as trees. Furthermore, only those with a genetic mutation with direct dependence on another mutation can be faithfully represented. The main difference is that CAPRESE reconstructs these models using a prob-ability raising notion of causation in the framework of Suppes' probabilistic causation.[4,39,40]

CPMs with frequency-dependent fitness

As already mentioned, frequency-dependent fitness refers to the phenomenon where the fitness of a phenotype or genotype depends on the frequency of other phenotypes or genotypes. Evolutionary game theory has been established as a quantitative way of reflecting frequency-dependent fitness within an interactive population. For instance, although is not usual in cancer biology, competitive fitness assays are a gold standard for studying bacteria. They are typically conducted with a single initial ratio of the two competing cell types. Kaznatcheev et. *al*., showed that if the initial proportion of parental to resistant cells is seen as a variable parameter represented by optical density, a hint of frequency dependence can be seen in both parental and resistant growth rates.[28] The use of different drug concentrations allowed them to define a game assay to quantitatively describe their system in the language of EGT. The difference

in the growth rate between resistant and parental cells as a function of the proportion of parental cells allowed them to describe the interaction as a matrix game. Finally, this allowed them to show the statistical significance that effective games allow to quantify frequency-dependent differences in growth rates in non-small cell lung cancer.

But then we have the CPMs that are supposed to encode implicitly all the possible mutational trajectories that are compatible with the restrictions specified.[36] Current literature offers little information about predictive models, such as CPM, considering the impact of the frequency of an individual within the population.

Diaz-Colunga, J. & Diaz-Uriarte, R., reported that, "CPMs could, under very specific combinations of genotype characteristics and fitness landscape characteristics, be used to obtain good predictions of the short-term evolution of a tumour" even when we work without explicit frequency-dependent fitness relationships. Thus, the CPMs performance to infer the complete evolutionary trajectories (i.e., from the root genotype with no driver mutations to a final fixated genotype) is limited.[36]

Although tumour prediction in processes with frequency-dependent fitness is a question that currently remains open, at this point it is questionable whether the prediction of the evolutionary path is the most remarkable objective. On this stage, the behaviour the host's cell with regard to the different tumour genotypes could be of greater importance than the prediction of the complete pathway. The fact that the models have frequency-dependent fitness could bias the outcome, so that key genotypes in tumour progression are no reflected in final genotype (e.g., not oncogenic cells which could ease the adaptation of the tumour environment).

Cell-autonomous vs. non-cell-autonomous

First, it is important to establish the meaning of cell autonomous and non-cell autonomous. Although the cells are part of the ensemble that is a human organism, cells behave autonomously. They have their own degradation and synthesis pathways, cell development and death, etc.[41] That a cell has a mutated gene and ends up developing into a tumour does not mean that its immediate neighbour will have it as well (although it is quite probable) since cells, in principle, are understood as autonomous systems. Hence, cancer diseases have been considered mechanistically as cell-autonomous pathologies, meaning that damage or defect within a selective population of affected cells suffice to produce disease.[41]

However, it is the whole set of cells of a certain environment that participates in different processes on a larger scale such as immunity or tissue remodelling. This is done to maintain a relative constancy in the composition and properties of the internal environment of an organism. This is what is known as homeostasis.[41,42]

Thus, non-cell autonomous is understood as communication existing between cells and their environment by means of product signalling molecules or by participating in the synthesis of signalling molecules, while in cell autonomous dogma do not participate in a process that involves cell-cell interactions. In addition, non-cell autonomous signalling pathways allow coordination of stress responses across tissues, thus ensuring maintenance of cellular homeostasis at an organismal level.[41]

Therefore, it is meaningful to account for interactions with the microenvironment when predicting cancer progression. Bardelli, A. *et al.*, established that the emergence of therapy resistance is inevitabe in colorectal cancer and significantly affects survival of cancer patients.[43] The tumour microenvironment can influence therapy efficacy because cancers are heterogeneous cellular entities whose growth is dependent upon reciprocal interactions between genetically altered cells and the dynamic microenvironment in which they live.[44] Marusyk, A. et *al*., point out that "recent data from tumour genome sequencing studies and single-cell based analyses has revealed substantial genetic heterogeneity within tumours", which contradicts the linear model of acquisition of mutations.[45,46] This also challenges the assumption that tumour evolution is driven by mutations that provide strong clone-specific selective advantages, since clone heterogeneity possibly arises from interaction between clones.[1,45]

Considering all the above-mentioned statements, we can easily trace the similarity with what Axelrod and his theory of cooperation propose.[31] Looking back briefly, the main idea is that different cell clones within the microenvironment of the tumour can be recruited, or evolve, through cooperation (i.e., paracrine mutualism and commensalism interactions) to become a fully complete tumour ensemble. Therefore, partially transformed cells with genotypes that would not be possible in a linear acquisition of mutations, would be allowed because other fully transformed or partially transformed cells would be sharing their phenotypic advantages.

