A Logical Framework for Modelling Breast Cancer Progression

Joëlle Despeyroux[1], Amy Felty[2], Pietro Lio[3], and Carlos Olarte[4]

[1] INRIA and CNRS, I3S Laboratory, Sophia-Antipolis, France
[2] School of Electrical Engineering and Comp. Science, University of Ottawa, Canada
[3] Department of Computer Science and technology, University of Cambridge, UK
[4] ECT – Universidade Federal do Rio Grande do Norte, Brazil

Abstract. Data streams for a personalised breast cancer programme could include collections of image data, tumour genome sequencing, likely at the single cell level, and liquid biopsies (DNA and Circulating Tumour Cells (CTCs)). Although they are rich in information, the full power of these datasets will not be realised until we develop methods to model the cancer systems and conduct analyses that transect these streams. In addition to machine learning approaches, we believe that logical reasoning has the potential to provide help in clinical decision support systems for doctors. We develop a logical approach to modelling cancer progression, focusing on mutation analysis and CTCs, which include the appearance of driver mutations, the transformation of normal cells to cancer cells in the breast, their circulation in the blood, and their path to the bone. Our long term goal is to improve the prediction of survival of metastatic breast cancer patients. We model the behaviour of the CTCs as a transition system, and we use Linear Logic (LL) to reason about our model. We consider several important properties about CTCs and prove them in LL. In addition, we formalise our results in the Coq Proof Assistant, thus providing formal proofs of our model. We believe that our results provide a promising proof-of-principle and can be generalised to other cancer types and groups of driver mutations.

1 Introduction

Cancer is a complex evolutionary phenomenon, characterised by multiple levels of heterogeneity (inter-patient, intra-patient and intra-tumour), multiscale events (i.e. changes at intracellular, intercellular, tissue levels), multiomics variability (i.e. changes to chromatine, epigenetic and transcriptomic levels) that affect all aspects of clinical decisions and practice. The most remarkable phenomenon is the occurrence of numerous somatic mutations, of which only a subset contributes to cancer progression. The dynamic genetic diversity, coupled with epigenetic plasticity, within each individual cancer induces new genetic architectures and clonal evolutionary trajectories.

A striking feature of cancer is the subclonal genetic diversity, i.e., the presence of clonal succession and spatial segregation of subclones in primary sites

and metastases. At the root of the subclonal expansion there are developmentally regulated potentially self-renewing cells. After initiation, multiple subclones often coexist, signalling parallel evolution with no selective sweep or clear fitness advantage that becomes evident with therapies. Fitness calculation for each subclonal is difficult as subclones can form ecosystems and can cooperate through paracrine loops or interact through stromal, endothelial and inflammatory cells.

Tumour stages describe the progress of the tumour cells. One widely adopted approach is the American Joint Committee on Cancer (AJCC) tumour node metastasis (TNM) staging system. It classifies tumours with a combined stage between I-IV using three values: T gives the size of the primary tumour and extent of invasion, N describes if the tumour has spread to regional lymph nodes and M is indicative of distant metastasis. In terms of prognosis, Stage I patients have the best prognosis with 5 year survival rates (80-95%). The survival rates progressively worsen with each stage. Even with advances in targeted therapies, Stage IV patients have survival rates of just over two years. The integration of blood tests, biopsies, medical imaging, with genomic data have allowed the classification of many subtypes of cancers with striking differences of driver mutations and survival patterns. Therefore, the current data stream goals for a personalised breast cancer program should include the generation of tumour whole genome sequencing (DNA and RNA) for patients with breast cancer. Another data stream goal could focus on liquid biopsies. This will consist of data obtained from circulating tumour DNA (ctDNA) and single cell analysis. However, the full power of these datasets will not be realised until we leverage advanced statistical, mathematical and computational approaches to devise the needed procedures to conduct analyses that transect these streams.

There is a very rich body of biomedical statistics, machine learning and epidemiological literature for cancer data analysis which includes methods ranging from survival analysis, i.e., the effect of a risk factor or treatment with respect to cancer progression, analyses of co-alteration and mutual exclusivity patterns for genetic alterations, gene expression analyses, to network science algorithms (see e.g., [4, 5, 8, 17, 19, 26, 33]). For example, in survival modelling, the data is referred to as the time to event date and the objective is to analyse the time that passes before an event occurs due to one or more covariates [22, 23]. We believe that together with machine learning and biostatistics, there is a role for a logical approach in guiding optimal treatment decisions and in developing a risk stratification and monitoring tool to manage cancer. In this work, we focus on the use of a formal logical framework (as described below) to provide a reliable hypothesis-driven decision making system based on molecular data.

1.1 Formal Methods for Systems Biology

Computational systems biology provides a variety of methods for understanding the structure of biological systems and for studying their dynamics. In order to capture the *qualitative* nature of dynamics, approaches include Petri nets [9], π-calculus [31], bio-ambients [30], and rule-based modelling languages such as Biocham [18] and Kappa [11]. Molecular Logic [37] uses boolean logic gates to

define regulations in networks. One of the most successful approaches to model and analyse signal transduction networks, inside the cells, is Pathway Logic [35] (a system based on rewriting rules). Our approach, based on logical frameworks, can be used to model biological networks both inside and outside the cell.

The dynamics from a *quantitative* point of view can be captured by means of ordinary or stochastic differential equations. More recent approaches include hybrid Petri nets [21] and hybrid automata [1], piecewise linear equations [24], stochastic π-calculus [29], and rule-based languages with continuous/stochastic dynamics such as Kappa [11], Biocham [18], or BioNetGen [7].

One of the most common approaches to the formal verification of biological systems is model checking [10] that exhaustively enumerates all the states reachable by the system. In order to apply such a technique, the biological system should be encoded as a finite transition system and relevant system properties should be specified using temporal logic.

1.2 Logical Frameworks

In contrast to the aforementioned approaches, we encode both biological systems and temporal properties in logic, and prove that the properties can be derived from the system. This approach is new, only proposed in two previous works up to now [12,28]. In the present work, we choose discrete modelling, with temporal transition constraints. We believe that discrete modelling is crucial in systems biology since it allows taking into account some phenomena that have a very low chance of happening (and could thus be neglected by differential approaches), but which may have a strong impact on system behaviour.

We advocate the use of logical frameworks as an *unified and safe* approach to both specifying and analysing biological systems. *Logical Frameworks* are logics designed to formally study a variety of systems. The formal study of such systems means providing both formal models and proofs of properties of the systems. In the case where the logical systems are themselves logics, the logical framework enables the proofs of both *meta-theoretical* theorems (about the logic being formalised) and *object level* theorems (about the systems being encoded in the formalised logic). We shall use *Linear Logic* (LL) [20] as the intermediate logic, formalised in the Calculus of Constructions, which is a type theory implemented in the Coq Proof Assistant [6]. We note that the Coq system has been (partially) proven correct in itself, extensive meta-theoretical studies of LL are available in the Coq system (see e.g., [38]), and our encoding of biological systems is *adequate* (Section 3.1). This means that we prove that the formal model of the system correctly encodes the intended biological system. The approach is thus an unified approach, fully based on logic, and a safe approach, as each step is proved correct, as far as it can be.

We leverage our formalisation of LL in Coq to give a natural and direct characterisation of the state transformations of Circulating Tumour Cells (CTCs). For instance, a rule describing the evolution of a cell n, in a region r, from a healthy cell (with no mutations) to a cell that has acquired a mutation $\texttt{TGF}\beta$ can be modelled by the linear implication $\texttt{C}(n, r, []) \multimap \texttt{C}(n, r, [\texttt{TGF}\beta])$. This formula

describes the fact that a state where a cell $\mathtt{C}(n, r, [])$ is present can evolve into a state where $\mathtt{C}(n, r, [\mathtt{TGF}\beta])$ holds. More interestingly, the LL specification can be used to prove some desired properties of the system. For instance, it is possible to prove *reachability properties*, i.e., whether the system can reach a given state (Sect. 4.1) or even more abstract (meta-level) properties such as checking *all* the possible evolution paths the system can take under certain conditions (Sect. 4.2). Finally, we attain a certain degree of automatisation in our proofs which opens the possibility of testing recent proposed hypotheses in the literature.

Organisation. The rest of the paper is organised as follows. Section 2 describes the most relevant properties related to cancer mutations and CTCs which we believe are key factors driving the model dynamics. In Section 3 we recall the theory of LL and we specify in it the dynamics of CTCs. We prove correct our model in Theorem 1 by showing that transitions in the system are in one-to-one correspondence with logical steps. Reachability and meta-level properties of the system are proved in Section 4. Finally we conclude with a discussion on challenges and opportunities of logical frameworks in cancer studies. There is a companion appendix with the proofs of the results presented here. Moreover, all the proofs of the properties of our model were certified in Coq and are available at `http://subsell.logic.at/bio-CTC/`.

2 Tumour Cells in Metastatic Breast Cancer

In this section we first describe the mutations involved in cancer in general and then, we focus on the evolution of circulating tumour cells described in this work.

