UNIVERSITÀ DEGLI STUDI DI CATANIA

Dipartimento di Matematica e Informatica

Dottorato di ricerca in informatica (Internazionale) XXXIV Ciclo

Giuseppe Sgroi

Mycobacterium tuberculosis - immune system dynamics through agent-based modeling methodology: an important step in accelerating vaccine discovery

Thesis

Tutor: Prof. Francesco Pappalardo

Anno Accademico 2020 - 2021

Declaration of Authorship

I, Giuseppe Sgroi, declare that this thesis titled, "Mycobacterium tuberculosis - immune system dynamics through agent-based modeling methodology: an important step in accelerating vaccine discovery" and the work presented in it is my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- This PhD position has been funded by European Commission under the Horizon 2020 framework program, project "STriTuVaD", Grant agreement ID: 777123. Results of this PhD work belongs to the STriTuVaD consortium.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Abstract

Tuberculosis (TB) has existed for millennia and remains a major global health problem. It causes ill-health in millions of people each year and in 2015 was one of the top 10 causes of death worldwide, ranking above HIV/AIDS as one of the leading causes of death from an infectious disease. TB is a contagious airborne disease caused by the bacillus Mycobacterium tuberculosis (MTB). It typically affects the lungs (pulmonary TB) but also affects other sites (extrapulmonary TB). Approximately 1/3 of the world's population is affected by the disease but without any symptoms (known as latent TB infection (LTBI)), and approximately 10% of these people will likely develop active disease during their lifetime and become capable of transmitting the mycobacterium. TB can be treated through the selection of four standards (first-line) drugs. However, there is currently no effective vaccine in preventing TB disease in adults, either before or after exposure to TB infection. In biomedical, pharmaceutical, and toxicology research, the safety and efficacy of biomedical products are ultimately tested on humans via clinical trials after prior laboratory testing in vitro and/or in vivo on animals. The complete development chain of a new biomedical product and its introduction to the market is very long and expensive. Alternative methodologies to reduce animal and human testing are needed to address the safety and efficacy issues of human clinical trials, the ethical ones, and the imperfection of predictions issued from laboratory and animal studies when applied to humans. Computer modeling and simulation are currently used to a certain degree in pharmacokinetics, pharmacodynamics, or mechanistic simulations. A research and technological roadmap on in-silico trials that use individualized computer simulations in testing interventional strategies is currently available, showing

both strong interest and potential benefit of expanding the computer modeling in drugs and other biomedical products research.

Most of the work carried out during the Ph.D. project falls under the main objectives of the STriTuVaD project. In this Ph.D. thesis, it has been extended the Universal Immune System Simulator (UISS) to simulate the dynamics of tuberculosis and its interactions (physiological model), i.e., the behavior of the Mycobacterium within the host organism and its interactions with the immune system (active and latent form scenario) (disease model). In addition, the mechanisms of action of isoniazid and RUTI® vaccine were developed (treatment model).

Abbreviations

Abbreviation	**Meaning**	**Page**
TB	Tuberculosis	6
MTB	Mycobacterium tuberculosis	6
WHO	World Health Organization	7
AIDS	Acquired Immune Deficiency Syndrome	7
HIV	Human Immunodeficiency Virus	7
Mdr-TB	Multidrug-resistant Tuberculosis	9
LTBI	Latent-type TB Infection	11
TST	Tuberculin Skin Test	13
PPD	Purified Protein Derivative	13
RR-TB	Rifampicin-Resistant TB	15
BCG	Bacillus Calmette – Guérin	16
ABM	Agent-Based Model	18
CA	Cellular Automata	28
CS	Celada-Seiden	30
MAS	Multi-Agent Systems	32
GIS	Geographic Information System	39
UISS	Universal Immune System Simulator	42
TLR	Toll-Like Receptor	44
PAMP	Pathogen-Associated Molecular Pattern	44
HPC	High-Performance Computing	45
DT	Digital Twin	46
GA	Genetic Algorithm	46
SA	Simulated Annealing	46
HAART	Highly Active Anti-Retroviral Therapies	47
RRMS	Relapsing-Remitting Multiple Sclerosis	47

Abbreviation	Meaning	Page
UISS-TB	Universal Immune System Simulator – Tuberculosis	48
AM	Alveolar Macrophage	49
N	Neutrophil	49
LXA4	Lipoxin A4	50
PGE2	Prostaglandin E2	50
DC	Dendritic Cell	50
MHC-I	Major Histocompatibility Complex of class I	50
MHC-II	Major Histocompatibility Complex of class II	50
Treg	Regulatory T cells	51
IgG	Immunoglobulins type G	51
IgA	Immunoglobulins type A	51
IgE	Immunoglobulins type E	51
IgM	Immunoglobulins type M	51
INH	Isoniazid	54
MoA	Mechanism of Action	59
SP-TB	Smear-Positive TB	63
SN-TB	Smear-Negative TB	63
PETAL	ParallEl paThways AnaLyzer	80
BFS	Breadth-First Search	81
SVG	Scalable Vector Graphics	81
CSS	Cascading Style Sheet	81
KEGG	Kyoto Encyclopedia of Genes and Genome	82
FIFO	First Input First Output	86
CML	Chronic Myelogenous Leukemia	93
NSCLC	Non-Small Cell Lung Cancer	93
HNSCC	Head and Neck Squamous Cell Carcinoma	93
TKI	Tyrosine Kinases Inhibitor	93
PEAK	Pattern rEcognition frAmeworK	98

Acknowledgements

I want to express my gratitude to my Ph.D. supervisor, Prof. Francesco Pappalardo, for being fundamental support during the three years.

Likewise, I would like to thank Dr. Giulia Russo for her advice and support in writing this thesis. I would also like to thank all my colleagues in the COMBINE group.

I would especially like to thank my partner Giovanna for her constant support and my whole family.

Finally, I would like to thank my friend Joshua because he also thanks to that I have come this far.

Contents

Chapter 1

Aim of the thesis

In Silico Trial for Tuberculosis Vaccine Development (STriTuVaD) is a Horizon 2020 project funded by the European Commission for the development of a computational model to test, predict and verify the *in silico* specific vaccination strategies against tuberculosis.
The project started on 1st February 2018 and will end on 31st January 2023. Within the consortium, there are seven international partners:

- **ETNA BIOTECH**: Etna Biotech is one of the two vaccine research centre of Zydus Group, which is part of Cadila Healthcare, the fourth largest pharmaceutical company in India. Dr Fichera is the coordinator of the STriTuVaD project.

- **University of Catania (UniCT)**: Prof Pappalardo is the scientific coordinator of the project and the research team supervised by him is developing, through the Universal Immune System Simulator (UISS), the predictive model consdiered the main core of the augmented *in silico* trial that will be tested in the STriTuVaD project.

- **University of Sheffield (USFD)**: Dr Juarez in close collaboration with Prof Viceconti and Prof Pappalardo's team, is developing the Bayesian model to be used to combine digital and physical patients, along with the necessary criteria for avoiding any bias. Dr Richmond is collaborating with Prof Pappalardo's team to speed up the execution of the UISS framework, and to enable stochastic explorations using Monte Carlo methods.

- **Alma Mater Studiorum - University of Bologna**: Prof Viceconti's research group will take care of internal and external communication about any project news, results and achievements, along with the potential exploitation activities. Prof Viceconti in close collaboration with Prof Pappalardo and Dr Juarez, is focusing on the *in silico* model validation from a regulatory point of view.

- **Archivel Farma**: Archivel Farma is a R&D biotech company that develops immunotherapeutic agents to respond to unmet medical needs, such as in multidrug-resistant tuberculosis (MDR-TB). The RUTI® vaccine was developed and produced by Archivel Farma. Within the STriTuVaD project, the company role is to design and run the clinical trial necessary to validate the *in silico* trial technology.

- **TuBerculosis Vaccine Initiative (TBVI)**: is a non-profit foundation that facilitates the discovery and development of new, safe and effective TB vaccines that are accessible and affordable for the whole population. In particular, TBVI provides support and knowledge about TB vaccine clinical development.

- **All India Institute Of Medical Sciences (AIIMS)**: is a medical college and medical research public University based in New Delhi, India. Within the

project, AIIMS will conduct a Phase II clinical trial to assess the safety, immunogenicity and efficacy of RUTI® therapeutic vaccine against drug-sensitive and multi-drug resistant TB for the validation of *in silico* trial.

The StriTuVaD multidisciplinary consortium is co-working to deliver UISS *in silico* trial platform (see chapter 4) to simulate the relevant individual human physiology and physiopathology in patients affected by Mycobacterium tuberculosis.
Digital twins cohorts were generated to study the effects of specific treatments against tuberculosis as agreed within the proposal, allowing the simulation of RUTI® mechanism of action, and predicting its treatment outcomes in a personalized medicine approach. Figure 1.1 briefly sketches the month-by-month main phases and concepts of the STriTuVaD proposal.