Once again, the authors of the main CPMs are not addressing these kinds of possibilities, where these dynamics may underlie the available cancer data. Diaz-Colunga, J. & Diaz-Uriarte, R., already noticed that "CPMs could, under very specific combinations of genotype characteristics and fitness landscape characteristics, be used to obtain good predictions of the short-term evolution of a tumour". However, there are some authors that computationally modeled dynamic microenvironment-dependent fitness alterations, whose stochastic approach allowed them to model non-linear cell dynamics and the complex relationship between cell-autonomous and non-cell-autonomous processes, finding out that some evolutionary events are driven by non-cell-autonomous processes.[47]

It is interesting to see how most of current carcinogenesis' models are dominated mainly by the premise that oncogenic mutations have defined advantageous fitness effects on recipient stem and progenitor cells. From this point of view, these mutations are capable of driving evolution when they appear, determining a selection process that leads to cancer.

I need to emphasize that this assumption disagrees with evolutionary theory. Fitness is a dynamic property of the phenotype, imposed and modulated by the environment.

CPMs applications

After all the above, the following question arises naturally: have CPMs been used, or can they be used clinically?

CPMs are supposed to enable prediction of the next step(s) of the disease evolution of any given patient, with evident implications on prognostic strategies and therapeutic interventions. Nowadays, the inference problem is complicated by the high levels of heterogeneity typically observed in most tumor types. This heterogeneity is a consequence not only of the existence of multiple independent evolutionary trajectories but also the complex network of events and dependencies that underlies oncogenesis (and therefore, in the data obtained).

Hence, some CPMs have overcome limitations by allowing one to assess the existence of arbitrary logical formulas connecting events, which require that such formulas are identified prior to the inference and provided as input.[38] All this requires having a deep biological knowledge about the evolutionary process or, alternatively, employing *ad hoc* computational strategies to identify specific patterns between events.

Regarding what has been done in the literature, there is no trial or proposal to guide in a real way an intervention in a patient to stop or prevent the progress of the disease. Although it is true that there is work that models the possible effects of an intervention in scenarios such as those proposed (i.e., cancer population evolution based on the evolutionary games with resources),[29] most CPMs provide nothing more than a hopeful phrase that CPMs will be able to reliably predict and possibly intervene to halt or slow down the disease progression,[38,40,48] having an "important impact on downstream clinical practices and therapeutic strategies".[49] The reality is that currently there is currently no such a thing as a therapeutic strategy led by CPMs.

At least, what does seem to be clear is the importance of integrating not only the dynamics that underlie CPMs in evolutionary processes with frequency-dependent fitness, but also the interventions themselves as players in the game.[29]

Conclusions

To our initial questions "Do subclone interactions drive cancer cell adaptation? Does this amplify intratumoral heterogeneity?", the experimental data seem to show that the observed effect of heterogeneity of cancer cells is not only at the population level but also at the level of single cells.[3,28,29,50] Cancer is not only a disease of the genome, but also of abnormal cellular interactions. Human cancers frequently display in distinguishable phenotypic features such as cellular morphology, metabolism, and metastatic potential, so probably the interaction between all these features causes a better adaptation to the tissue microenvironment. For example, the fitness of a clone depends on its genotype and the tissue environment the cells live in.[48] In addition, the tissue microenvironment has many cellular components that can modify tumor progression and evolution. Hence, the tissue microenvironment is a complex dynamic system that influences cancer cells and vice versa (i.e., acidification, vascularization).

Regarding to the second question "How CPMs behave when we have data with FDF and when they are cell-autonomous?", unfortunately the literature does not seem to have explored this possibility so much. Despite this, it is interesting to ponder on the ability of CPMs to predict cancer. As said in Diaz-Uriarte, R. & Vasallo, C., we can predict the likely course of tumor progression using CPMs "only with moderate success and only under representable fitness landscapes and with very large sample sizes".[4]

In general, the literature on these topics is straightforward. The models are made up of two or three well-differentiated cell types (tumor vs infiltrated cells) and the dynamics that are established between them can be seen clear. But there are not simulations in which new genotypes (new cell types) appear over time. This leads us on how to define fitness. How do we see the fitness of cell A that benefits from B (via the payoff matrix) vs the mutant AB. As mentioned, Kaznatchev et *al*. already modelled changes in the cancer ecosystem in the context of different therapeutic strategies but authors focus on the interdependence of growth-factor production and acidification instead of studying the resulting dynamics with more than three phenotypes or with a broad definition of fitness.[28]

Further perspectives

Accordingly, how to determine a kind of game that cancer cells play and how to experimentally find parameters for game models still is an open question. However, we want to propose an approach to address this problem. We propose to simulate different scenarios with and without frequency-dependent fitness, analysing the data from the simulations through different CPMs and giving biological meaning to the different inferences obtained.