Cancer Mutations. Cancer mutations can be divided into drivers and non-drivers (or passengers). The accumulation of evidence of clonal heterogeneity and the observation of the arrival of drug resistance in clonal sub-populations suggest that mutations usually classified as non-drivers may have an important role in the fitness of the cancer cell and in the evolution and physiopathology of cancer. Similarly, mutations that alter the metabolism and the epigenetics may modify the fitness of the cancer cells. A meaningful way to identify drivers and passenger mutations is to use a statistical estimator of the impact of mutations such as FATHMM-MKL and a very large mutation database such as Cosmic (`http://cancer.sanger.ac.uk/`) [34].

The mutation process (causing tumour evolution) generates intra-tumour heterogeneity. The subsequent selection and Darwinian evolution (including immune escape) on intra-tumour heterogeneity is the key challenge in cancer medicine. Those clones that have progressed more than the others will have larger influence on patient survival. The amount of heterogeneity can vary between zero and up to over several thousand mutations, found to be heterogeneous within primary tumours or between primary and metastatic. The heterogeneity could be investigated through molecular biology techniques such as single cell sequencing, in situ PCR and could also be phenotypically classified using microscope image

analysis. In case of large heterogeneity we could assume that the survival of the patient strongly depends on the mutations of the most aggressive clones/cells.

Circulating Tumour Cells. We follow the study of the evolution of Circulating Tumour Cells in metastatic breast cancer in [4], where the authors use differential equations. This reference has an extensive discussion on the modelling choices, in particular concerning the driver mutations.

In [4] the probability for a cell in a duct in the breast to metastasise in the bone depends on the following mutational events:

1. A mutation in the `TGF`β pathways frees the cell from the surrounding cells.
2. A mutation in the `EPCAM` gene makes the cell rounded and free to divide. Then the cell enters the blood stream and becomes a circulating cancer cell.
3. In order to survive, this cell needs to over express the gene `CD47` that prevents attacks from the immune system.
4. Finally, there are two mutations that allow the circulating cancer cell to attach to the bone tissue and start the deadly cancer there: `CD44` and `MET`.

Hence, a cancer cell has four possible futures: (a) acquiring a driver mutation; (b) acquiring a passenger mutation which does not cause too much of a viability problem: it simply increases a sort of "counter to apoptosis"; (c), the new (i.e., last) mutation brings the cell to apoptosis; and (d), moving to the next compartement, or seeding in the bone.

The behaviour of the cells depends on the *compartments* the cells live in (here the breast, the blood and the bone), the other cells (i.e., the *environment*: the availability of food/oxygen or the pressure by the other cells), and the behaviour of the surface proteins (the *mutations*). In this work, we shall formalise the compartments and the mutations, and leave the formalisation of the environment to future work. Note that this environment plays a role only in the breast.

The *phenotype* of a cell is characterised by both the number of its mutations and its fitness. In biology the *fitness* is the capability of the cell to survive and produce offsprings. The cell's viability is particularly dependent on metabolic health and energy level. Most of the cell metabolic health depends on the accumulation of mutations that affect the production of enzymes involved in catalysing energetic expensive reactions and cell homeostasis. The fitness is particularly altered by the occurrence of driver mutations: each driver mutation provides the cell with additional fitness. Non-driver mutations, on the other hand, may accumulate in large numbers, and may affect cell stress response due to the altered metabolism and the competition with other neighboring cells [15]. Wet-lab tests for cell fitness and stress responsiveness have been recently developed, see for instance [3, 32, 36]. In our formalisation, the fitness will be a parameter of the cells. Physicians see the appearance of the cell (round, free, etc.), while biologists see the mutations. Our model can take both into account.

We extend the model in [4] with a few rules modelling DNA repair - of passenger mutations. These rules, only available for cells with `TGF`β or `EPCAM` mutations, (i.e. before `CD47` mutation), represent DNA repair by increasing the fitness by one. Note that this addition introduces *cycles* in our model. (i.e., it is possible for a cell to go back to a previous state).

3 Specification in Linear Logic

Linear Logic (LL) [20] is a *resource conscious* logic particularly well suited for describing state transition systems. LL has been successfully used to model such diverse systems as planning, Petri nets, process calculi, security protocols, multiset rewriting, graph traversal algorithms, and games. In this section we formalise in LL the behaviours of the Circulating Tumour Cells and we prove this formalisation correct. In each step of the formalisation, we shall give an intuitive description of the LL connectives and their proof rules that should suffice to understand our developments. The reader may find in [20] a more detailed account on the proof theory of LL.

LL is a substructural logic where there is an explicit control over the number of times a formula can be used in a proof. Formulas can be split into two sets: classical (those that can be used as many times as needed) or linear (those that are consumed after being used). Using a dyadic system for LL [2], sequents take the form $\Gamma \ ; \ \Delta \vdash G$ where G is the formula (goal) to be proved (examples in Section 4), Γ is the set of classical formulas and Δ is the multiset of linear formulas. We store in Γ the formulas representing the rules of the system and in Δ the atomic predicates representing the state of the system, namely:
- $\underline{\mathtt{C}(\mathbf{n}, \mathbf{c}, \mathbf{f}, \mathbf{lm})}$, denoting a cell n (a natural number used as an id), in a compartment c (breast, blood, or bone), with a phenotype given by a *fitness* degree $f \in 0 \,..\, 12$ and a list of driver mutations lm. The list of driver mutations lm is built up from mutations $\mathtt{TGF}\beta$, `EPCAM`, `CD47`, `CD44`, and/or `MET`, to which we add `seeded`, for the cells seeded in the bone. As $\mathtt{TGF}\beta$ is required before any further mutations, a list of mutations $[\mathtt{EPCAM}, \ldots]$ will by convention mean $[\mathtt{TGF}\beta, \mathtt{EPCAM}, \ldots]$;
- $\underline{\mathtt{A}(\mathbf{n})}$, representing the fact that the cell n has gone to apoptosis;
- and $\underline{\mathtt{T}(\mathbf{t})}$, stating that the current time-unit is t.

The linear implication $F_1 \multimap F_2$ models a state transformation where F_1 is consumed to later produce F_2. The proof rules are:

$$\frac{\Gamma; \Delta, F_1 \vdash F_2}{\Gamma; \Delta \vdash F_1 \multimap F_2} \multimap_R \qquad \frac{\Gamma; \Delta_1 \vdash \Downarrow F_1 \quad \Gamma; \Delta_2, \Downarrow F_2 \vdash G}{\Gamma; \Delta_1, \Delta_2, \Downarrow F_1 \multimap F_2 \vdash G} \multimap_L$$

In $\multimap_R$, the proof of $F_1 \multimap F_2$ requires the use of the resource F_1 to conclude F_2. This rule is invertible (i.e., the premise is provable if and only if the conclusion is provable). Hence, this rule belongs to the *negative* phase of the construction of a proof, where, without losing provability, we can apply all the invertible rules in any order. The rule $\multimap_L$ shows the resource awareness of the logic: part of the context (Δ_1) is used to prove F_1 and the remaining resources (Δ_2) must be used to prove G. The classical context Γ is not divided but copied in the premises. The rule $\multimap_L$ is non-invertible and then, it belongs to the *positive* phase. The notation $\Downarrow F_1 \multimap F_2$ means that we *decide* to *focus on* that formula and then, we have to keep working on the subformulas F_1 and F_2 (notation $\Downarrow F_1$ and $\Downarrow F_2$).

Each rule of the biological system is associated with a delay (see the terms of the form d_i in Fig. 1), which depends on the fitness parameter. Such parameters decrease marginally with passenger mutations and increase drastically with driver mutations. A typical rule in our model is then as follows:

$$\mathtt{rl}(br_{e0.1}) \stackrel{\text{def}}{=} \forall t, n. \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{20}(1)) \otimes \mathtt{C}(n, breast, 0, [\mathtt{EPCAM}])$$

This rule describes a cell acquiring passenger mutations. Its fitness is decreased by one in a time-delay $d_{20}(1)$ and its driver mutations remain unchanged. In this formula, we have introduced two new connectives: the universal quantifier allowing us to instantiate the same rule for any cell n and any time-unit t and the multiplicative conjunction $\otimes$ whose rules are:

$$\frac{\Gamma; \Delta_1 \vdash \Downarrow F_1 \quad \Gamma; \Delta_2 \vdash \Downarrow F_2}{\Gamma; \Delta_1, \Delta_2 \vdash \Downarrow F_1 \otimes F_2} \otimes_R \qquad \frac{\Gamma; \Delta, F_1, F_2 \vdash G}{\Gamma; \Delta, F_1 \otimes F_2 \vdash G} \otimes_L$$

The rule $\otimes_R$ belongs to the positive phase and it says that the proof of $F_1 \otimes F_2$ requires to split the linear context in order to prove both F_1 and F_2. The left rule belongs to the negative phase and the resource $F_1 \otimes F_2$ is simply transformed into two resources (F_1 and F_2).

Most of the rules in our model are parametric on the fitness degree. Hence, a rule of the form:

$$\boxed{\mathtt{rl}(br_{t1}) \stackrel{\text{def}}{=} \forall t, n.\ \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{TGF}\beta]) \multimap \mathtt{T}(t + d_{11}(f)) \otimes \mathtt{A}(n),\ f \in 0..2}$$

represents, in fact, three rules (one for each value of $f \in 0..2$). This family of rules describes three cases of apoptosis. In this particular example, any cell located in the breast, with fitness degree 0, 1, or 2 and list of mutations $[\mathtt{TGF}\beta]$ may go to apoptosis and the time needed for such a transition is $d_{11}(f)$. Note that $d(\cdot)$ is a function that depends on f. If such $d(\cdot)$ does not depend on f, we shall simply write d instead of $d()$.