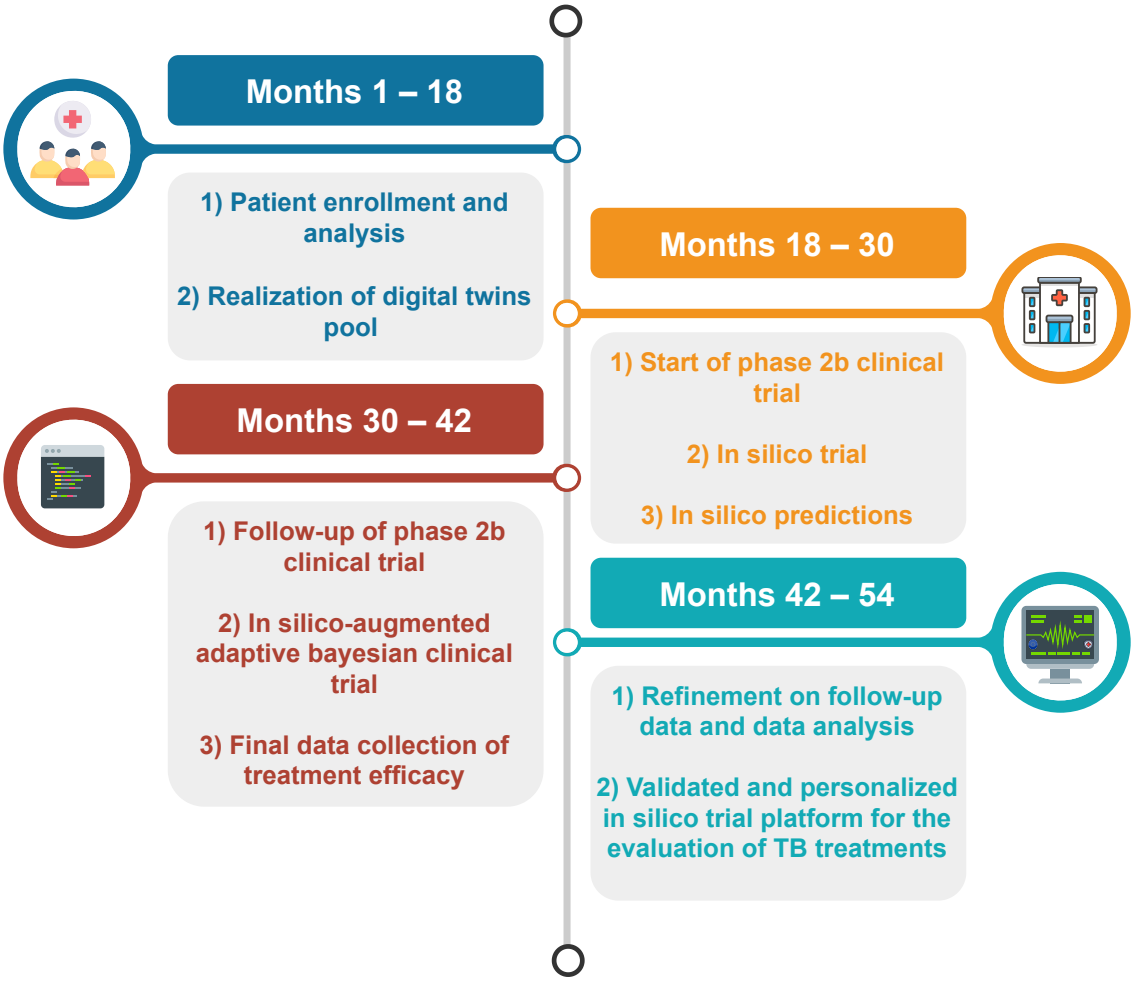


Figure 1.1: A sketch depicting the main steps of the STriTuVaD project development.

The main objective of this Ph.D. project is to provide a new advanced version of UISS *in silico* platform which also implements the treatment layer.

As a first step, the TB cellular and molecular pathway was studied through ParallEl paThways AnaLyzer (PETAL). It is a Python tool that automatically explores and detects the most relevant nodes within a KEGG pathway, scanning and performing an in-depth search. In order to clarify potential links between the genes of our interest, PETAL was performed to allow users to find hidden interactions among

significant proteins belonging to the same pathway and other proteins within possible linked pathways. From a biological point of view, this depth search helps in retrieving certain interesting genes potentially hidden and involved in an important biological and cellular process. Thus, PETAL can lead to the identification of possible pathways for drug resistance, providing novel insights about TB resistance.
Secondly, for the three modeling layers (physiological, disease, and treatment), MetricUISS has been used to faithfully generate digital twin cohorts based on observed in vivo data. This tool performs statistical analysis, and studies UISS results from different perspectives, e.g., correctly simulating disease dynamics, studying treated patients, and studying retrospective and prospective data.
In addition, software that explores, correlates, analyzes, and classifies data straightforwardly was developed under the codename PEAK (Pattern rEcognition frAmeworK). It allows users to reduce the time needed for data analysis and discovering unknown relationships between different data. PEAK was used in a specific case study dealing with a well-defined dataset representing a cohort of 10,000 digital twins affected by COVID-19 with different immunological characteristics. This dataset was created to evaluate the correlations of specific components (e.g., cytotoxic T cells, antibodies, Interferon-γ, Lung epithelial cells, and Interleukin 6) activated during exposure to COVID-19.

Chapter 2

Tuberculosis

2.1 Introduction

Tuberculosis (TB) is an infectious disease caused by Mycobacterium tuberculosis (MTB) [1], an organism that belongs to the Mycobacteriaceae family, including other mycobacteria genetically related.

Mycobacteria are thin bacilli that are 2 to 4 µm long. Several species of mycobacteria are harmless and live in the surface layers of the soil.

Tuberculosis usually affects the lungs in pulmonary tuberculosis, although, in a percentage of cases (up to 1/3), other organs are involved, such as in extrapulmonary tuberculosis.

Tuberculosis is a disease among the top 10 causes of mortality worldwide, especially in underdeveloped areas of the African and Asian continents [2]; it has existed for millennia and remains a major global health problem.

Most importantly, tuberculosis represents a clinical and public health problem worldwide. However, its incidence and prevalence have significantly decreased in more developed nations. In contrast, it has increased in less developed countries due to the emergence of strains resistant to many antibacterial drugs.

The World Health Organization (WHO) estimates that tuberculosis is the second leading cause of death from infectious disease, behind only AIDS (Acquired Immune Deficiency Syndrome). Each year, tuberculosis kills about two million people, mostly concentrated in developing countries.

2.2 Epidemiology

Tuberculosis is widespread throughout the world. Since 1980, the disease has worsened, which can be explained, in part, by the global spread of Human Immunodeficiency Virus (HIV) infection and the emergence of drug-resistant strains [3].

In particular, in 2016, the largest number of new tuberculosis cases occurred in Asia, with 45% of new cases, followed by Africa, with 25% new cases (Figure 2.2).

The WHO estimates that about one-third of the world's population is currently infected with TB, and tuberculosis represents one of the leading causes of death (in 2016, it ranks ninth).

In this context, approximately within 10.4 million of new cases (incident cases), 90% involve adults, 65% males and 10% people with HIV infection (74% in Africa) (Figure 2.1) [4].

The total estimation of patients who died from TB is 95%, registered in the middle- or low-income countries, 56% of new cases are concentrated in seven countries: India, Indonesia, China, Philippines, Pakistan, Nigeria, and South Africa.

Therefore, India ranks as the country with the highest-burden of TB and has the highest number of deaths worldwide [5].