After an extensive bibliographic search to find data and models capable of representing an oncogenic scenario with frequency-dependent fitness in a realistic way (i.e., with patient data, real numbers and / or kinetic constants), we found out that very little is known, being an heterogeneous breast tumour progression model the only example of it.[51]

The model describes the progression through the grades of ductal carcinoma in situ (DCIS) 1, 2, and 3, and through the grades of invasive ductal carcinoma (IDC) 1, 2, and 3 in four different pathways.

The first three pathways, denominated linear, nonlinear, and branched, describe DCIS as a progenitor of IDC, and grades of DCIS progressing into grades of IDC. The fourth pathway, termed parallel, describes DCIS and IDC as diverging from a common progenitor and progressing through grades in parallel. The objective is not so much to replicate what they collect in their model, but to propose a credible scenario, so the 4 scenarios could be adapted as follows:

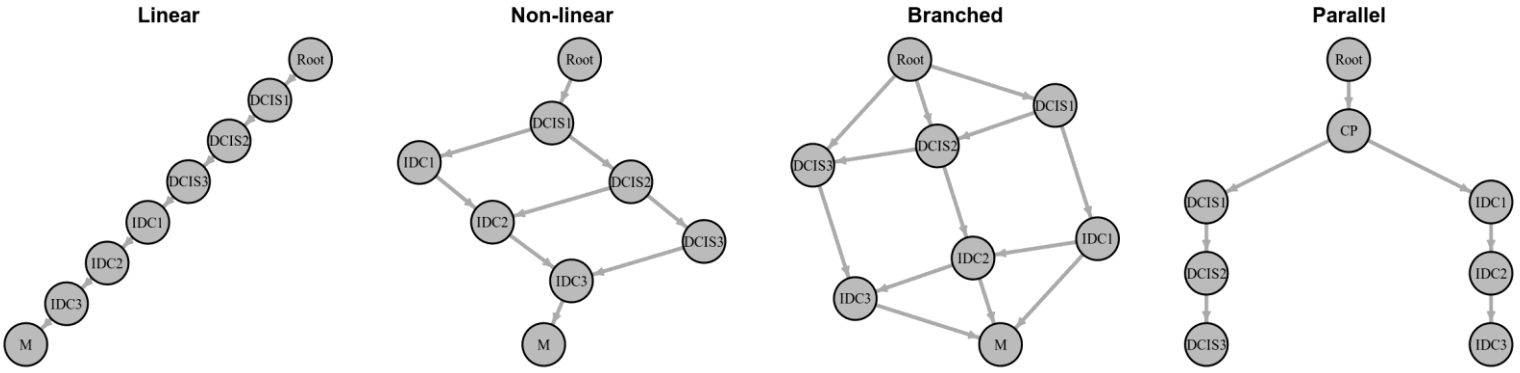

Figure 6. *Different scenarios proposed by Axelrod. Each node stands for different degrees of a disease. ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), and metastasis (M). Grades of DCIS and IDC are indicated by 1, 2, and 3.*

- Linear pathway. The linear pathway allowed progression consecutively through the three grades of DCIS, from DCIS 3 to IDC 1, and then through the three grades of IDC.
- Nonlinear pathway. The nonlinear pathway considers the possibility that transitions occur between equivalent grades of DCIS and IDC. Hence, this pathway has nine different transition rate constants, one for each of the possible transitions.
- Branched pathway. The branched pathway differs from the nonlinear pathway in that it includes additional transitions between the wildtypes (WT) and each grade of DCIS and between each grade of IDC and metastasis (M).
- Parallel pathway. In the parallel pathway, transitions do not occur between grades of DCIS and grades of IDC. DCIS is not a progenitor of IDC, but rather, DCIS and IDC diverge from a common progenitor. There is progression from DCIS 1 to DCIS 2, to DCIS 3, and in parallel, there is progression at about the same rate from IDC 1 to IDC 2, to IDC 3.

