



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie

**From the mechanistic modeling of signaling pathways in
cancer to the interpretation of models and their
contributions: clinical applications and statistical
evaluation**

Soutenance prévue par

Jonas BEAL

Le 23/09/2020

École doctorale n°515

Complexité du Vivant

Spécialité

Génomique

Composition du jury :

Adeline LECLERQ-SAMSON
Professeur, Université Grenoble Alpes *Rapporteur*

Lodewyk WESSELS
Professeur, Netherlands Cancer Institute *Rapporteur*

Émilie LANOY
Ingénierie de recherche, Gustave Roussy *Examinateur*

Denis Thieffry
Professeur, ENS, PSL *Examinateur*

Emmanuel BARILLOT
Directeur de Recherche, Institut Curie, PSL *Directeur de thèse*

Aurélien LATOUCHÉ
Professeur, Institut Curie, Cnam *Directeur de thèse*

Laurence CALZONE
Ingénierie de recherche, Institut Curie, PSL *Membre invitée, co-encadrante de thèse*

Abstract

Beyond its genetic mechanisms, cancer can be understood as a network disease that often results from the interaction between different perturbations in a cellular regulatory network. The dynamics of these networks and associated signaling pathways are complex and require integrated approaches. One approach is to design mechanistic models that translate the biological knowledge of networks in mathematical terms to simulate the molecular features of cancers in a computer-readable form. However, these models only reflect the general mechanisms at work in cancers.

This thesis proposes to define personalized mechanistic models of cancer. A generic model is first defined in a logical (or Boolean) formalism, before using omics data (mutations, RNA, proteins) from patients or cell lines in order to make the model specific to each one profile. These personalized models can then be compared with the clinical data of patients in order to validate them. The response to treatment is investigated in particular in this thesis. The explicit representation of the molecular mechanisms by these models allows to simulate the effect of different treatments according to their targets and to verify if the sensitivity of a patient to a drug is well predicted by the corresponding personalized model. An example concerning the response to BRAF inhibitors in melanomas and colorectal cancers is thus presented.

The comparison of mechanistic models of cancer, those presented in this thesis and others, with clinical data also encourages a rigorous evaluation of their possible benefits in the context of medical use. The quantification and interpretation of the value of certain prognostic models is briefly presented before focusing on the particular case of models able to recommend the best treatment for each patient according to his molecular profile. A theoretical framework is defined to extend causal inference methods to the evaluation of such precision medicine algorithms. An illustration is provided using simulated data and patient derived xenografts.

All the methods and applications put forward a possible path from the design of mechanistic models of cancer to their evaluation using statistical

models emulating clinical trials.

Key-words: Modeling, Cancer, Mechanistic model, Biostatistics, Causal inference, Precision medicine

Résumé

Au delà de ses mécanismes génétiques, le cancer peut-être compris comme une maladie de réseaux qui résulte souvent de l'interaction entre différentes perturbations dans un réseau de régulation cellulaire. La dynamique de ces réseaux et des voies de signalisation associées est complexe et requiert des approches intégrées. Une d'entre elles est la conception de modèles dits mécanistiques qui traduisent mathématiquement la connaissance biologique des réseaux afin de pouvoir simuler le fonctionnement moléculaire des cancers informatiquement. Ces modèles ne traduisent cependant que les mécanismes généraux à l'oeuvre dans certains cancers en particulier.

Cette thèse propose en premier lieu de définir des modèles mécanistiques personnalisés de cancer. Un modèle générique est d'abord défini dans un formalisme logique (ou Booléen), avant d'utiliser les données omiques (mutations, ARN, protéines) de patients ou de lignées cellulaires afin de rendre le modèle spécifique à chacun. Ces modèles personnalisés peuvent ensuite être confrontés aux données cliniques de patients pour vérifier leur validité. Le cas de la réponse clinique aux traitements est exploré en particulier dans cette thèse. La représentation explicite des mécanismes moléculaires par ces modèles permet en effet de simuler l'effet de différents traitements suivant leur mode d'action et de vérifier si la sensibilité d'un patient à un traitement est bien prédite par le modèle personnalisé correspondant. Un exemple concernant la réponse aux inhibiteurs de BRAF dans les mélanomes et cancers colorectaux est ainsi proposé.

La confrontation des modèles mécanistiques de cancer, ceux présentés dans cette thèse et d'autres, aux données cliniques incite par ailleurs à évaluer rigoureusement leurs éventuels bénéfices dans la cadre d'une utilisation médicale. La quantification et l'interprétation de la valeur de certains modèles à visée pronostique est brièvement présentée avant de se focaliser sur le cas particulier des modèles capables de sélectionner le meilleur traitement pour chaque patient en fonction des ses caractéristiques moléculaires. Un cadre théorique est proposé pour étendre les méthodes d'inférence causale à l'évaluation de tels algorithmes de médecine de précision. Une illustration

est fournie à l'aide de données simulées et de xénogreffes dérivées de patients.