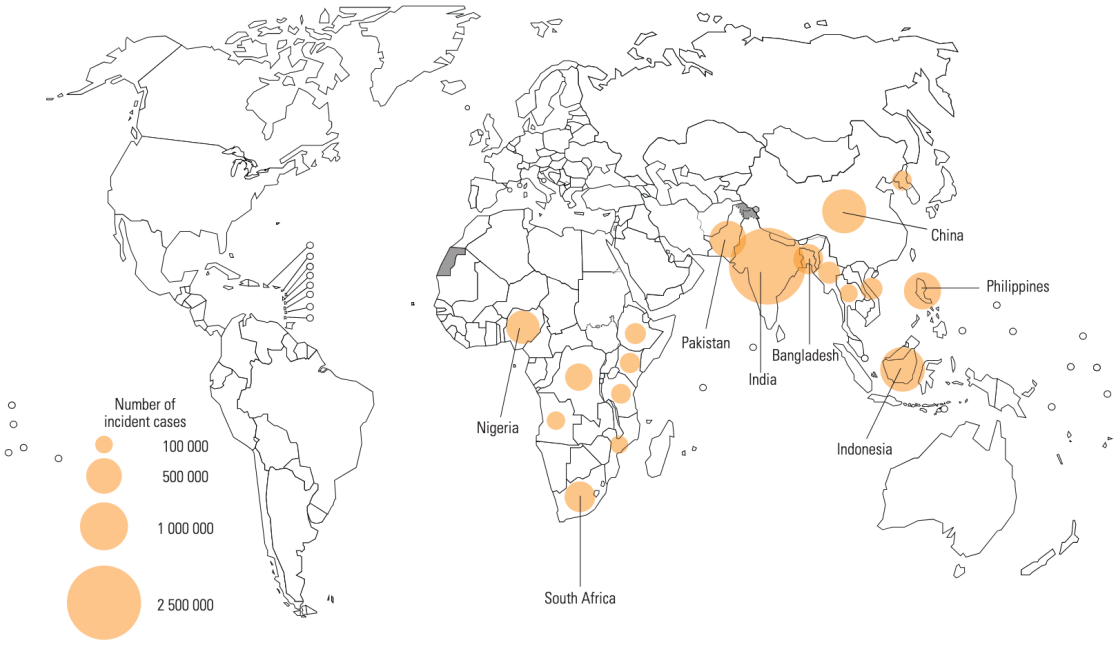


Figure 2.1: Countries that had at least 100 000 incident cases of TB in 2019 [4].

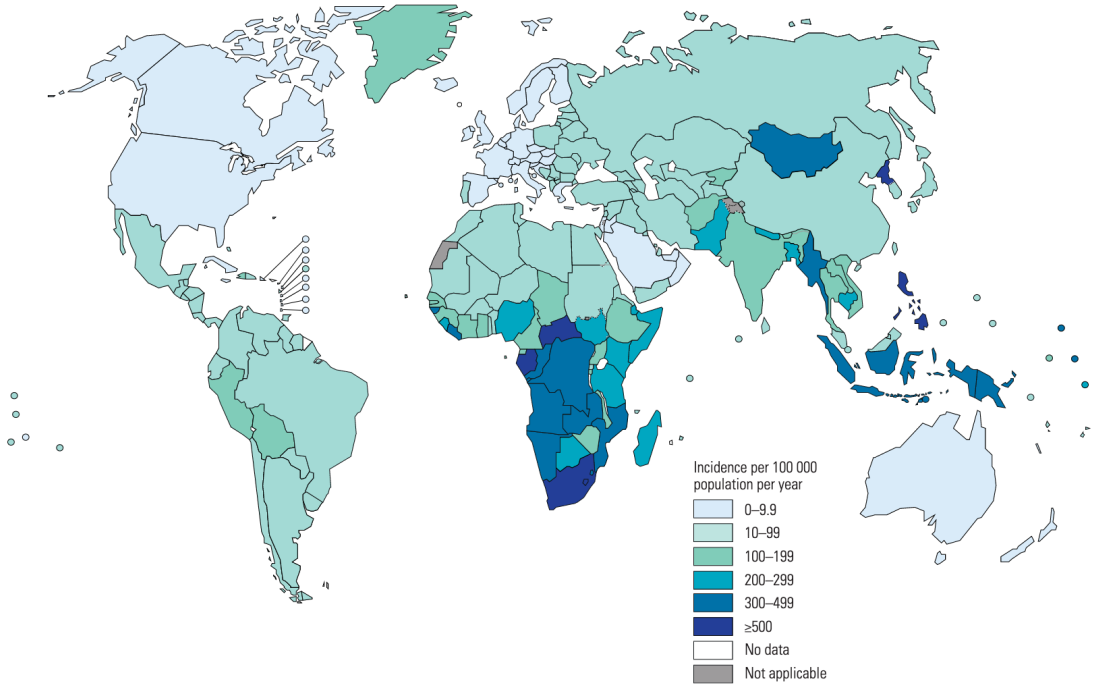


Figure 2.2: Estimated tuberculosis incidence rate in 2019 [4].

In Europe, an "action plan" called Tuberculosis Action Plan for the WHO European Region 2016-2020 has been outlined [6].
The plan defines and recommends the European States a series of activities to circumscribe and contain the spread of Multidrug-resistant tuberculosis (Mdr-TB). Mdr-TB is a form of tuberculosis that is very difficult to treat. Less than 50% of those affected can recover, as it is resistant to isoniazid and rifampicin, the most popular antibiotics against the infection.
The plan's proposed strategy for halting the disease's spread is to ensure universal access to prevention practices, diagnosis, and treatment in all states in the region. Therefore, by 2020, the plan set a goal to:

- reduce TB deaths by 35%;
- reduce TB incidence by 25%;
- increase successful cases among individuals with Mdr-TB by 75%.

These results can only be achieved by improving action strategies at the level of treatment, prevention and research, and sharing the knowledge and experience gained by each European country. According to WHO estimates, Italy is one of the countries with a low incidence of the disease (about 20 cases per 100,000 inhabitants), and health authorities continuously monitor its spread (Figure 2.3).
In 2016, there were approximately 4072 registered TB cases, down slightly from the last ten years. Specifically, 1.9% of cases involved children within four years of age, 2% deals with children of 5-14 years of age, 18% is related to 15-24 years old, 35% of cases involved individuals of 25-44 years of age, another 18% involved individuals of 45-64 years of age, and the final 14% involved those over 65.
Of the total cases, only 3778 were classified as new ones (not previously treated),

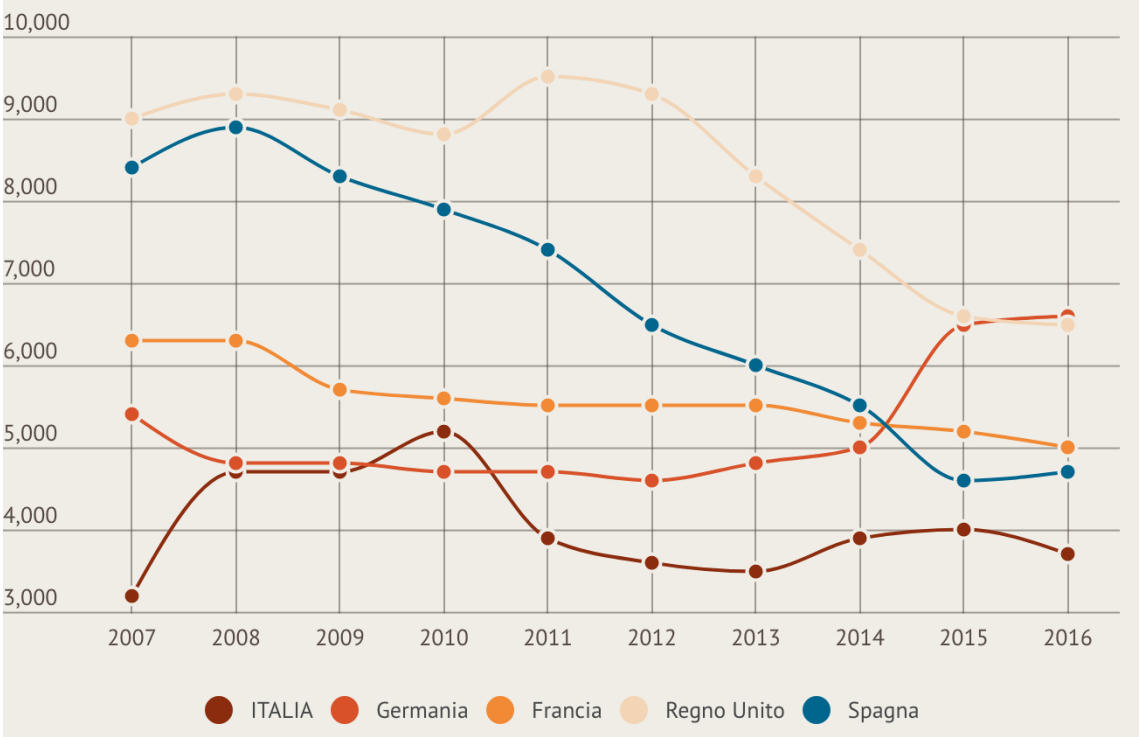


Figure 2.3: Estimated incidence of tuberculosis cases between 2007 and 2016 [7].

and 300 occurred in pediatric age. Of the total number of patients, 70% presented pulmonary tuberculosis. The estimated number of deaths per year is 330, or nearly one death per day.

2.3 Etiology and Pathogenesis

TB is a primarily airborne infectious disease caused by M. tuberculosis. Infected individuals disperse germs into the air via saliva droplets from sneezing or coughing (approximately 3000 bacciferous droplet nuclei per cough) [8, 9].

Intensively exposed individuals face the highest risk of acquiring TB infection to prolonged periods in closed, overcrowded, and poorly ventilated environments.

This increases the likelihood of infection because these germs can remain suspended

for several hours and then be inhaled through the airways.