A typical rule describing a cell acquiring a driver mutation is:

$$\boxed{\mathtt{rl}(br_{t2.1}) \stackrel{\text{def}}{=} \forall t, n.\ \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, [\mathtt{TGF}\beta]) \multimap \mathtt{T}(t + d_{12}) \otimes \mathtt{C}(n, breast, 2, [\mathtt{EPCAM}])}$$

This rule says that a cell in the breast with a fitness degree $f = 1$ may acquire a new mutation (EPCAM), which increases its fitness by 1.

Another kind of rule describes a cell moving from one compartment to the next. The following rule describes an intravasating CTC:

$$\boxed{\mathtt{rl}(br_{e2}) \stackrel{\text{def}}{=} \forall t, n.\ \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{22}(f)) \otimes \mathtt{C}(n, blood, 2, [\mathtt{EPCAM}]),\ f \in 1..3}$$

Finally, a last kind of rule describes a DNA repair of passenger mutations:

$$\boxed{\mathtt{rl}(br_{e0r}) \stackrel{\text{def}}{=} \forall t, n. \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{20r}(f)) \otimes \mathtt{C}(n, breast, f + 1, [\mathtt{EPCAM}]),\ f \in 1..2}$$

The complete set of rules is in Fig. 1. For the sake of readability, we omit the universal quantification on t and n in the formulas. We shall use system to denote the set of rules and then, sequents take the form:

$$\mathtt{system}\ ;\ \mathtt{T}(t), \mathtt{C}(\cdot), \cdots, \mathtt{C}(\cdot) \vdash G$$

where G is a property to be proved (Sect. 4). Let s be a multiset of the form $\{\mathtt{C}_1(n_1, c_1, f_1, lm_1), ..., \mathtt{C}_n(n_n, c_n, f_n, lm_n), \mathtt{A}(n'_1), ..., \mathtt{A}(n'_k)\}$ representing the state of different cells. The multiset $\{\mathtt{T}(t), \mathtt{C}_1(n_1, c_1, f_1, lm_1), ..., \mathtt{C}_n(n_n, c_n, f_n, lm_n), \mathtt{A}(n'_1), ..., \mathtt{A}(n'_k)\}$ of atomic formulas, formalising the state of the system at time-unit t, is denoted as $[\![s]\!]_t$. Observe that a cell that has gone to apoptosis cannot evolve any further (as no rule has a $\mathtt{A}(n)$ on the left hand side).

We note that our rules are asynchronous: only one rule can be fired at a time. As in Biocham, we choose an asynchronous semantics in order to eliminate the risk of affecting fundamental biological phenomena such as the masking of a relation by another one and the consequent inhibition/activation of biological processes. Finally, we note that the delays depend on the fitness and more

In the breast	
$\mathtt{rl}(br_0)$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, []) \multimap \mathtt{T}(t + d_{00}) \otimes \mathtt{C}(n, breast, 0, [])$
$\mathtt{rl}(br_1)$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, []) \multimap \mathtt{T}(t + d_{01}(f)) \otimes \mathtt{A}(n)\ f \in 0 \mathinner{..} 1$
$\mathtt{rl}(br_2)$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, []) \multimap \mathtt{T}(t + d_{02}) \otimes \mathtt{C}(n, breast, 1, [\mathtt{TGF}\beta])$
$\mathtt{rl}(br_{t0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, [\mathtt{TGF}\beta]) \multimap \mathtt{T}(t + d_{10}) \otimes \mathtt{C}(n, breast, 0, [\mathtt{TGF}\beta])$
$\mathtt{rl}(br_{t0r})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, [\mathtt{TGF}\beta]) \multimap \mathtt{T}(t + d_{10r}) \otimes \mathtt{C}(n, breast, 2, [\mathtt{TGF}\beta])$
$\mathtt{rl}(br_{t1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{TGF}\beta]) \multimap \mathtt{T}(t + d_{11}(f)) \otimes \mathtt{A}(n),\ f \in 0 \mathinner{..} 2$
$\mathtt{rl}(br_{t2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{TGF}\beta]) \multimap \mathtt{T}(t + d_{12}) \otimes \mathtt{C}(n, breast, f + 1, [\mathtt{EPCAM}])\ f \in 1 \mathinner{..} 2$
$\mathtt{rl}(br_{e0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{20}(f)) \otimes \mathtt{C}(n, breast, f - 1, [\mathtt{EPCAM}]),\ f \in 1 \mathinner{..} 3$
$\mathtt{rl}(br_{e0r})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{20r}(f)) \otimes \mathtt{C}(n, breast, f + 1, [\mathtt{EPCAM}]),\ f \in 1 \mathinner{..} 2$
$\mathtt{rl}(br_{e1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{21}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 3$
$\mathtt{rl}(br_{e2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, breast, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{22}(f)) \otimes \mathtt{C}(n, blood, f + 1, [\mathtt{EPCAM}]), f \in 1 \mathinner{..} 3$
In the blood	
$\mathtt{rl}(bl_{e0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{30}(f)) \otimes \mathtt{C}(n, blood, f - 1, [\mathtt{EPCAM}]), f \in 1 \mathinner{..} 4$
$\mathtt{rl}(bl_{e0r})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{30r}(f)) \otimes \mathtt{C}(n, blood, f + 1, [\mathtt{EPCAM}]), f \in 1 \mathinner{..} 3$
$\mathtt{rl}(bl_{e1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{31}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 4$
$\mathtt{rl}(bl_{e2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}]) \multimap \mathtt{T}(t + d_{32}(f)) \otimes \mathtt{C}(n, blood, f + 2, [\mathtt{EPCAM}, \mathtt{CD47}]), f \in 1 \mathinner{..} 4$
$\mathtt{rl}(bl_{ec0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}, \mathtt{CD47}]) \multimap \mathtt{T}(t + d_{40}(f)) \otimes \mathtt{C}(n, blood, f - 1, [\mathtt{EPCAM}, \mathtt{CD47}]), f \in 1 \mathinner{..} 6$
$\mathtt{rl}(bl_{ec1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}, \mathtt{CD47}]) \multimap \mathtt{T}(t + d_{41}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 6$
$\mathtt{rl}(bl_{ec2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}, \mathtt{CD47}]) \multimap \mathtt{T}(t + d_{42}(f)) \otimes \mathtt{C}(n, blood, f + 2, [\mathtt{EPCDCD}]), f \in 1 \mathinner{..} 6$
$\mathtt{rl}(bl_{ec3})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCAM}, \mathtt{CD47}]) \multimap \mathtt{T}(t + d_{43}(f)) \otimes \mathtt{C}(n, blood, f + 2, [\mathtt{EPCDME}]), f \in 1 \mathinner{..} 6$
$\mathtt{rl}(bl_{ecc0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDCD}]) \multimap \mathtt{T}(t + d_{50}(f)) \otimes \mathtt{C}(n, blood, f - 1, [\mathtt{EPCDCD}]), f \in 1 \mathinner{..} 6$
$\mathtt{rl}(bl_{ecc1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDCD}]) \multimap \mathtt{T}(t + d_{51}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 8$
$\mathtt{rl}(bl_{ecc2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDCD}]) \multimap \mathtt{T}(t + d_{52}(f)) \otimes \mathtt{C}(n, blood, f + 2, [\mathtt{EPCDCDME}]), f \in 1 \mathinner{..} 8$
$\mathtt{rl}(bl_{ecm0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDME}]) \multimap \mathtt{T}(t + d_{60}(f)) \otimes \mathtt{C}(n, blood, f - 1, [\mathtt{EPCDME}]), f \in 1 \mathinner{..} 8$
$\mathtt{rl}(bl_{ecm1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDME}]) \multimap \mathtt{T}(t + d_{61}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 8$
$\mathtt{rl}(bl_{ecm2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDME}]) \multimap \mathtt{T}(t + d_{62}(f)) \otimes \mathtt{C}(n, blood, f + 2, [\mathtt{EPCDCDME}]), f \in 1 \mathinner{..} 8$
$\mathtt{rl}(bl_{eccm0})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDCDME}]) \multimap \mathtt{T}(t + d_{70}(f)) \otimes \mathtt{C}(n, blood, f - 1, [\mathtt{EPCDCDME}]), f \in 1 \mathinner{..} 10$
$\mathtt{rl}(bl_{eccm1})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDCDME}]) \multimap \mathtt{T}(t + d_{71}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 10$
$\mathtt{rl}(bl_{eccm2})$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, blood, f, [\mathtt{EPCDCDME}]) \multimap \mathtt{T}(t + d_{72}) \otimes \mathtt{C}(n, bone, f + 1, [\mathtt{EPCDCDME}]), f \in 1 \mathinner{..} 10$
In the bone	
$\mathtt{rl}(bo_0)$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, bone, f, [\mathtt{EPCDCDME}]) \multimap \mathtt{T}(t + d_{80}(f)) \otimes \mathtt{C}(n, bone, f - 1, [\mathtt{EPCDCDME}]), f \in 1 \mathinner{..} 11$
$\mathtt{rl}(bo_1)$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, bone, f, [\mathtt{EPCDCDME}]) \multimap \mathtt{T}(t + d_{81}(f)) \otimes \mathtt{A}(n), f \in 0 \mathinner{..} 11$
$\mathtt{rl}(bo_2)$	$\stackrel{\text{def}}{=} \mathtt{T}(t) \otimes \mathtt{C}(n, bone, f, [\mathtt{EPCDCDME}]) \multimap \mathtt{T}(t + d_{82}(f)) \otimes \mathtt{C}(n, bone, f + 1, [\mathtt{EPCDCDME}, \mathtt{seeded}]), f \in 1 \mathinner{..} 11$

Fig. 1. Complete set of rules. Variables t and n are universally quantified. EPCDCDME, EPCDCD and EPCDME are shorthand, respectively, for the list of mutations [EPCAM, CD47, CD44, MET], [EPCAM, CD47, CD44] and [EPCAM, CD47, MET].

accurate values can be found using data. DNA mutational processes have been succesfully modeled as compound poisson processes (see for instance [16]). Delays could be seen as events waiting times which are well-known measures for Poisson processes (see for instance [25]). In our model, delays are (uninterpreted) logical constants that can be later tuned when experimental results are available. The proofs presented here remain the same regardless such values.