L'ensemble des méthodes et applications décrites tracent donc un chemin, de la conception de modèles mécanistiques de cancer à leur évaluation grâce à des modèles statistiques émulant des essais cliniques.

Mots-clés: Modélisation, Cancer, Modèle mécanistique, Biostatistiques, Inférence causale, Médecine de précision

Acknowledgements

Many persons to thanks. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

Preface

The present thesis is structured in three parts, each subdivided into three chapters. Since the whole thesis is about cancer modeling, the first part aims at defining the type of model to be referred to, and in particular models that will be called mechanistic, as well as the object of the modeling, i.e. the molecular networks involved in cancer. So the first part answers the question

The second part will be devoted to the methods developed during this thesis to transform qualitative models of molecular networks, known as logic models, into personalized models that can be interpreted clinically. In short, **how can a mathematical representation of biological knowledge be transformed into a tool that contributes to the understanding of the clinical manifestations of cancer?**

Finally, the third and last part will look at how the clinical relevance of all the above-mentioned models can be rigorously evaluated, both in their ability to predict the evolution of the disease and in their ability to recommend the most appropriate treatments for each patient. **How to quantify and interpret the value of the clinical information delivered by these models?**

Moreover, this thesis also exists in an online version that allows to take advantage of the interactivity of some graphs and applications: <https://jonasbeal.github.io/files/PhdThesis/>.

Scientific content

Except for the first part, essentially introductory and based on scientific literature, the different chapters are based on original scientific work done during this thesis (2017-2020) and mentioned at the beginning of each chapter in a box similar to this one.

The main articles behind this thesis are indicated below with one published article and two pre-prints currently under review:

- Béal, Jonas, Arnau Montagud, Pauline Traynard, Emmanuel Barillot, and Laurence Calzone. “Personalization of logical models with multi-omics data allows clinical stratification of patients.” *Frontiers in physiology* 9 (2019): 1965. [Link](#).
- Béal, Jonas, Lorenzo Pantolini, Vincent Noël, Emmanuel Barillot, and Laurence Calzone. “Personalized logical models to investigate cancer response to BRAF treatments in melanomas and colorectal cancers.” *bioRxiv* (2020). [Link](#).
- Béal, Jonas, and Aurélien Latouche. “Causal inference with multiple versions of treatment and application to personalized medicine.” *arXiv preprint arXiv:2005.12427* (2020). [Link](#).

These three articles were described or completed in oral presentations, respectively in International Conference of Systems Biology 2018, conference on Intelligent Systems for Molecular Biology (ISMB/ECCB 2019, [Video](#)) and conference of International Society of Clinical Biostatistics (ISCB41, coming in August 2020).

Table of contents

	Page
List of Tables	xii
List of Figures	xiii
I Cells and their models	1
1 Scientific modeling: abstract the complexity	3
1.1 What is a model?	4
1.1.1 In your own words	4

List of Tables

Table	Page
-------	------

List of Figures

Figure	Page
1.1 A scientist and his model	4

Part I

Cells and their models

Scientific modeling: abstract the complexity

"Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable."

Paul Valéry (Mauvaises pensées et autres, 1942)

The notion of modeling is embedded in science, to the point that it has sometimes been used to define the very nature of scientific research.

What is called a model can, however, correspond to very different realities which need to be defined before addressing the object of this thesis which will consist, if one wants to be mischievous, in analyzing models with other models. This semantic elucidation is all the more necessary as this thesis is interdisciplinary, suspended between systems biology and biostatistics. In order to convince the reader of the need for such a preamble, he is invited to ask a statistician and a biologist how they would define what a model is.

The present thesis is structured in three parts, each subdivided into three chapters. Since the whole thesis is about cancer modeling, the first part aims at defining the type of model to be referred to, and in particular models that will be called mechanistic, as well as the object of the modeling, i.e. the molecular networks involved in cancer. So the first part answers the question

The second part will be devoted to the methods developed during this thesis to transform qualitative models of molecular networks, known as logic

CHAPTER 1. SCIENTIFIC MODELING: ABSTRACT THE COMPLEXITY



Figure 1.1: **A scientist and his model.** Joseph Wright of Derby, *A Philosopher Giving a Lecture at the Orrery (in which a lamp is put in place of the sun)*, c. 1763-65, oil on canvas, Derby Museums and Art Gallery

models, into personalized models that can be interpreted clinically. In short, **how can a mathematical representation of biological knowledge be transformed into a tool that contributes to the understanding of the clinical manifestations of cancer?**

1.1 What is a model?

1.1.1 In your own words

A model is first of all an ambiguous object and a polysemous word. It therefore seems necessary to start with a semantic study. Among the many meanings and synonymous proposed by the dictionary, while some definitions are more related to art, several find echoes in scientific practice. It is sometimes a question of the physical representation of an object, often on a reduced scale as in Figure 1.1, and sometimes of a theoretical description intended to facilitate the understanding of the way in which a system works [?]. It is even sometimes an ideal to be reached and therefore an ambitious prospect for an introduction.