The 4 scenarios were represented as DAG (Fig. 6). The different nodes are understood in this work as that each genotype (or phenotype) has a series of mutations that must accumulate to give rise to the phenotype itself, although we are not interested in which mutations are them. For example, in the linear pathway, to present the IDC1 phenotype, we need the genotype / phenotype above and its progenitors, so IDC1 will have the DCIS1, DCIS2, DCIS3 and IDC1 mutations. If an individual did not present in its genome, for example, DCIS3, it would not be able to reach IDC1, or, on the contrary, its fitness would be lower so it would become extinct.

The interesting thing now, would be to integrate frequency-dependent fitness. This would allow that, for example, if we have a genotype that accumulates mutations for DCIS1 and DCIS3, this genotype could be viable if there were some genotypes in the population with the relevant mutations to have DCIS3. This would allow surviving "incomplete" genotypes and even advancing to the next canonic evolutionary step, which, in the linear scenario, would be to reach the IDC1 mutation. In this way, a frequency-dependent fitness scenario could be implemented, with the biological significance that the genotypes with greater frequency are sharing resources with the others (e.g., via diffusible factor).

Code availability

The code used both to create the different scenarios and fitness landscape is available in a GitHub repository (https://github.com/maragu04/TFM).

REFERENCES

1. Marusyk, A. *et al.* Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58 (2014).
2. Tomlinson, I. & Bodmer, W. Modelling the consequences of interactions between tumour cells. *Br. J. Cancer* **75**, 157–160 (1997).
3. Inda, M.-M. *et al.* Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. *Genes Dev.* **24**, 1731–1745 (2010).
4. Diaz-Uriarte, R. & Vasallo, C. Every which way? On predicting tumor evolution using cancer progression models. *bioRxiv* 1–29 (2018) doi:10.1101/371039.
5. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. cancer* **136**, E359-86 (2015).
6. Haberkorn, U. What is cancer? *Adv. Nucl. Oncol.* **62**, 1–16 (2007).
7. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
8. Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14330–14335 (2016).
9. Kutschera, U. & Niklas, K. J. The modern theory of biological evolution: An expanded synthesis. *Naturwissenschaften* **91**, 255–276 (2004).
10. Kimura, M. The neutral theory of molecular evolution: A review of recent evidence. *Japanese J. Genet.* **66**, 367–386 (1991).
11. Chuang, H.-Y., Hofree, M. & Ideker, T. A Decade of Systems Biology. *Annu Rev Cell Dev Biol* **26**, 721–744 (2010).
12. Wagner, A. *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*. (Oxford University Press., 2011).
13. Krug, J. Accessibility percolation in random fitness landscapes. *Probabilistic Struct. Evol.* 1–22 (2021) doi:10.4171/ecr/17-1/1.
14. Gillespie, J. H. Molecular Evolution Over the Mutational Landscape. *Evolution (N. Y).* **38**, 1116–1129 (1984).
15. Desai, M. M. & Fisher, D. S. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
16. Sigmund, K. & Nowak, M. A. Evolutionary game theory. *Curr. Biol.* **9**, R503–R505 (1999).
17. Smith, J. M. *Evolution and the Theory of Games*. (Cambridge University Press, 1982). doi:DOI: 10.1017/CBO9780511806292.