3.1 Adequacy

In this section, we prove the adequacy of our encoding: a single transition in the state of the system corresponds, exactly, to a complete focused phase (a positive phase followed by a negative phase) in focused LL [2]. Focusing organises proofs in phases where a negative phase introduces all the invertible rules. In a positive

phase, we choose one of the formulas (notation $\Downarrow F$) whose principal connective has a non-invertible rule. Introducing such a connective is a *decision* in the proof search procedure and then, the order in which we apply them may lead to a proof or not. Proofs must finish in the positive phase with an initial rule:

$$\frac{}{\Gamma; p \vdash \Downarrow p}\ I \qquad \frac{}{\Gamma, p; \cdot \vdash \Downarrow p}\ I$$

where p is an atomic formula (e.g, $\mathtt{C}_1(n, c, f, lm)$). Note that the proof ends when p is the only atom in the linear context or when the linear context is empty and p is in the classical context.

Our results are based on the following observations:

1. It is not possible to focus on the formulas resulting from $[\![s]\!]_t$ since those are atoms. Hence, no focus step can start by focusing on those formulas.
2. Once we focus on one of the formulas in the classical context `system` (modelling the rules of the system), what we observe is that one of the $\mathtt{C}(n, c, f, lm)$ formulas is consumed as well as the predicate $\mathtt{T}(t)$. The focus phase ends by producing the needed $\mathtt{C}(n, c', f', lm')$ and $\mathtt{T}(t')$ atoms (or $\mathtt{A}(n)$ in the case of apoptosis rules). This means that focused derivations are in one-to-one correspondence with steps in the system.

In the following theorem, we shall use the notation $s \xrightarrow{(r,d)} s'$ to denote that the system may evolve from state s to state s' by applying the rule r that takes d time-units. Hence, $S_s = \{(s', r, d) \mid s \xrightarrow{(r,d)} s'\}$ represents the set of possible transitions starting from s. We shall show that all transitions in S_s match exactly one focused derivation of the encoded system. More precisely (proof in Appendix)

Theorem 1 (Adequacy). *Let s be a state and $S_s = \{(s', r, d) \mid s \xrightarrow{(r,d)} s'\}$. Then, $(s', r, d) \in S_s$ iff focusing on the encoding of r leads to the following derivation.*

$$\frac{\mathtt{system}\ ;\ [\![s']\!]_{t+d} \vdash G}{\mathtt{system}\ ;\ [\![s]\!]_t \vdash G}$$

The following corollaries are immediate consequences of Theorem 1.

Corollary 1 (Adequacy). *Let s and s' be two states. Then $s \xrightarrow{(r,d)} s'$ iff the sequent $\mathtt{system}; \cdot \vdash [\![s]\!]_t \multimap [\![s]\!]_{t+d}$ is provable.*

The above results allow us to use the whole positive-negative phase as macro rules in the logical system. Hence, during proofs, we shall abuse notation and we shall use, e.g., $\mathtt{rl}(bo_2)$ as a logical inference rule. Formally, we can show that the corresponding logical rule is admissible in the system, i.e., if the premise is provable then the conclusion is also provable.

Corollary 2 (Macro rules). *Assume that $s \xrightarrow{(r,d)} s'$. Then, the following macro rule is admissible:*

$$\frac{\mathtt{system}\ ;\ \Delta, [\![s']\!]_{t+d} \vdash G}{\mathtt{system}\ ;\ \Delta, [\![s]\!]_t \vdash G}\ r$$

4 Verifying Properties of the Model

The goal of this section is twofold: testing our rules, but also testing some hypothesis of our model—as these are recent proposals in the literature. We shall detail some of the proofs here. The others can be found in the proof scripts and the documentation of our Coq formalisation (`http://subsell.logic.at/bio-CTC/`).

4.1 Reachability and Existence of Cycle Properties

Recall that a *Circulating Tumour Cell (CTC)* is a cancer cell in the blood, i.e., a cell $\mathtt{C}(n, \mathit{blood}, f, m)$. An *extravasating CTC* is a CTC that has reached the bone, i.e., a cell of the shape $\mathtt{C}(n, \mathit{bone}, f, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])$. A first property of interest might be the following one: "is it possible for a CTC, with mutations $[\mathtt{EPCAM}, \mathtt{CD47}]$ and fitness 3, to become an extravasating CTC with fitness 8? What is the time delay for such a transition?" This is formalised as follows:

Property 1. The following sequent is provable:

$$\mathtt{system}\ ;\ . \vdash \forall n, t.\ \mathtt{T}(t) \otimes \mathtt{C}(n, \mathit{blood}, 3, [\mathtt{EPCAM}, \mathtt{CD47}]) \multimap \exists d.\ \mathtt{T}(t+d) \otimes \mathtt{C}(n, \mathit{bone}, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])$$

In our Coq formalisation, we have implemented several tactics (e.g., `solveF` and `applyRule` used below) to automate the process of proving properties and make the resulting scripts compact and clear. This should ease the testing/proving of new hypotheses in our model. For instance, the proof of the previous property is as follows (F below is the formula in Property 1):

```
Lemma Property1: forall n t, |- System ; F
Proof with solveF . (* solves the "trivial" goals in a focused proof *)
  intros. (* introducing the quantified variables n and t *)
  applyRule (blec2 3). (* application of macro rules -- corollary 2-- *)
  applyRule (blecc2 5).
  applyRule (bleccm2 7).
  eapply tri_dec1 ...    (* decision rule, focusing on the goal *)
  eapply tri_tensor ...  (* tensor *)
  eapply tri_ex with (t:= (d72 7) s+ (d52 5) s+ (d42 3) s+ (Cte t)) ...  (* existential quantif. *)
  eapply Init1...  (* initial rule *)
  eapply Init1...  (* initial rule *)
Qed.
```

The reader may compare the steps in the script above with the proof (by hand) of Property 1 in Appendix B.

Our next property is the following: "what is the time delay for a CTC with mutations $[\mathtt{EPCAM}, \mathtt{CD47}]$ and fitness 3 or 4 to become an extravasating CTC with fitness between 6 and 9?"

Property 2. The following sequent is provable:

$$\begin{aligned}\mathtt{system}\ ;\ \vdash \forall n, t.\ \mathtt{T}(t) \otimes (\mathtt{C}(n, \mathit{blood}, 3, [\mathtt{EPCAM}, \mathtt{CD47}]) \oplus \mathtt{C}(n, \mathit{blood}, 4, [\mathtt{EPCAM}, \mathtt{CD47}])) \multimap \\ \exists t_d.\ \mathtt{T}(t+t_d) \otimes \\ (\mathtt{C}(n, \mathit{bone}, 6, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus \mathtt{C}(n, \mathit{bone}, 7, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus \\ \mathtt{C}(n, \mathit{bone}, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus \mathtt{C}(n, \mathit{bone}, 9, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]))\end{aligned}$$

Due to the $\oplus$ connective (additive conjunction), the proof of this property entails two proof obligations (see the details in Appendix B): the case

when the fitness is 3 and the case when the fitness is 4. The proof of the first case reveals that the rules $\mathtt{rl}(bl_{ec0})$ and $\mathtt{rl}(bl_{ecc0})$ could be used zero or more times—as long as the fitness remains positive. In the end, when the state $\mathtt{C}(n, bone, 6, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])$ is reached, we also obtain in our proof the delay needed ($t_d = d_{40}(3) + d_{42}(2) + d_{50}(4) + d_{52}(3) + d_{72}(5)$) to reach that state. The second proof obligation can be discharged by considering several paths, depending on the order of mutations CD44 and MET and the eventually many passenger mutations (rules $bl_{ec0}, bl_{ecc0}, bl_{ecc0}$, and bl_{ecm0}). We illustrate some of those paths in Appendix B. Finally, we note that along with the time delay t_d we are looking for, the proof provides also the fitness of the extravasating CTC.