TB is a potentially severe disease and, if not treated properly, can lead to death. If the infection goes untreated, each person with active TB infects, on average, 10-15 new individuals each year (Global WHO TB report 2016). However, anyone with a compromised immune system, such as people with diabetes and or HIV, has a much higher risk of becoming ill [10, 11, 12]. In particular, TB is a disease strongly associated with the conditions in which people live. Therefore, the lowering of immune defences may depend on living in deplorable hygienic conditions and suffering from a state of malnutrition and poor health.

After exposure, about 1-2% of people affected immediately develop active TB, also called "primary TB"; 1/3 develop latent-type TB infection (LTBI), typical of most individuals affected by TB; 2/3 do not contract the disease.

Among individuals with latent infection, only 5-10% will develop the active form of TB. The risk of contracting TB is related to age, which is one of the essential factor in the disease's evolution. It can affect people of all ages, but the risk of getting tuberculosis again is about 15%. Infection is more likely to develop during:

- childhood (risk of becoming ill is between 30% - 50%);
- late adolescence;
- early adulthood.

The immune system encountering M. tuberculosis bacilli can trigger two possible scenarios; first, the bacillus can be killed instantly by the innate immune response [13]. Secondly, out of approximately ten persons infected with tuberculosis, one may develop an active infection within a period of 1 to 3 years. In the latter, the immune system can be probably not avle to control early infection or enabled to acquire a

protective response in time to prevent disease.

Preventive drug treatment for individuals already infected and at high risk of relapse can significantly decrease the likelihood of developing an active TB form.

2.4 Clinical Manifestations

Any organ in the human body can be affected by tuberculosis. Specifically, pulmonary manifestations are the most frequent in HIV-negative individuals (70 - 80% of cases). Pulmonary and extrapulmonary TB can occur many years after an individual's exposure to the infectious agent and can be caused by temporary or permanent immunity impairment. Only on rare occasions, individuals develop symptoms soon after the primary infection.

However, not all infected people develop the disease; in fact, the immune system can fight the infection, and the bacteria can remain dormant for years. This specific condition is called LTBI, and about a quarter of the world's population is affected [14, 15]. People with a latent infection have no symptoms and are not contagious. It has been estimated that about 5-15% of people with latent infection develop the disease in an active form in their lifetime.

The typical main symptoms of active pulmonary tuberculosis are fever, weakness, chest pain, weight loss, night sweats, and coughing with sputum and blood sometimes [16]. These symptoms can be mild for months.

In contrast, if the tuberculosis is extrapulmonary, the symptoms depend on the site involved.

2.5 Diagnosis

Early diagnosis of tuberculosis has positive repercussions on the recovery of the affected person and the prevention other people. Symptoms and signs of tuberculosis depend on the disease location and extent, and vary according to the development stage of the disease. There are several tools for diagnosis, including:

- *Mantoux Test or Tuberculin Skin Test (TST)*: represents the most common diagnostic tool and consists on a simple skin test. A small amount of a substance called purified protein derivative (PPD) of tuberculin is injected just under the forearm skin. After 48-72 hours, the patient should undergo a forearm examination by a health care provider to detect, evaluate and determine reactions occurring at the tuberculin inoculation site, i.e., the appearance of swelling erythematous patches. However, what must be analyzed is not so much the extent of the red patch but the diameter of the dermal induration at the injection site to establish the diagnostic accuracy of the result. Based on the diameter of the inoculation site, it is possible to evaluate if the test is positive or negative according to the following criteria:
  - no hardening: negative test;
  - hardening with diameter less than 2 mm: negative test;
  - hardening with diameter between 2 and 4 mm: doubtful test;
  - hardening with a diameter of 5 mm or more: positive test.

  A new examination is usually required after a couple of months in a negative test and a doubtful test. The advantages of the Mantoux test are represented by the fact that is an inexpensive test and at the same time easy to administer.

Otherwise, the disadvantages include the need for a demanding reading time (strictly dependent on the operator) and the potentiality that the outcome depends on the temporary conditions of patient's natural immuno-repression, any pharmacological inductions or immuno-activation.

- *A blood test or interferon-gamma test*: can be used to obtain confirmation of diagnostic suspicion. These tests use sophisticated technology to measure the immune system reaction to Mycobacterium tuberculosis and provide a faster and more accurate result in comparison to the Mantoux test.

- *Instrumental examinations*: if the preliminary tests are positive, further assessments to determine the disease stage (active or latent) are needed, along with the evaluation of mycobacteria resistance to antibiotics. One of the most common tests, given the lesions primary localization, is the chest X-ray. This latter can highlight the presence of small white spots in the lung regions where the immune system has confined the pathogens (in an inactive form).

2.6 Treatments

The discovery of the first antitubercular drugs in the 1940s, together with the improvement in the population of socio-sanitary conditions, raised rosy prospects for eradicating tuberculosis. The pharmacological treatment of tuberculosis is quite complicated.

Tuberculosis can be treated with a panel of four standards (first-line) medications. There is usually a two-month intensive treatment phase using all medicines, followed by a four-month continuation phase with only two. Due to the too-long treatment and hepatotoxic effects, many people do not finish the course of medication, and as a

result, resistance to TB drugs can be developped. Therefore, following WHO guidelines published in May 2016, all rifampicin-resistant TB cases (RR-TB), including those with multidrug-resistant TB (MDR-TB), should be treated with second-line MDR-TB treatment. Thus, the treatment duration is much longer for MDR-TB (at least 9-12 months) than for drug-sensitive TB (between six and nine months), with a significantly higher risk of adverse drug reactions and unsuccessful treatment outcomes, particularly death.

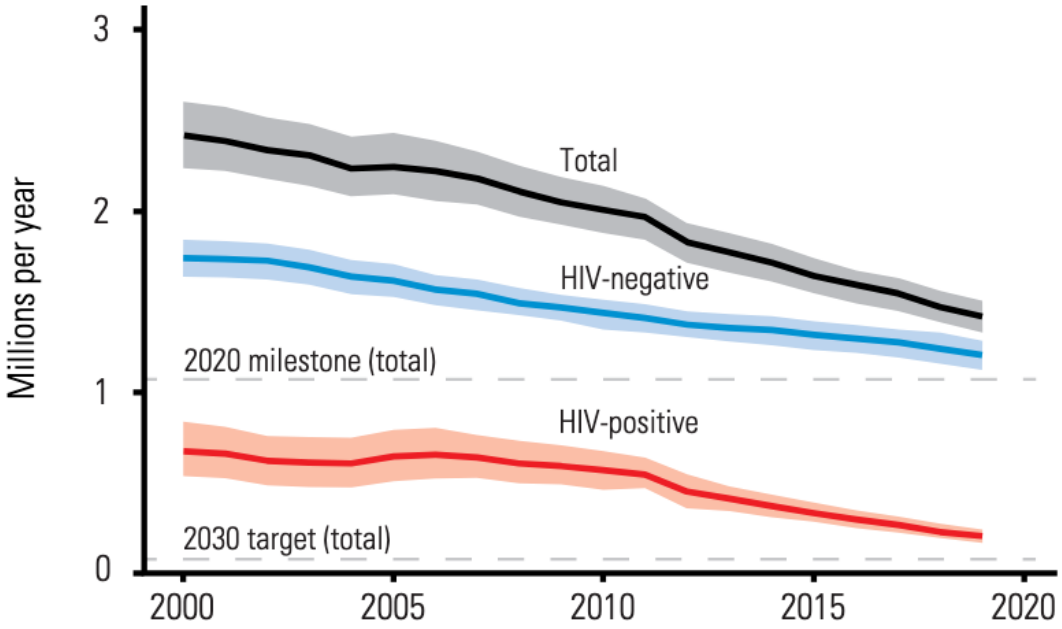


Figure 2.4: The global trend in the estimated number of TB deaths, during the period 2000–2019. The shaded areas represent the uncertainty intervals. Horizontal dashed lines highlight the 2020 milestone and 2030 target of the End TB Strategy [4].

The outcome of MDR-TB treatment is poor. Within the MDR-TB cases worldwide that started therapy in 2010, only 48% had a favourable outcome. Preventive treatment of chemoprophylaxis is essential to avoid the worsening of tuberculosis

from the latent phase to the active one. In this sense, isoniazid is particularly indicated in patients with AIDS and infants. Chemoprophylaxis therapy should be performed in patients with previously untreated tuberculosis, significantly when immunocompromised. Among chemoprophylactic drugs, isoniazid is used in treating both pulmonary and extrapulmonary tuberculosis.

The current prophylactic vaccine against human tuberculosis is Bacillus Calmette – Guérin (BCG) vaccine, developed almost 100 years ago, and able to prevent severe forms of tuberculosis in children. However, no vaccine is currently effective in preventing tuberculosis disease in adults, either before or after exposure to tuberculosis infection [17].