1.1. WHAT IS A MODEL?

A model is first of all an ambiguous object and a polysemous word. It therefore seems necessary to start with a semantic study. Among the many meanings and synonymous proposed by the dictionary, while some definitions are more related to art, several find echoes in scientific practice. It is sometimes a question of the physical representation of an object, often on a reduced scale as in Figure 1.1, and sometimes of a theoretical description intended to facilitate the understanding of the way in which a system works [?]. It is even sometimes an ideal to be reached and therefore an ambitious prospect for an introduction.

RÉSUMÉ

Au delà de ses mécanismes génétiques, le cancer peut-être compris comme une maladie de réseaux qui résulte souvent de l'interaction entre différentes perturbations dans un réseau de régulation cellulaire. La dynamique de ces réseaux et des voies de signalisation associées est complexe et requiert des approches intégrées. Une d'entre elles est la conception de modèles dits mécanistiques qui traduisent mathématiquement la connaissance biologique des réseaux afin de pouvoir simuler le fonctionnement moléculaire des cancers informatiquement. Ces modèles ne traduisent cependant que les mécanismes généraux à l'oeuvre dans certains cancers en particulier.

Cette thèse propose en premier lieu de définir des modèles mécanistiques personnalisés de cancer. Un modèle générique est d'abord défini dans un formalisme logique (ou Booléen), avant d'utiliser les données omiques (mutations, ARN, protéines) de patients ou de lignées cellulaires afin de rendre le modèle spécifique à chacun. Ces modèles personnalisés peuvent ensuite être confrontés aux données cliniques de patients pour vérifier leur validité. Le cas de la réponse clinique aux traitements est exploré en particulier dans cette thèse. La représentation explicite des mécanismes moléculaires par ces modèles permet en effet de simuler l'effet de différents traitements suivant leur mode d'action et de vérifier si la sensibilité d'un patient à un traitement est bien prédictive par le modèle personnalisé correspondant. Un exemple concernant la réponse aux inhibiteurs de BRAF dans les mélanomes et cancers colorectaux est ainsi proposé. La confrontation des modèles mécanistiques de cancer, ceux présentés dans cette thèse et d'autres, aux données cliniques incite par ailleurs à évaluer rigoureusement leurs éventuels bénéfices dans la cadre d'une utilisation médicale. La quantification et l'interprétation de la valeur de certains modèles à visée pronostique est brièvement présentée avant de se focaliser sur le cas particulier des modèles capables de sélectionner le meilleur traitement pour chaque patient en fonction des ses caractéristiques moléculaires. Un cadre théorique est proposé pour étendre les méthodes d'inférence causale à l'évaluation de tels algorithmes de médecine de précision. Une illustration est fournie à l'aide de données simulées et de xénogreffes dérivées de patients

L'ensemble des méthodes et applications décrites tracent donc un chemin, de la conception de modèles mécanistiques de cancer à leur évaluation grâce à des modèles statistiques émulant des essais cliniques.

MOTS CLÉS

Modélisation, Cancer, Modèle mécanistique, Biostatistiques, Inférence causale, Médecine de précision.

ABSTRACT

Beyond its genetic mechanisms, cancer can be understood as a network disease that often results from the interaction between different perturbations in a cellular regulatory network. The dynamics of these networks and associated signaling pathways are complex and require integrated approaches. One approach is to design mechanistic models that translate the biological knowledge of networks in mathematical terms to simulate the molecular features of cancers in a computer-readable form. However, these models only reflect the general mechanisms at work in cancers.

This thesis proposes to define personalized mechanistic models of cancer. A generic model is first defined in a logical (or Boolean) formalism, before using omics data (mutations, RNA, proteins) from patients or cell lines in order to make the model specific to each one profile. These personalized models can then be compared with the clinical data of patients in order to validate them. The response to treatment is investigated in particular in this thesis. The explicit representation of the molecular mechanisms by these models allows to simulate the effect of different treatments according to their targets and to verify if the sensitivity of a patient to a drug is well predicted by the corresponding personalized model. An example concerning the response to BRAF inhibitors in melanomas and colorectal cancers is thus presented.

The comparison of mechanistic models of cancer, those presented in this thesis and others, with clinical data also encourages a rigorous evaluation of their possible benefits in the context of medical use. The quantification and interpretation of the value of certain prognostic models is briefly presented before focusing on the particular case of models able to recommend the best treatment for each patient according to his molecular profile. A theoretical framework is defined to extend causal inference methods to the evaluation of such precision medicine algorithms. An illustration is provided using simulated data and patient derived xenografts.

All the methods and applications put forward a possible path from the design of mechanistic models of cancer to their evaluation using statistical models emulating clinical trials.

KEYWORDS

Modeling, Cancer, Mechanistic model, Biostatistics, Causal inference, Precision medicine.