Existence of Cycle. Rules for passenger mutations decrease the fitness of the cell by one, while rules for DNA repair increase the fitness. Hence, we may observe loops and oscillations in our model. This can be exemplified in the following property: "a cell in the breast, with mutation [EPCAM], might have its fitness oscillating from 1 to 2 and back."

Property 3. The following sequents are provable:

$\mathtt{system}\ ;\ . \vdash \forall n, t.\ \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 1, [\mathtt{EPCAM}]) \multimap \exists d.\ \mathtt{T}(t+d) \otimes \mathtt{C}(n, breast, 2, [\mathtt{EPCAM}])$ and

$\mathtt{system}\ ;\ . \vdash \forall n, t.\ \mathtt{T}(t) \otimes \mathtt{C}(n, breast, 2, [\mathtt{EPCAM}]) \multimap \exists d.\ \mathtt{T}(t+d) \otimes \mathtt{C}(n, breast, 1, [\mathtt{EPCAM}])$

4.2 Meta-level Properties

In a first experiment on using LL for biology on the computer [12], we defined the set of biological rules as an inductive type in Coq, and proved some of their properties by induction on the set of fireable rules. Here, we choose a different approach. We have defined the biological rules by formulas in LL, and we use focusing, along with adequacy, to look for the fireable rules at a given state. Properties whose proofs need meta-reasoning will be formalised at the level of derivations. In this section, we give two examples of these.

Let us first give a simple property concerning apoptosis, namely: "any cell having a null fitness must go to apoptosis".

Property 4. Let Δ be a multiset of atoms of the form $\mathtt{C}(\cdot)$. Then, in any derivation of the form

$$\dfrac{\mathtt{system}; \Delta, \mathtt{T}(t+d), St \vdash G}{\mathtt{system}; \Delta, \mathtt{T}(t), \mathtt{C}(n, c, 0, m) \vdash G}\ \mathtt{rl}(\cdot)$$

we have $St = \mathtt{A}(n)$.

In our Coq formalisation, the above property can be discharged with few lines of code:

```
Lemma Property4: forall  n t c lm , F.
Proof by solveF .
 intros H. (* the first sequent is assumed to be provable *)
 apply FocusOnlyTheory in H;auto.  (* The proof H must start by focusing on one of the rules *)
 destruct H as [R] ;destruct H.
 repeat first [ CaseRule | DecomposeRule; FindUnification | eauto ].
Qed.
```

The `FocusOnlyTheory` lemma says that the proof of the sequent must start by focusing on one of the formulas in `System`. The `destruct` tactic simplifies the hypotheses after the use of lemma `FocusOnlyTheory`. The interesting part is the last line of the script. The `CaseRule` tactic tests each of the rules of the system. Then, `DecomposeRule; FindUnification` decomposes (positive-negative phase) the application of the rule. Finally, `eauto` proves the desired goal after the application of the rule. This is a very general scheme, where we do case analysis on all possible rules. Some of them cannot be fired in the current state and then, the proof follows by contradiction. In the rest of the cases, the `eauto` tactic is able to conclude the goal.

The following property states one of the key properties of our model: "any cell in the blood, with mutations including `CD47`, has four possible evolutions:

1. acquiring passenger mutations: its fitness decreases by one and the driver mutations remain unchanged;
2. going to apoptosis;
3. acquiring a driver mutation: its fitness increases by two;
4. moving to the bone: its fitness increases by one and the driver mutations ([`EPCAM`, `CD47`, `CD44`, `MET`]) remain unchanged."

Property 5. Let Δ be a multiset of atoms of the form $\mathtt{C}(\cdot)$. Then, in any derivation of the form

$$\frac{\mathtt{system}; \Delta, \mathtt{T}(t + t_d), St \vdash G}{\mathtt{system}; \Delta, \mathtt{T}(t), \mathtt{C}(n, blood, f, m) \vdash G}\ \mathtt{rl}(\cdot)$$

with m containing `CD47`, it must be the case that

1. either $St = \mathtt{C}(n, blood, f - 1, m)$,
2. or $St = \mathtt{A}(n)$,
3. or $St = \mathtt{C}(n, blood, f+2, m')$ with m' being as m plus an additional mutation,
4. or $St = \mathtt{C}(n, bone, f + 1, m)$ with $m = [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]$.

The proof follows the same rationale as in Property 4. In fact, the proof in Coq is exactly the same but, in the last line, we have

```
repeat (first [ CaseRule  | DecomposeRule; FindUnification | SolveGoal]) .
```

where `SolveGoal` (instead of `eauto`) is able to finish the resulting cases.

5 Concluding Remarks

Our goal is to study cancer progression, aiming at a better understanding of it, and, in the long term, help in finding, and testing, new targeted drugs, a priori much more efficient than most of the drugs used so far. This paper describes the use of linear logic in modelling the multi compartment role of driver mutations in breast cancer. This work is innovative but also proof-of-principle. It can clearly be generalised to other cancer types where driver mutations are known. It also makes evident the capability of this logical approach to integrate

different types of data and output a diagnosis with higher interpretability than many currently fashionable machine learning methods such as deep learning. Note however that building a system for cancer/disease diagnosis and therapy prognosis would require both automatic proof search and taking the size of the tumour into account. Also note that, although all the properties considered so far only deal with the evolution of one single cell, our approach allows us to consider a state with many cells. We believe that the paper and the rich sections in the online supplementary material would become an important resource for other similar studies.

For example we can believe there is a complementary of our work with respect to mathematical models such as [4] that include survival data and quantitative results. Logic allows to model the evolution of cells across scales and compartments, while ODEs require parameter estimation (qualitative vs quantitative).

While temporal logics have been very successful in practice with efficient model checking tools, these logics do not enjoy standard proof theory. In contrast, LL has a very traditional proof theoretic pedigree: it is naturally presented as a sequent calculus that enjoys cut-elimination and focusing. A further advantage of our approach wrt model checking is that it provides a unified framework to encode both transition rules and (both statements and proofs of) temporal properties. Observe also that we do not need to build the set of states of the transition system. We view model checking as a useful first step before proofs: testing the model before trying to prove properties of it. The interested reader can find a detailed comparison of the approaches in [12]. See also [14] for an adequate encoding of Temporal Logic in LL.

In order to describe *constrained* transitions systems (by timed or spatial restrictions for example), previous work have proposed different extensions of Linear Logic, e.g., HyLL (a modal extension of ILL) [13] and SELL$^{\pi}$ (LL with with quantifiers on subexponentials). See [14] for a formal comparison of the two logics, and an adequate encoding of Temporal Logic in LL extended with fixpoints. These Logical Frameworks were then used to specify and analyse small biology systems in two initial experiments: [12] in HyLL and [28] in SELL$^{\pi}$. In the present work, we chose to use pure LL, and define a predicate to encode time. This approach avoids the extra complication of copying the "unused" information to the next time-unit (HyLL world or SELL subexponential), thus benefiting from the usual compositional nature of the logic.

Future perspectives. The ongoing revolution in AI is accelerating the development of software that enables computers to perform intelligent clinical and medical tasks. Machine learning algorithms find hidden patterns in data, classify and associate similar patients/diseases/drugs based on common features (e.g., the IBM Watson system which is used to analyse genomic and cancer data). Future challenges in medicine include understanding bias in data collection (and also in doctor's experience) and fostering the ability to integrate evidence from heterogeneous datasets, from different omics and clinical data, from several lines of independent data. We believe that machine learning could satisfy well these needs and that there is also a need to develop methods that offer a hypothesis-

driven approach, so that doctors do not feel that they are going to be replaced. Such methods could provide them with a personalised and easily interpretable clinical support decision-making tool that could perform a synthesis of qualitative and quantitative multi-modal evidence. Examples of decision trees used in current practice for breast cancer diagnosis can be found at pages 598–603 of [27]. Our logical approach, although focused on driver mutations, goes in such a direction and could be used with continuous and discrete mixed variables. This information could be obtainable through single cell experiments on cancer biopsies (although with large variance), which is now at the stage of passing from basic science to clinical protocols. Machine learning could analyse cancer mutation patterns and feed our logic approach with this information that could be integrated with other rules such as changes on the metabolic networks or on epigenetics. Other rules could be derived from other levels of cancer clinical investigation such as from image data (changes in fMRI, CT-scans and microscopy samples), blood analyses (identification and counts of circulating cancer cells) and other types of medical observations. The long term plan is to build a portable resource that facilitates diagnostic and therapeutic decision making and promotes a cost-effective personalised patient workup. This would represent a new paradigm in personalised and precision cancer treatment which integrates multi-modality analyses and clinical characteristics in a near-real time manner, improving clinical management of cancer. Finally we believe that logical approaches could improve the harmonisation and standardisation of the reporting and interpretation of clinically relevant data.

Acknowledgements. We thank the anonymous reviewers for their valuable comments on an earlier draft of this paper. The work of Olarte was supported by CNPq, the project FWF START Y544-N23 and by CAPES, Colciencias, and INRIA via the STIC AmSud project EPIC (Proc. No 88881.117603/2016-01).