In 2014, the World Health Assembly adopted the WHO Tuberculosis Strategy to eliminate the global TB epidemic by 2035, reducing TB cases by 90% (compared to the 2015 baseline) (Figure 2.4). The achievement of this goal requires a comprehensive approach including new and more effective vaccines and improved diagnostics and treatments. Vaccines are considered the most effective and cost-effective means of addressing the disease. For this particular disease, multiple vaccine development strategies are being pursued:

- infection prevention: vaccines are administered before exposure to MTB to prevent initial infection and therefore disease;
- disease prevention: after exposure to MTB, vaccines are administered to infected individuals that are asymptomatic and with a medium-high risk of developing the disease in the future. The main objective is to prevent the manifestation of active disease and thus, reduce transmission.

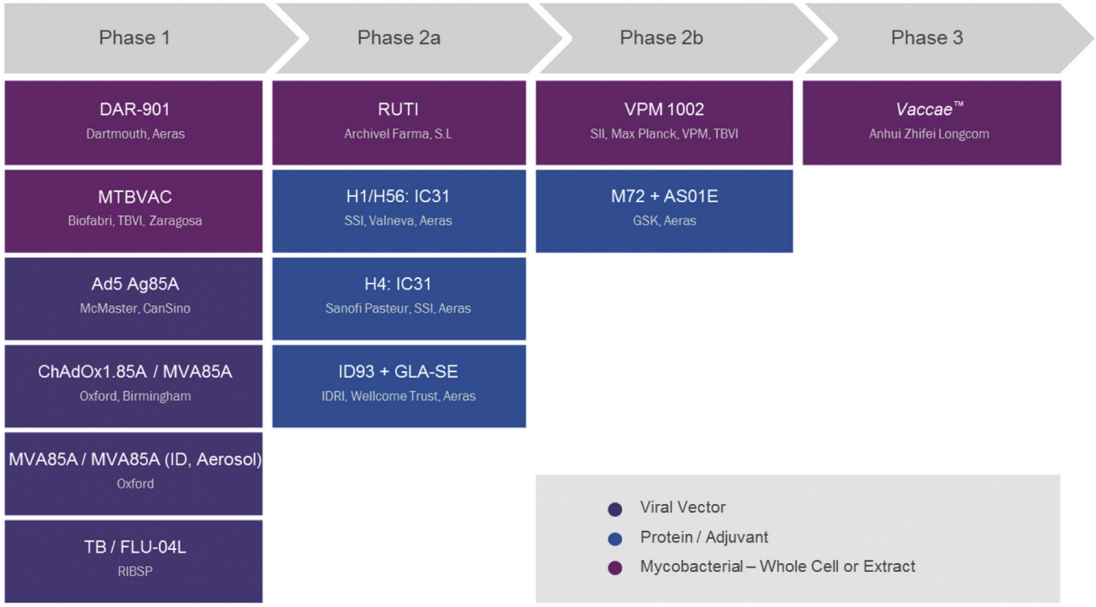


Figure 2.5: Global Clinical Pipeline of TB Vaccine Candidates (information updated to December 2015).

- relapse prevention: vaccines are administered after MTB infection and anti-tubercular drugs administration to prevent reactivation and subsequent transmission of the infection [18].

In addition to classical approaches, immunotherapeutic vaccines are being developed for individuals with active TB combined with TB drug therapy to shorten the duration of treatment and reduce relapse rates after the completion of treatment. The RUTI® vaccine is one of the most advanced developed therapeutic vaccine against Mdr-TB. Figure 2.5 shows the development pipeline for the new candidate TB vaccines [19, 20, 21].

Chapter 3

Agent-Based Model (ABM) paradigm

3.1 Simulation and modeling of natural phenomena

Simulation is an essential instrument for studying physical and natural phenomena allowing, beginning from the actual state of knowledge, the determination of specific possible future condition through a mathematical model. Numerous physical and biological processes are still described using deterministic laws. Using these laws, it should be possible to obtain a univocal and predictable result from an "exact description" of a natural phenomenon. In 1960, Edward Lorenz [22] conducted a study on the simulation of weather forecasts to demonstrate the interactions between the main meteorological phenomena. By slightly varying the input parameters for the determination of similar models to those already found, Lorenz unexpectedly obtained completely different results from the previous ones. This led him to realize that "complex dynamic systems" often produce significant effects on minor changes dealing with the starting conditions and laying the groundwork for what will be

later described as "Chaos Theory" [23, 24, 25].

In the modeling approach, deterministic laws are used, but the results obtained are unpredictable, in other words, not known a priori. The apparent contradiction can be quickly solved because the sensitivity of a phenomenon, in comparison to its initial conditions, plays a fundamental role. Indeed, the description of the current state of a phenomenon is always subjected to approximations and limits of precision. These limits can compromise the deterministic nature because they cause errors and uncertainties within the results. This is particularly evident in the study of susceptible phenomena. In this case, even small approximations of the input lead to specific behaviors; the latters, if described with finite precision, show unpredictable results of non-deterministic and therefore, chaotic nature. Researchers are used to formally describe the phenomenon they would like to study and also own essentially two needs:

(a) to understand how the current observed situation is;

(b) to be able to predict what will happen in the future.

From this point forward, a mathematical model for the physical or biological phenomenon is needed to describe, as faithful as possible, the overall behavior. The modeling process, therefore, aims to provide this mathematical model for the studied phenomenon.

The main steps in the process of modeling include:

- the targeted study to obtain the set of characteristics (e.g., existence, uniqueness, regularity) that the solution owns within the dynamic system [26];

- the identification of an algorithm able to provide an universal method for calculating such solution or, more generally, to obtain information about the solution in a possibly efficient and complete way;

- the conversion of the algorithm into a programming language and its execution.

One can notice how the use of a computer assumes considerable importance within the whole procedure. It allows obtaining an approximation of the phenomenon object of the study results, avoiding the continuous searching of the real solution. This latter may be, computationally speaking, too costly or not physically feasible. There are, however, many issues to be considered. Firstly, the algorithm used must be "usable", in the sense that it must require a reasonable amount of memory resources and computation time because, in such complex problems, they play a fundamental role. A summary scheme of the entire process is provided in Figure 3.1.

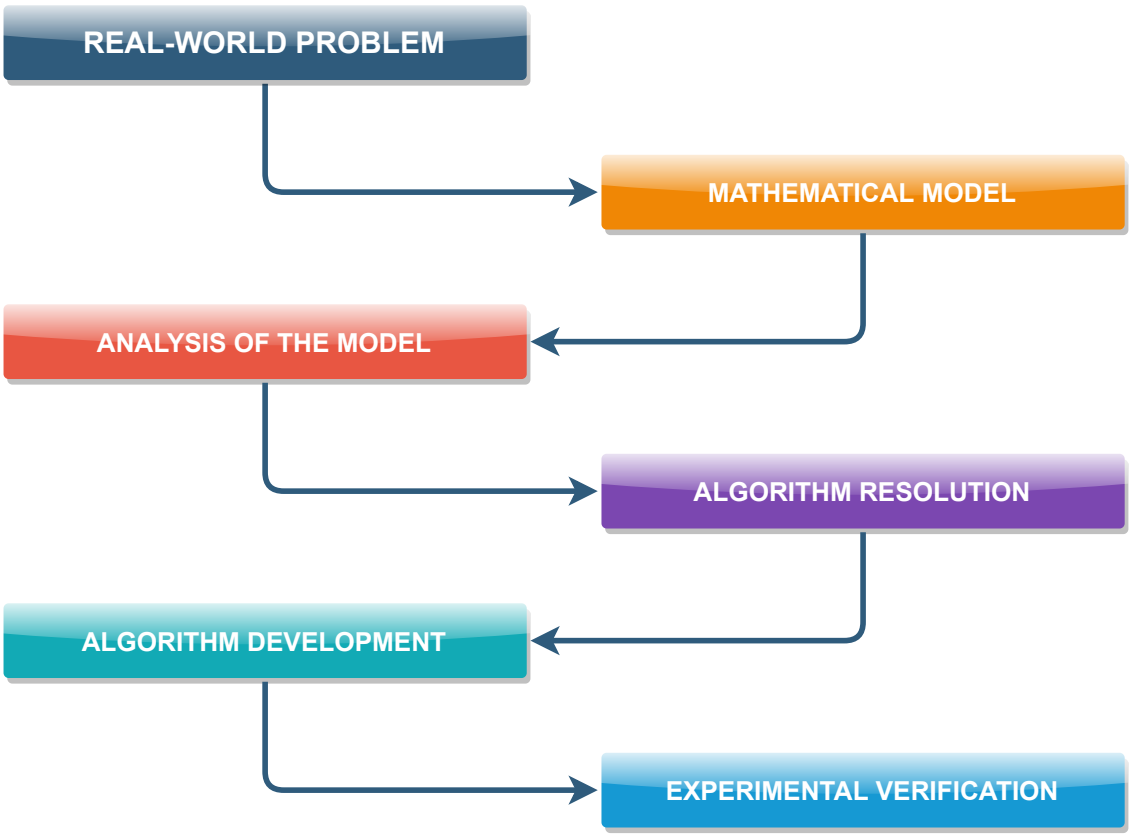


Figure 3.1: The logical scheme from the real world-problem to the experimental model.