18. Swierniak, A., Krzeslak, M. & Borys, D. *Games with resources and their use in modeling control processes in heterogeneous populations*. *2019 24th International Conference on Methods and Models in Automation and Robotics, MMAR 2019* (Springer International Publishing, 2019). doi:10.1109/MMAR.2019.8864644.
19. Kaznatcheev, A., Vander Velde, R., Scott, J. G. & Basanta, D. Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *Br. J. Cancer* **116**, 785–792 (2017).
20. Lodish, Harvey F.; Arnold Berk, Paul Matsudaira, Chris A. Kaiser, Monty Krieger, Mathew P. Scott, S. Lawrence Zipursky, J. D. Molecular Cell Biology. in 230, 231 (W.H. Freeman and Company., 2004).
21. Joly, E. & Hudrisier, D. What is trogocytosis and what is its purpose? *Nat. Immunol.* **4**, 815 (2003).
22. Heino, M., Metz, J. A. J. & Kaitala, V. The enigma of frequency-dependent selection. *Trends Ecol. Evol.* **13**, 367–370 (1998).
23. Smith, J. M. & Price, G. R. The Logic of Animal Conflict. *Nature* **246**, 15–18 (1973).
24. Hofbauer, J., Schuster, P. & Sigmund, K. A note on evolutionary stable strategies and game dynamics. *J. Theor. Biol.* **81**, 609–612 (1979).
25. Fisher, J. C. Multiple-mutation theory of carcinogenesis. *Nature* **181**, 651–652 (1958).
26. Anderson, A. R. A., Rejniak, K. A., Gerlee, P. & Quaranta, V. Microenvironment driven invasion: A multiscale multimodel investigation. *J. Math. Biol.* **58**, 579–624 (2009).
27. Basanta, D., Hatzikirou, H. & Deutsch, A. Studying the emergence of invasiveness in tumours using game theory. *Eur. Phys. J. B* **63**, 393–397 (2008).
28. Kaznatcheev, A., Peacock, J., Basanta, D., Marusyk, A. & Scott, J. G. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nat. Ecol. Evol.* **3**, 450–456 (2019).
29. Swierniak, A., Krzeslak, M., Borys, D. & Kimmel, M. The role of interventions in the cancer evolution - An evolutionary games approach. *Math. Biosci. Eng.* **16**, 265–291 (2019).
30. Basanta, D. & Anderson, A. R. A. Exploiting ecological principles to better understand cancer progression and treatment. *Interface Focus* **3**, (2013).
31. Axelrod, R., Axelrod, D. E. & Pienta, K. J. Evolution of cooperation among tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13474–13479 (2006).
32. Esposito, A. Cooperation of partially transformed clones : an invisible force behind the early stages of carcinogenesis. *R. Soc. Open Sci.* **8**, 201532 (2021).
33. Diaz-Uriarte, R. Cancer progression models and fitness landscapes: A many-to-many relationship. *Bioinformatics* **34**, 836–844 (2018).
34. Kundra, R. *et al.* OncoTree: A Cancer Classification System for Precision Oncology. *JCO Clin. Cancer Informatics* 221–230 (2021) doi:10.1200/CCI.20.00108.
35. Beerenwinkel, N., Eriksson, N. & Sturmfels, B. Conjunctive Bayesian networks. *Bernoulli* **13**, 893–909 (2007).
36. Diaz-Colunga, J. & Diaz-Uriarte, R. Conditional prediction of consecutive tumor evolution using cancer progression models: What genotype comes next? *bioRxiv* (2020) doi:10.1101/2020.12.16.423099.
37. Schill, R., Solbrig, S., Wettig, T. & Spang, R. Modelling cancer progression using Mutual Hazard Networks. *Bioinformatics* **36**, 241–249 (2020).
38. Ramazzotti, D. *et al.* CAPRI: Efficient inference of cancer progression models from cross-sectional data. *Bioinformatics* **31**, 3016–3026 (2015).
39. Suppes, P. *A Probabilistic Theory of Causality*. (North-Holland Pub. Co., 1968).
40. Loohuis, L. O. *et al.* Inferring tree causal models of cancer progression with probability raising. *PLoS One* **9**, (2014).
41. Fenouille, N. *et al.* To be or not to be cell autonomous? Autophagy says both. *Essays Biochem.* **61**, 649–661 (2017).
42. Davies, K. Adaptive homeostasis. *Mol. Aspects Med.* **49**, (2016).
43. Bardelli, A. *et al.* Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer. *Cancer Discov.* **3**, 658–673 (2013).
44. Tlsty, T. D. & Coussens, L. M. Tumor stroma and regulation of cancer development. *Annu. Rev. Pathol.* **1**, 119–150 (2006).
45. Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta* **1805**, 105–117 (2010).
46. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**, 525–532 (1988).
47. Rozhok, A. I., Salstrom, J. L. & DeGregori, J. Stochastic modeling indicates that aging and somatic evolution in the hematopoietic system are driven by non-cell-autonomous processes. *Aging (Albany. NY).* **6**, 1033–1048 (2014).
48. Beerenwinkel, N., Schwarz, R. F., Gerstung, M. & Markowetz, F. Cancer evolution: mathematical models and computational inference. *Syst. Biol.* **64**, e1--e25 (2015).
49. Angaroni, F. *et al.* PMCE: efficient inference of expressive models of cancer evolution with high prognostic power. 1–14 (2014) doi:10.1101/008326.
50. Archetti, M. Heterogeneity and proliferation of invasive cancer subclones in game theory models of the {Warburg} effect. *Cell Prolif.* **48**, 259–269 (2015).
51. Sontag, L. & Axelrod, D. E. Evaluation of pathways for progression of heterogeneous breast tumors. *J. Theor. Biol.* **232**, 179–189 (2005).