References

1. Alur, R., Belta, C., Ivaničić, F., Kumar, V., Mintz, M., Pappas, G.J., Rubin, H., Schug, J.: Hybrid modeling and simulation of biomolecular networks. In: Proc of the HSCC. LNCS, vol. 2034, pp. 19–32. Springer (2001)
2. Andreoli, J.: Logic programming with focusing proofs in linear logic. J. Log. Comput. **2**(3), 297–347 (1992)
3. Antczak, C., Mahida, J., Singh, C., Calder, P., Djaballah, H.: A high content assay to assess cellular fitness. Combinatorial Chemistry & High Throughput Screening **17**(1), 12–24 (jan 2014).
4. Ascolani, G., Occhipinti, A., Lio, P.: Modeling circulating tumour cells for personalised survival prediction in metastatic breast cancer. PLoS Computational Biology. **11**(5) (2015)
5. Bellomo, N., Preziosi, L.: Modelling and mathematical problems related to tumor evolution and its interaction with the immune system. Mathematical and Computer Modelling **32**(3), 413–452 (2000)
6. Bertot, Y., Castéran, P.: Interactive Theorem Proving and Program Development. Coq'Art: The Calculus of Inductive Constructions. Springer (2004)

7. Blinov, M.L., Faeder, J.R., Goldstein, B., Hlavacek, W.S.: BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics **20**(17), 3289–3291 (2004)
8. Caravagna, G., Giarratano, Y., Ramazzotti, D., Tomlinson, I., Graham, T.A., Sanguinetti, G., Sottoriva, A.: Detecting repeated cancer evolution from multi-region tumor sequencing data. Nature Methods **15**(9), 707–714 (2018).
9. Chaouiya, C., Naldi, A., Remy, E., Thieffry, D.: Petri net representation of multi-valued logical regulatory graphs. Natural Computing **10**(2), 727–750 (2011).
10. Clarke, E.M., Henzinger, T.A., Veith, V., Bloem, R.: Handbook of Model Checking. Springer (2018)
11. Danos, V., Laneve, C.: Formal molecular biology. Theoretical Computer Science **325**(1), 69–110 (2004)
12. De Maria, E., Despeyroux, J., Felty, A.: A logical framework for systems biology. In: Springer (ed.) Proc. of FMMB. LNCS, vol. 8738, pp. 136–155 (2014)
13. Despeyroux, J., Chaudhuri, K.: A hybrid linear logic for constrained transition systems. In: Post-Proceedings of TYPES'13. Leibniz Intl. Proceedings in Informatics, vol. 26, pp. 150–168. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2014)
14. Despeyroux, J., Olarte, C., Pimentel, E.: Hybrid and subexponential linear logics. Electronic Notes in Theoretical Computer Science **332**, 95–111 (2017)
15. Gregorio, A.D., Bowling, S., Rodriguez, T.A.: Cell competition and its role in the regulation of cell fitness from development to cancer. Developmental Cell **38**(6), 621–634 (sep 2016).
16. Ding, J., Trippa, L., Zhong, X., Parmigiani, G.: Hierarchical bayesian analysis of somatic mutation data in cancer. Ann. Appl. Stat. **7**(2), 883–903 (06 2013).
17. Enderling, H., Chaplain, M.A., Anderson, A.R., Vaidya, J.S.: A mathematical model of breast cancer development, local treatment and recurrence. Journal of theoretical biology **246**(2), 245–259 (2007)
18. Fages, F., Soliman, S., Chabrier-Rivier, N.: Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM. Journal of Biological Physics and Chemistry **4**(2), 64–73 (2004)
19. Gavaghan, D., Brady, J.M., Behrenbruch, C., Highnam, R., Maini, P.: Breast cancer: Modelling and detection. Computational and Mathematical Methods in Medicine **4**(1), 3–20 (2002)
20. Girard, J.Y.: Linear logic. Theoretical Computer Science **50**, 1–102 (1987)
21. Hofestädt, R., Thelen, S.: Quantitative modeling of biochemical networks. In Silico Biology, vol. 1, pp. 39–53. IOS Press (1998)
22. Iuliano, A., Occhipinti, A., Angelini, C., Feis, I.D., Lió, P.: Cancer markers selection using network-based cox regression: A methodological and computational practice. Frontiers in Physiology **7** (2016)
23. Iuliano, A., Occhipinti, A., Angelini, C., Feis, I.D., Liò, P.: Combining pathway identification and breast cancer survival prediction via screening-network methods. Frontiers in Genetics **9** (2018).
24. de Jong, H., Gouzé, J.L., Hernandez, C., Page, M., Sari, T., Geiselmann, J.: Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bulletin of Mathematical Biology **66**(2), 301–340 (2004)
25. Kingman, J.F.C.: Poisson processes, Oxford Studies in Probability, vol. 3. The Clarendon Press Oxford University Press, New York (1993), Oxford Science Publications
26. Knutsdottir, H., Palsson, E., Edelstein-Keshet, L.: Mathematical model of macrophage-facilitated breast cancer cells invasion. J. of theoretical biology (2014)

27. Mushlin, S.B., Greene, H.L.: Decision making in medicine: An algorithmic approach, 3e (clinical decision making series) 3rd edition (2009)
28. Olarte, C., Chiarugi, D., Falaschi, M., Hermith, D.: A proof theoretic view of spatial and temporal dependencies in biochemical systems. Theor. Comput. Sci. **641**, 25–42 (2016)
29. Phillips, A., Cardelli, L.: A correct abstract machine for the stochastic pi-calculus. In: BioConcur: Workshop on Concurrent Models in Molecular Biology (2004)
30. Regev, A., Panina, E.M., Silverman, W., Cardelli, L., Shapiro, E.: Bioambients: An abstraction for biological compartments. Theoretical Computer Science **325**(1), 141–167 (2004)
31. Regev, A., Silverman, W., Shapiro, E.Y.: Representation and simulation of biochemical processes using the π-calculus process algebra. In: Proceedings of the 6th Pacific Symposium on Biocomputing. pp. 459–470 (2001)
32. Rogers, Z.N., McFarland, C.D., Winters, I.P., Naranjo, S., Chuang, C.H., Petrov, D., Winslow, M.M.: A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nature Methods **14**(7), 737–742 (may 2017).
33. Savage, N.: Computing cancer software models of complex tissues and disease are yielding a better understanding of cancer and suggesting potential treatments. Nature **491**, s62–s63 (2012)
34. Shihab, H.A., Rogers, M.F., Gough, J., Mort, M.E., Cooper, D.N., Day, I.N.M., Gaunt, T.R., Campbell, C.: An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics **31**(10), 1536–1543 (2015)
35. Talcott, C., Dill, D.: Multiple representations of biological processes. Transactions on Computable Systems Biology pp. 221–245 (2006)
36. Venkataram, S., Dunn, B., Li, Y., Agarwala, A., Chang, J., Ebel, E.R., Geiler-Samerotte, K., Hérissant, L., Blundell, J.R., Levy, S.F., Fisher, D.S., Sherlock, G., Petrov, D.A.: Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. Cell **166**(6), 1585–1596.e22 (sep 2016).
37. Wynn, M.L., Consul, N., Merajver, S.D., Schnell, S.: Logic-based models in systems biology: a predictive and parameter-free network analysis method. Integrative Biology **4**(11), 1323 (2012).
38. Xavier, B., Olarte, C., Reis, G., Nigam, V.: Mechanizing linear logic in Coq. Electr. Notes Theor. Comput. Sci. 338: 219-236 (2018)

Appendix

A Proof of the adequacy results

Theorem 1: Adequacy. Let s be a state and $S_s = \{(s', r, d) \mid s \xrightarrow{(r,d)} s'\}$. Then, $(s', r, d) \in S_s$ iff focusing on the encoding of r leads to the following derivation.

$$\frac{\texttt{system} \,;\, [\![s']\!]_{t+d} \vdash G}{\texttt{system} \,;\, [\![s]\!]_t \vdash G}$$

Proof. The encoding of the rule r is a bipole [2] (i.e., a formula that, being focused, will produce a single positive and a single negative phases) of the form

$$\forall t, n.\ \mathtt{T}(t) \otimes \mathtt{C} \multimap \mathtt{T}(t+d) \otimes \mathtt{C}'$$

Focusing on this formula (stored in $\mathtt{system}$) necessarily produces the following derivation, starting with rule D_C (decision on the classical context):

$$\dfrac{\dfrac{\dfrac{\overset{\pi}{\mathtt{system}\,;\Delta \vdash \Downarrow \mathtt{T}(t)\otimes \mathtt{C}} \quad \overset{\psi}{\mathtt{system}\,;\Delta', \Downarrow \mathtt{T}(t+d)\otimes \mathtt{C}'\vdash G}}{\mathtt{system}\,;[\![s]\!]_t, \Downarrow \mathtt{T}(t)\otimes \mathtt{C} \multimap \mathtt{T}(t+d)\otimes \mathtt{C}'\vdash G}\ \multimap_L}{\mathtt{system}\,;[\![s]\!]_t, \Downarrow \forall t,n.\ \mathtt{T}(t)\otimes \mathtt{C} \multimap \mathtt{T}(t+d)\otimes \mathtt{C}'\vdash G}\ \forall_L \times 2}{\mathtt{system}\,;[\![s]\!]_t \vdash G}\ D_C$$

Here (Δ, Δ') is a partition of the atoms in $[\![s]\!]_t$. Since r is fireable in the state s, then Δ must contain all the atoms needed to prove $\mathtt{T}(t)$ and $\mathtt{C}$. Moreover, Δ' must correspond to the components not affected by the application of r, i.e., $\Delta' = [\![s]\!]_t \setminus \Delta$. Hence, derivation π takes the form:

$$\dfrac{\dfrac{}{\mathtt{system}\,;\mathtt{T}(t)\vdash \Downarrow \mathtt{T}(t)}\ I \quad \dfrac{}{\mathtt{system}\,;\mathtt{C}\vdash \Downarrow \mathtt{C}}\ I}{\mathtt{system}\,;\Delta \vdash \Downarrow \mathtt{T}(t)\otimes \mathtt{C}}\ \otimes_R$$

This means that $\Delta = \{\mathtt{T}(t), \mathtt{C}\}$. On the other hand, derivation ψ starts with the release rule R (since $\otimes_L$ must be introduced in the negative phase and then, focusing is lost) and we have

$$\dfrac{\mathtt{system}\,;\Delta', \mathtt{T}(t+d), \mathtt{C}'\vdash G}{\mathtt{system}\,;\Delta', \Downarrow \mathtt{T}(t+d)\otimes \mathtt{C}'\vdash G}\ R, \otimes_L$$

In the last sequent, the negative phase ends. Note that the set $\{\mathtt{T}(t+d), \mathtt{C}'\}$ corresponds to $[\![s']\!]_{t+d}$.