The approximate results must deviate from the valid values in a non-significant way to ensure usability. It is necessary to limit the presence of errors as much as possible. To do this, it is important to use an algorithm that is "stable", in other words, that does not amplify errors during calculation. The problem lies in the fact that errors are already introduced in the modeling phase, especially when it pass from the physical phenomenon to the corresponding experimental model. It should be noted that some natural phenomena are so complex that it is almost impossible to create a mathematical model that entirely describes them. When using a mathematical model is feasible, the excessive complexity of the model will not make it more accessible to find a numerical solution. Moreover, it is often necessary to

consider a simplification that describes only some aspects of the problem, leaving some variables out.

Further truncation and rounding errors are introduced in the mathematical operators during the numerical representation phase; other errors can be introduced when the input are inserted into the computer memory. Error analysis is almost entirely outside the typical nature and characteristics of the phenomenon under investigation. In addition, a chaotic phenomenon remains unpredictable and occurs independently from the mathematical model chosen to describe the phenomenon and the available degree of precision.

3.2 Modeling and simulation in biomedicine

Within the field of biomedical phenomena the association of a mathematical model to the experimental process is now widespread practice, acquiring a bidirectional nature. This means that information is mutually exchanged between the two systems to understand natural phenomena in a better way. This is particularly evident when observing the ideal research process [27]. From observing a particular phenomenon in real patients, the researcher attempts to develop biological, in vitro, and in vivo models that lead to an easier study of the phenomenon. The analysis of the properties of the solution can then provide important information about the occurring dynamics within the phenomenon. The resulting data are then used for the simulations of the *in silico* models.

Finally, the obtained results are tested on multiple experimental steps of increasing complexity and applied to clinical practice. This modelling process is shown in Figure 3.2. The mathematical model can accelerate the entire process, helping to

reveal potential hidden mechanisms of the phenomenon which are difficult to catch through ordinary experimental studies, which also are excessively long and costly, or impractical. Consequently, such a biomedical modelling process suggests new therapeutic interventions and provides significant refinements for the experimental protocols, identifying in advance the most promising pathways from a clinical perspective. It is worth noting that the evolution of many natural phenomena, such as tumor growth, is incredibly complex. Often, to obtain an exhaustive analysis, it is not sufficient to observe the phenomenon, but it is necessary to use different scale levels. The description of the phenomenon is, in fact, strongly related to the degree of the scale to be used [27].

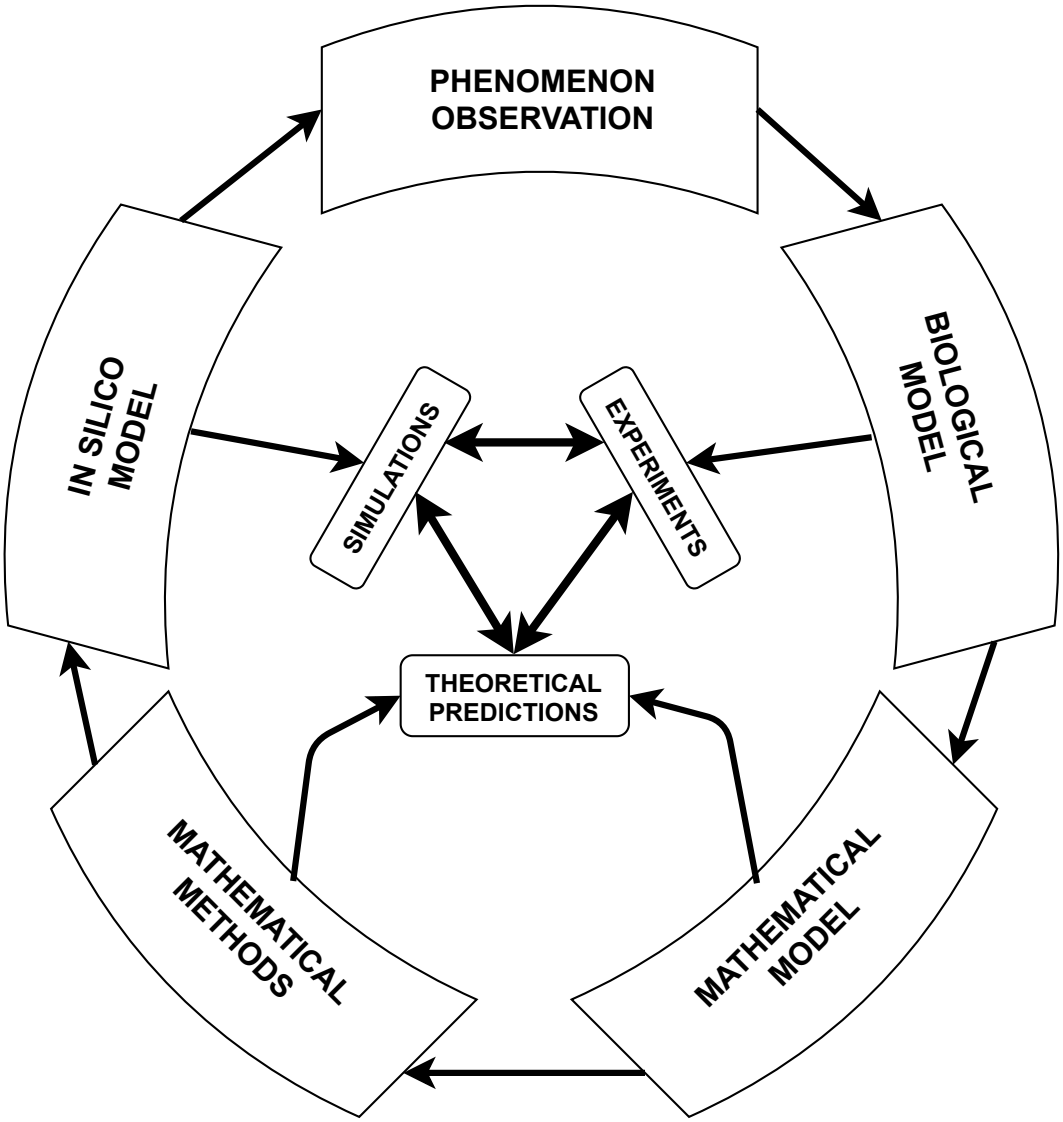


Figure 3.2: The biomedical modeling process.

Specifically, in the analysis of a biological and biomedical phenomenon, researchers should consider three natural levels of scales (as shown in Figure 3.3):

1. **Sub-cellular level**: it describes specific mechanisms and phenomena a on a microscopic scale that occur within the cell and its membranes, such as

receptor activation, macromolecule synthesis, degradation of DNA sequences, and expression of specific protein sequences on the cell membrane.

2. **Cellular level**: it describes the interactions between the cell population on a mesoscopic scale, such as the interaction between tumor cells and other ones such as macrophages and lymphocytes.

3. **Tissue level**: it provides a macroscopic view of typical continuous phenomena such as the migration of sets of cells or the main biochemical reactions for the diffusion of substances within the organism.

In particular, the different levels describe the same phenomenon observed from different points of view. The variation of a phenomenon in one level is directly influenced by the other ones, and affects the same phenomena in the other levels.

Therefore, a multi-scale analysis can provide a more detailed and complete description of the phenomenon, facilitating the modeling of the biomedical phenomenon itself.