Corollary 1: Adequacy. Let s and s' be two states. Then $s \xrightarrow{(r,d)} s'$ iff the sequent $\mathtt{system}; \cdot \vdash [\![s]\!]_t \multimap [\![s]\!]_{t+d}$ is provable.

Proof. Note that after the negative phase, we have:

$$\dfrac{\mathtt{system}; \Delta \vdash [\![s]\!]_{t+d}}{\mathtt{system}; \cdot \vdash [\![s]\!]_t \multimap [\![s]\!]_{t+d}}$$

where Δ is the multiset of atoms in $[\![s]\!]_t$. We cannot focus on those atoms (since they are positive). Moreover, we cannot focus on $[\![s]\!]_{t+d}$ (since the atom $\mathtt{T}(t+d)$ is not in Δ nor in $\mathtt{system}$). Hence, we can only focus on the formulas in $\mathtt{system}$. We conclude by focusing on the encoding of r and using Theorem 1.

The proof of Corollary 2 follows easily from Theorem 1.

B Proof of the properties of the model

Property 1. The following sequent is provable:

$$\mathtt{system}\ ;\ . \vdash \forall n, t.\ \mathtt{T}(t)\otimes \mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}]) \multimap \exists d.\ \mathtt{T}(t+d)\otimes \mathtt{C}(n, bone, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])$$

Proof. After the negative phase (using the rules $\forall_R$, $\otimes_L$, $\multimap_R$), we have only one proof obligation: $\mathtt{system}\ ; \mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}]), \mathtt{T}(t) \vdash G$, where G is

$$\exists t_d.\mathtt{T}(t + t_d) \otimes \mathtt{C}(n, bone, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])$$

Note that the only non-atomic formulas are G and those formulas in $\mathtt{system}$. The proof proceeds by focusing, several times, on the formulas in $\mathtt{system}$ thus transforming the state $\mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}])$. In the end, we focus on G and the proof ends. Using the rules of the system as macro logical rules (see Corollary 2), we have the following:

$$\dfrac{\dfrac{\dfrac{\mathtt{system}\ ; \mathtt{C}(n, bone, 8, [\mathtt{EPCDCDME}]), \overset{\pi}{\mathtt{T}}(t + d_{42}(3) + d_{52}(5) + d_{72}(7)) \vdash G}{\mathtt{system}\ ; \mathtt{C}(n, blood, 7, [\mathtt{EPCDCDME}]), \mathtt{T}(t + d_{42}(3) + d_{52}(5)) \vdash G}\ \mathtt{rl}(bl_{eccm2.7})}{\mathtt{system}\ ; \mathtt{C}(n, blood, 5, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}]), \mathtt{T}(t + d_{42}(3)) \vdash G}\ \mathtt{rl}(bl_{ecc2.5})}{\mathtt{system}\ ; \mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}]), \mathtt{T}(t) \vdash G}\ \mathtt{rl}(bl_{ec2.3})$$

In the above derivation, we note that, in the last sequent (bottom-up) we already reach the state $\mathtt{C}(n, bone, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])$, with delay $t_d = d_{42}(3) + d_{52}(5) + d_{72}(7)$. Hence, the derivation π corresponds to focusing and decomposing entirely the formula G:

$$\dfrac{\dfrac{\dfrac{}{\mathtt{system}\ ; \mathtt{T}(t + t_d) \vdash\Downarrow \mathtt{T}(t + t_d)}\ I \quad \dfrac{}{\mathtt{system}\ ; \mathtt{C}(n, bone, 8, [\mathtt{EPCDCDME}]) \vdash\Downarrow \mathtt{C}(n, bone, 8, [\mathtt{EPCDCDME}])}\ I}{\mathtt{system}\ ; \cdots \vdash\Downarrow \mathtt{T}(t + t_d) \otimes \mathtt{C}(n, bone, 8, [\mathtt{EPCDCDME}])}\ \otimes_R}{\mathtt{system}\ ; \mathtt{C}(n, bone, 8, [\mathtt{EPCDCDME}]), \mathtt{T}(t + t_d) \vdash\Downarrow G}\ \exists_R$$

Property 2. The following sequent is provable:

$$\begin{aligned}\mathtt{system}\ ; \ \vdash\ & \forall n, t.\ \mathtt{T}(t) \otimes (\mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}]) \oplus \mathtt{C}(n, blood, 4, [\mathtt{EPCAM}, \mathtt{CD47}])) \multimap \\ & \exists t_d.\ \mathtt{T}(t + t_d) \otimes \\ & \quad (\mathtt{C}(n, bone, 6, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus \mathtt{C}(n, bone, 7, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus \\ & \quad \mathtt{C}(n, bone, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus \mathtt{C}(n, bone, 9, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]))\end{aligned}$$

Proof. After the negative phase (using the rules $\forall_R$, $\otimes_L$, $\oplus_L$, $\multimap_R$), we have two proof obligations (due to $\oplus_L$)

$$\begin{aligned}&\text{(PO1)}\ \ \mathtt{system}\ ; \mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}]), \mathtt{T}(t) \vdash G \\ &\text{(PO2)}\ \ \mathtt{system}\ ; \mathtt{C}(n, blood, 4, [\mathtt{EPCAM}, \mathtt{CD47}]), \mathtt{T}(t) \vdash G\end{aligned}$$

where G is the goal

$$\begin{aligned}\exists t_d.\mathtt{T}(t + t_d) \otimes (\ & \mathtt{C}(n, bone, 6, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus && \text{first choice} \\ & \mathtt{C}(n, bone, 7, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus && \text{second choice} \\ & \mathtt{C}(n, bone, 8, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]) \oplus && \text{third choice} \\ & \mathtt{C}(n, bone, 9, [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}])) && \text{last choice}\end{aligned}$$

Let us start with the proof obligation (PO1). Similar to the proof of Property 1, we start by focusing on the formulas in $\mathtt{system}$ so that we may later focus on G. One of the possible paths/proofs leading to the conclusion of the goal G is the following:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathtt{system}\ ; \mathtt{C}(n, bone, 6, [\mathtt{EPCDCDME}]), \mathtt{T}(t + d_{40}(3) + d_{42}(2) + d_{50}(4) + d_{52}(3) + d_{72}(5)) \vdash G}{\mathtt{system}\ ; \mathtt{C}(n, blood, 5, [\mathtt{EPCDCDME}]), \mathtt{T}(t + d_{40}(3) + d_{42}(2) + d_{50}(4) + d_{52}(3)) \vdash G}\ \mathtt{rl}(bl_{eccm2.5})}{\mathtt{system}\ ; \mathtt{C}(n, blood, 3, [\mathtt{EPCDCD}]), \mathtt{T}(t + d_{40}(3) + d_{42}(2) + d_{50}(4)) \vdash G}\ \mathtt{rl}(bl_{ecc2.3})}{\mathtt{system}\ ; \mathtt{C}(n, blood, 4, [\mathtt{EPCDCD}]), \mathtt{T}(t + d_{40}(3) + d_{42}(2)) \vdash G}\ \mathtt{rl}(bl_{ecc0.4})}{\mathtt{system}\ ; \mathtt{C}(n, blood, 2, [\mathtt{EPCAM}, \mathtt{CD47}]), \mathtt{T}(t + d_{40}(3)) \vdash G}\ \mathtt{rl}(bl_{ec2.2})}{\mathtt{system}\ ; \mathtt{C}(n, blood, 3, [\mathtt{EPCAM}, \mathtt{CD47}]), \mathtt{T}(t) \vdash G}\ \mathtt{rl}(bl_{ec0.3})$$

In such a derivation, the rules $\texttt{rl}(bl_{ec0})$ and $\texttt{rl}(bl_{ecc0})$ could be used zero or more times - as long as the fitness remains positive. Moreover, in the last sequent (bottom-up) we have already reached the state $\texttt{C}(n, bone, 6, [\texttt{EPCAM}, \texttt{CD47}, \texttt{CD44}, \texttt{MET}])$, with delay $t_d = d_{40}(3)+d_{42}(2)+d_{50}(4)+d_{52}(3)+d_{72}(5)$ and derivation π proceeds as in the proof of Property 1.

The proof obligation (PO2) can be discharged similarly by several paths, depending (as in PO1) on the order of mutations CD44 and MET and the eventually many passenger mutations (rules $bl_{ec0}, bl_{ecc0}, bl_{ecc0}$, and bl_{ecm0}). We give here the shortest path and one of the longest paths, as an illustration.

$$\begin{array}{l}
\texttt{C}(n, blood, 4, [\texttt{EPCAM}, \texttt{CD47}]) \otimes \texttt{T}(t) \\
\multimap \texttt{T}(t + d_{42}(4)) \otimes \texttt{C}(n, blood, 6, [\texttt{EPCAM}, \texttt{CD47}, \texttt{CD44}]) - \texttt{rl}(bl_{ec2.4}) \\
\multimap \texttt{T}(t + d_{42}(4) + d_{52}(6)) \otimes \texttt{C}(n, blood, 8, [\texttt{EPCAM}, \texttt{CD47}, \texttt{CD44}, \texttt{MET}]) - \texttt{rl}(bl_{ecc2.6}) \\
\multimap \texttt{T}(t + d_{42}(4) + d_{52}(6) + d_{72}(8)) \otimes \texttt{C}(n, bone, 9, [\texttt{EPCAM}, \texttt{CD47}, \texttt{CD44}, \texttt{MET}]) - \texttt{rl}(bl_{eccm2.8})
\end{array}$$

$$\begin{array}{l}
\texttt{C}(n, blood, 4, [\texttt{EPCAM}, \texttt{CD47}]) \otimes \texttt{T}(t) \\
\multimap \texttt{T}(t + d_{40}(4)) \otimes \texttt{C}(n, blood, 3, [\texttt{EPCAM}, \texttt{CD47}]) - \texttt{rl}(bl_{ec0.4}) \\
\multimap \texttt{T}(t + d_{40}(4) + d_{43}(3)) \otimes \texttt{C}(n, blood, 4, [\texttt{EPCAM}, \texttt{CD47}, \texttt{MET}]) - \texttt{rl}(bl_{ec3.3}) \\
\multimap \texttt{T}(t + d_{40}(4) + d_{43}(3) + d_{60}(4) \otimes \texttt{C}(n, blood, 3, [\texttt{EPCAM}, \texttt{CD47}, \texttt{MET}]) - \texttt{rl}(bl_{ecm0.4}) \\
\multimap \texttt{T}(t + d_{40}(4) + d_{43}(3) + d_{60}(4) + d_{62}(3)) \otimes \texttt{C}(n, blood, 5, [\texttt{EPCAM}, \texttt{CD47}, \texttt{CD44}, \texttt{MET}]) - \texttt{rl}(bl_{ecm2.3}) \\
\multimap \texttt{T}(t + d_{40}(4) + d_{43}(3) + d_{60}(4) + d_{62}(3) + d_{72}(5)) \otimes \texttt{C}(n, bone, 6, [\texttt{EPCAM}, \texttt{CD47}, \texttt{CD44}, \texttt{MET}]) - \texttt{rl}(bl_{ecm2.5})
\end{array}$$

Note that along with the time delay t_d we are looking for, the proof provides also the fitness of the extravasating CTC.

Property 3. The following sequents are provable:

$\texttt{system}\ ;\ . \vdash \forall n, t.\ \texttt{T}(t) \otimes \texttt{C}(n, breast, 1, [\texttt{EPCAM}]) \multimap \exists d.\ \texttt{T}(t + d) \otimes \texttt{C}(n, breast, 2, [\texttt{EPCAM}])$ and

$\texttt{system}\ ;\ . \vdash \forall n, t.\ \texttt{T}(t) \otimes \texttt{C}(n, breast, 2, [\texttt{EPCAM}]) \multimap \exists d.\ \texttt{T}(t + d) \otimes \texttt{C}(n, breast, 1, [\texttt{EPCAM}])$

Proof. In this case we present the Coq script needed to discard this proof. We prove separately the two sequents above:

```
Lemma Property3_Seq1: forall n t,
    exists d,
 |−F− Theory ; [ E{ fun _ x ⇒ perp TX{ fc1 d (var x)}} ** (C{ n; breast; 2; EP} )   ;
 Atom T{ Cte t} ; Atom C{ n ; breast ; 1 ; EP}]   ; UP [] .
Proof with solveF .
  idtac "Property3: Proving Cycle 1" .
  intros.
  eexists.
  applyRule (bre0r 1).
  (* Proving the goal *)
  eapply tri_dec1
    with (F:= E{ fun (T : Type) (x : T) ⇒ perp TX{ DX{ d20 2, var x}}} ** C{ n; breast; 2; EP}  ) ...
  eapply tri_tensor ...
  eapply tri_ex with (t:= (Cte t)) ...
  eapply Init1...
  eapply Init1...
Qed.

Lemma Property3_Seq2: forall n t,
    exists d,
 |−F− Theory ; [ E{ fun _ x ⇒ perp TX{ fc1 d (var x)}} ** (C{ n; breast; 1; EP} )   ;
 Atom T{ Cte t} ; Atom C{ n ; breast ; 2 ; EP}]   ; UP [] .
Proof with solveF .
  idtac "Property3: Proving Cycle 2" .
  intros.
  eexists.
  applyRule (bre0 2).
  (* Proving the goal *)
  eapply tri_dec1 with
```

```
    (F:= E{ fun (T : Type) (x : T) ⇒ perp TX{ DX{ d20 2, var x}}} ** C{ n; breast; 1; EP}  ) ...
  eapply tri_tensor ...
  eapply tri_ex with (t:= (Cte t)) ...
  eapply Init1...
  eapply Init1...
Qed.
```

Property 4. Let Δ be a multiset of atoms of the form $\mathtt{C}(\cdot)$. Then, in any derivation of the form

$$\frac{\mathtt{system}; \Delta, \mathtt{T}(t+d), St \vdash G}{\mathtt{system}; \Delta, \mathtt{T}(t), \mathtt{C}(n, c, 0, m) \vdash G}\ \mathtt{rl}(\cdot)$$

we have $St = \mathtt{A}(n)$.

Proof. We know that the above derivation must start by focusing on one of the formulas in `system` (Theorem `FocusOnlyTheory` in our formalisation). Then, we proceed by case analysis on all of the rules. If the rule is not fireable, then we cannot focus on that rule since the initial rule cannot be applied (and the above derivation is not valid). If the rule can be fired, due to Corollary 2, we know that the resulting St is necessarily the one-step transformation of $\mathtt{C}(n, c, 0, m)$, that, in this case, satisfies $St = \mathtt{A}(n)$.

Property 5. Let Δ be a multiset of atoms of the form $\mathtt{C}(\cdot)$. Then, in any derivation of the form

$$\frac{\mathtt{system}; \Delta, \mathtt{T}(t+t_d), St \vdash G}{\mathtt{system}; \Delta, \mathtt{T}(t), \mathtt{C}(n, blood, f, m) \vdash G}\ \mathtt{rl}(\cdot)$$

with m containing `CD47`, it must be the case that

1. either $St = \mathtt{C}(n, blood, f-1, m)$,
2. or $St = \mathtt{A}(n)$,
3. or $St = \mathtt{C}(n, blood, f+2, m')$ with m' being as m plus an additional mutation,
4. or $St = \mathtt{C}(n, bone, f+1, m)$ with $m = [\mathtt{EPCAM}, \mathtt{CD47}, \mathtt{CD44}, \mathtt{MET}]$.

Proof. In this case we present the Coq script needed to discard this proof. Definition `GoalP5` is just a shorthand to denote the goal we need to prove.

```
Definition GoalP5 (f n t m:nat) :=
  |−F− Theory ; [ Atom  T{ Cte t} ; Atom C{ n ;  blood; f ;  m} ] ;  UP []  →
                                                                In m [EPCDCD ;EPCDCDME;EPCDCDMEse] →
  exists (d m' :nat),
  |−F− Theory ; [ Atom  T{ d s+ Cte t} ; Atom C{ n ;blood;  f  −1; m } ] ;  UP []  ∨
  |−F− Theory ; [ Atom  T{ d s+ Cte t} ; Atom A{ n} ] ;  UP []  ∨
  (|−F− Theory ; [ Atom  T{ d s+ Cte t} ; Atom C{ n ;blood;  f  +2; m' } ] ;  UP []  ∧ (plusOne m m')  ) ∨
  (m = EPCDCDME ∧ |−F− Theory ; [ Atom  T{ d s+ Cte t} ; Atom C{ n ;bone;  f   + 1; EPCDCDME } ] ;  UP [])  .

Proposition Property5 : forall (f n t m:nat) ,
    GoalP5 f n t m.
  idtac "Proving Property 5".
  intros f n t m HProof HCaseM.
  apply FocusOnlyTheory in HProof;auto;
    destruct HProof as [R HProof];  destruct HProof as [HIn HProof];
      time "Solve:" repeat (first [  CaseRule  | DecomposeRule; FindUnification | SolveGoal]) .
Qed.
```