3.3 Mathematical models in immunology

Computational methods currently used for immunology modeling are classified into:

- **Continuous models**
- **Discrete models**

Continuous models use differential equations to describe the system under investigation and can be used to describe, for example, cell populations. These methods are called continuous methods for the use of actual variables. However, the values of the phenomenon can be considered as discrete ones, such as the number

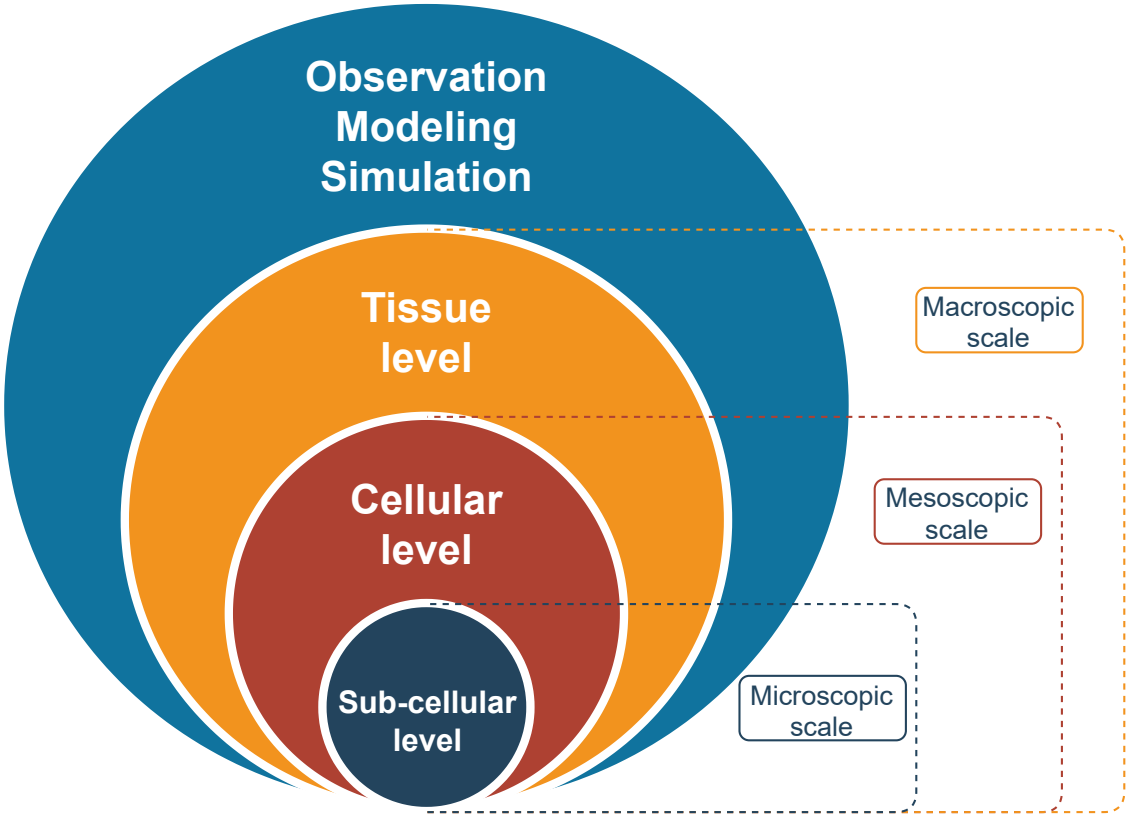


Figure 3.3: Representation of the potential levels of scale in the observation of biomedical phenomena.

of cells. These models own several advantages. Firstly, real parameters are used, there is no measurement limit, so asymptotic extrapolation is straightforward. In addition, differential equation theory owns a very solid mathematical basis. However, they can result in a complicated method to be used due to the many cases of non-linearity widely involved in biological systems. Furthermore, in continuous model the abstraction is much greater than for discrete systems, and the approximations introduced could not have apparent biological significance. Oppositely, discrete models allow the interaction among entities and they can be described in a defined space. One of the advantages of the discrete models is that a biological

scenario could be described precisely. In addition, the non-linearity does not represent a problem. Complex cases can require more time for computation but do not prevent their solution. System size is generally small compared to the reality, and consequently, one should pay close attention to those problems arising from the finite size. In addition, asymptotic extrapolation is often difficult or even impossible. It should be pointed out that both type of models are not necessarily mutually exclusive. Their common use can help to describe better the biological phenomena. In particular: i) continuous models have been used for the description of phenomena on the macroscopic scale, since they have behavioral connotations typically constant while ii) the discrete models are suitable for typical phenomena on the cellular and sub-cellular scale, allowing to describe their behavior accurately.

3.3.1 Continuous models

In continuous models, each population is characterized by a certain fixed activity, and intracellular interactions modify the number of cells related to that specific activity. Large numbers represent these populations. It is justifiable to consider their concentration in space as a continuous variable. Considering the distribution of the elements in space as a homogeneous solution in which interactions occur by random collisions, it is possible to use a system of nonlinear differential equations to describe their behavior. In this case, the equation system shows the following structure [28]:

$$x(t) = gain–loss \tag{3.1}$$

Where $x(t)$ represents the change in concentrations of $x = xi$ over time, and semi-linear or nonlinear terms express the gain or loss of molecular complexes due to

stimulation of cells or new complexes (cell-cell or cell-molecule). In contrast, "the gain" of new cells or their "loss", due to end-of-life, is expressed by linear time variables. A trend (or curve) represents the solution of $x(t)$, and the set of curves formed at different initial conditions describes a flow. If x is stationary ($x = x_0$), it is called as a fixed point. Stable fixed points, the one for which a small change in the variables over time decreases to zero, are referred to as flow attractors. After looking for flow attractors, the overall behavior can be analysed numerically.

3.3.2 Discrete models

As previously mentioned, the discrete models present numerous characteristics that make them suitable for modeling and simulating biological phenomena. Generally, the mathematical knowledge required does not affect the modeling, so the approximations made are generally more biological than mathematical. Finally, all cases of specific phenomena can be described easily, making a close correspondence between the model and the biological phenomenon. One of the most important examples of a discrete model is a cellular automaton (CA). Nowadays, CA is considered one of the most influential paradigms for studying complex systems. Along with neural networks and genetic algorithms, CA forms a set of methodological supports indispensable for scientific investigation. It is worth to mention that CA is a discrete model both in space and time. All the rules that constitute the transitions from one state to another one are applied simultaneously in the same "time-step". The cellular automata can be briefly defined as a set of cells arranged on a regular lattice in which the following properties are applied:

- the state of a cellular automaton is determined by the contents of all the cells located into the lattice in a given time-step;

- the contents of the cells can contain a finite number of states according to deterministic local rules;
- a "local rule" is similar to a function defined for all the cells that accept as arguments the value contained in the cell itself and the ones belonging to neighboring cells, returning a new content for the cell;
- the update of the cell contents takes place simultaneously.

From the structure of the CA, two fundamental properties can be inferred. Firstly, a CA is a "parallel" computing model that can be appropriately applied to new systems. Secondly, in a CA model, agents interact with each other depending on local information and then, they pursue global objectives.

3.3.3 Cellular automata

Cellular automata represent spatial models [29, 30, 31] based on relatively simple notions about the effect of space and related dynamics. In their most severe and complex form, they simulate spatial diffusion around a certain point where the dimensions of time and space are treated as a whole.

The temporal scale level and spatial aggregation show the tendency to be quite flexible in these models, although the scales tend to be significant. Moreover, these models are shown to be better adapted to the study of specific dynamics rather than others. Such models do not belong to a specific class of dynamics, due to external inputs, and they do not depend on the dynamics implemented within the model [32]. Agents do not physically move in space but can be associated with different locations, and their change over time may reflect an implicit process of movement [33]. A typical CA is a two-dimensional grid or lattice subdivided into more cells.

Each cell assumes a finite number of states at any given time and is characterized by a value determined by a set of simple rules based on the previous state.

3.3.4 A Cellular Automata for the immune system: the Celada-Seiden model

The aim of the Celada-Seiden (CS) model [34, 35, 36] is to create a system in which the "hypotheses" accurately reflect the immune phenomenon and can be tested through the interactions between its elements. Therefore, this model does not take into account the modeling through continuous systems and opts for modeling through discrete systems due to specific reasons.

Firstly, the discrete model allows the representation of all the simulation elements through a language with more biological than mathematical feature.

Besides, it becomes simpler to modify the interactions between the system entities without adding new and more complicated equations, hiding the practical complexity of the continuous models shown to the naked eye.

However, it should be noted that the actual complexity of the immunological phenomenon currently makes it impossible to implement on a computer a "large scale" representation of the phenomenon analyzed. The number of possible antigens that can be recognized, and therefore, the potential receptors that the system may contain, have been estimated theoretically to a value in the order of 10^{20}. The human organism under conditions of extreme activity can contain up to 1012 specific cells. To simulate such an event, the only solution is to consider a small sample of a population whose size can reach considerable proportions.

For this reason, the results obtained cannot replace in vitro and in vivo experiments but, on the contrary, should be considered as proper support for them. In particular,

the CS model bases its structure on the model of cellular automata. However, there are significant differences regarding the definition of cellular automata described above:

1. In the CS model, the rules are not always deterministic but probabilistic.
2. The value of a local cell is determined only by taking into account the cell itself and not considering the cells in the neighborhood.
3. The entities can spread from one cell to another one.

The model uses a bidirectional grid that represents a small portion of the organism. The cells are triangular in size instead of square, thus bringing the number of neighboring cells from eight to six. Instead of containing a binary value in various cellular automata implementations, each cell can be populated by several distinct entities of different types. It was also chosen to address the most demanding problem encountered during the modeling process, namely the description of the receptor, in the simplest way, that is, using a string of bits. This offers the possibility to move within the set of possible receptors affordably and straightforwardly at the memory level. A few bits are sufficient to represent a discrete range of possibilities, directly proportional to the number of bits used.

3.4 Definition of the Agent-Based Model concept and its essential features

Agent-based models (ABMs) are a class of computational models aimed at computer simulation of specific actions and interactions of autonomous agents to evaluate their effects on the system as a whole. ABMs have, to some extent, evolved from cellular

automata (CA).

ABMs can combine elements of game theory, complex systems, emergent behavior, computational sociology, and multi-agent systems. A recent literature review on individual-based, agent-based, and multi-agent systems models shows that ABMs are used in non-computationally related scientific domains such as biology, ecology, and social sciences [37]. Agent-based modeling is related to but distinct from the concept of multi-agent systems. ABM aims to seek explanatory information about the collective behavior of agents following simple rules, typically in natural systems. Multi-Agent Systems (MAS) aims to design agents or solve specific practical or engineering problems [37]. Moreover, agent-based models are a type of microscale model [38] that simulates simultaneous operations and interactions of multiple agents to recreate and predict the appearance of complex phenomena. The process deals with an emergent behavior from the microscopic level to the macroscopic one. Regarding the topic above mentioned, an essential notion should be introduced: simple behavioral rules generate complex behaviors, as enunciated by the "Keep it simple, stupid" (KISS) principle, adopted extensively in community modeling. Individual agents are typically characterized as rationally constrained, presumably acting according to what they perceive to be their interests, such as reproduction, economic benefits, or social status, using heuristics or simple decision rules. ABM agents can experience "learning," adaptation, and reproduction processes [39]. Agent-based models typically consist of:

- numerous agents defined at different scale levels;
- heuristic principles aimed at decision making;
- learning rules and adaptation processes;

Software	***Scope***	***License***	***Progr, Language***	***GIS***	***3D***	***Link***
UISS	Immune System dynamics	Closed-source	C	No	No	https://www.combine-group.org/software
Netlogo	Social, Natural, and Economic Sciences	GPL	Netlogo	No	No	http://ccl.northwestern.edu/netlogo/
BioDynaMo	Dynamic research purpose	Open-source	C++	No	Yes	https://biodynamo.org/

Table 3.1: The table shows just a few examples of the available simulators.

- topological system in which interactions occur;
- a specific environment.

ABMs are typically implemented in the form of computer simulations through specific programs or ABM development tools. Such models can be used to evaluate the impact on the emerging behavior of the system as a result of specific changes in individual behavior.

3.5 Typical ABM structure

The logic characterizing agent-based modeling has been defined in detail by two authors. The first definition given by Ferber [40] deals with MAS but can also be extended to the ABM approach. Ferber identifies five key components:

1. **the environment (E)** that represents the space of interaction of the system components;
2. **a set of objects (O)** with different characteristics associated with the environment (E);
3. **a collection of agents (A)** representing a subset of the objects and the active entities of the system;

4. **a set of relationships (R)** connecting objects and agents;

5. **a group of operations (Op)** that allows the agents to perceive, produce, transform, and manipulate the system objects.

The second definition was proposed by Macal and North [41, 42], who reduces the system components to three elements and focuses on the concept of Complex Adaptive System (CAS). In particular, the complex system consists of autonomous agents that interact with each other and adapt themselves to the variations produced by the system.
According to Macal and North, a typical agent-based model consists of three elements:

1. **a set of agents** characterized by their attributes and behaviors;

2. **a set of relationships** defined by specific rules that describe the topology of the interaction between the agents;

3. **the environment** that represents the place of interaction of the agents.

Although the two definitions diverge in some of the components that Ferber formalizes oppositely to North and Macal, they both agree on the presence of the agents, the environment, and the interactions between the agents and the environment. Focusing on the first of these three elements, the agents need to investigate the characteristics that distinguish their functioning with the environment and among the agents. In the next paragraph, the necessary factors an agent must incorporate to coordinate all the system agents involved will be defined. Three main features must be described for managing the interaction between the agents:

1. **Decision making**: what kind of decision-making system can the agents take? What is the logic behind their perception-representation-action?

2. **Communication**: what type of message the agents receive from the environment and the other agents? What is the formal communication protocol used?

3. **control**: Is there a hierarchical relationship between the agents?

3.5.1 Agents features

Two of the main issues that need to be addressed when applying agent-based methodologies are:

1. the interaction between the agents;

2. what other agents can interact with the agents themselves;

3. which dynamics trigger these interactions;

One of the paradigms underlying the agent-based modeling approach is that only local information is available to agents. The agent-based system is decentralized. No central entity globally disseminates the same information to all agents or controls their behavior to increase overall system performance. Agents interact with each other, but they do not interact all at the same time. The interaction of the agents typically happens through a subset of agents close to each other (neighbors). Likewise, the interaction between the agents and the environment is localized. The agents do not interact with each part of the environment. Therefore, the information obtained from the agents is derived from the subset of agents with which they interact and the portion of the environment in which they are located. If the model

is not static, the subset of agents and the environment vary over time. The main feature an agent should own is its ability to act autonomously. In the literature, there is no precise definition regarding the autonomy characteristic of the agent. The concept of autonomy can range from a simple passive behavior, consisting of reacting to external events that respond to a basic logic of the "if-then" type, to a more complex behavior constituted by models of adaptive artificial intelligence. One of the available definitions is the one suggested by Casti [43], which provides the presence within the agent of two levels of autonomy:

- **Basic level**: determines the behavioral rules of the agents;
- **Advanced level**: includes the agent ability to change its internal set of rules.

From a modeling-oriented point of view and considering the main practical implications of agent-based modeling, the agents own the following essential characteristics:

- **Self-sufficiency, modularity, and identifiability**: these properties specify the boundaries of the agent, the components that are part of the agent and those that are not, and the presence of shared attributes. The attributes of an agent make it identifiable (unique) for possible interactions with other agents.
- **Autonomy**: an agent acts independently of its interactions with other agents and the environmental context in which it is embedded (at least in a limited range of situations that are of interest to the model). The information rules the agent behavior acquired through the interactions with other agents and the environment. The behavior of an agent can range from a simple set of logical rules to more complex behavioral systems based on abstract models.
- **Variation of the state in time**: the state of model component variables determines the state of a system. In particular, the state of an agent is composed

of the variables associated with it at a given time. These variables are referred to as the attributes of the agent. As for the single-agent that contains it, the model state is determined from the total state of its components: agents and environment in an ABM. The agent state determines its behavior. Therefore, the more significant is the variety of conditions in which an agent can be found, the greater are the behavioral choices that this agent can make. In ABMs, the state represents the specific set of information necessary to modify the system, evolving it from one point to the successive one.

- **Social Dynamics**: agents own a specific social dynamic that manages the interactions between them. Each agent must use a standard communication protocol to enable communication, movement, response to environmental stimuli, and other behaviors. All agents within the system must know the chosen protocol.

North and Macal, in addition to the above essential characteristics, identified other valuable features that agents should present:

- **Adaptability**: an agent should adapt, in other words, its own rules or abstract decision-making models capable of modifying its behavior.

- **Goal Orientation**: an agent should be oriented towards achieving its internal goal, which does not necessarily correspond to goal maximization, considering the decision-making process. Completing the internal goal encourages the agent to change his behavior if it registers a deviation between the current situation and the achievement of the goal.

- **Heterogeneity**: agents can be heterogeneous with each other. Each agent owns several attributes that define its behavior. Behaviour is determined by

the agent ability to retrieve and process information. The agent processes information depending on the behavior developed to interpret the world around it and its previous experiences.

As one can notice, the agents show different attributes that can extensively vary the behavioral characteristics of an agent compared to another one; then, in an equally significant way, also the reactions to stimuli from the external environment can vary.

3.5.2 Interactions among agents

One of the assumptions at the basis of agent-based modeling, is given by the modality through which the information is made available to the agents. In particular, each agent owns the information it can retrieve locally through neighboring agents. There is no central entity in possession of all the information in the system. It follows that the interactions between agents are also limited. Agents do not interact with all other agents in the system but only with spatially close agents (as previously described in section 3.3.3). During the simulation, the number and type of agents in a specific space can vary because they can move from one space to another. The typology of connection between the agents takes the name of the topology of the model. Therefore, the topology of an ABM refers to what information is exchanged and between whom. Figure 3.4 collects the connection types between the more diffuse agents and the differences between the connections are:

- **spatial grids**: where space is divided into sub-portions of geometric shape and each grid owns specific characteristics (Panel A);

- **representations on a Euclidean plane**: where each agent can move in a plane with 2 or 3 dimensions and the position is determined through the use of specific coordinates on the axes (Panel B);
- **network model**: where the individual interactions between agents are more defined than in the Euclidean plane representation (Panel C);
- **geographic information system (GIS)**: where each area represents a simplification of an actual portion of the territory (Panel D);
- **aspatial representation**: the connection between the agents does not involve the representation through a spatial plane (Panel E).

The abovementioned topologies represent the most used by modelers and it is worth to mention that different connection modalities are often applied within the same model.

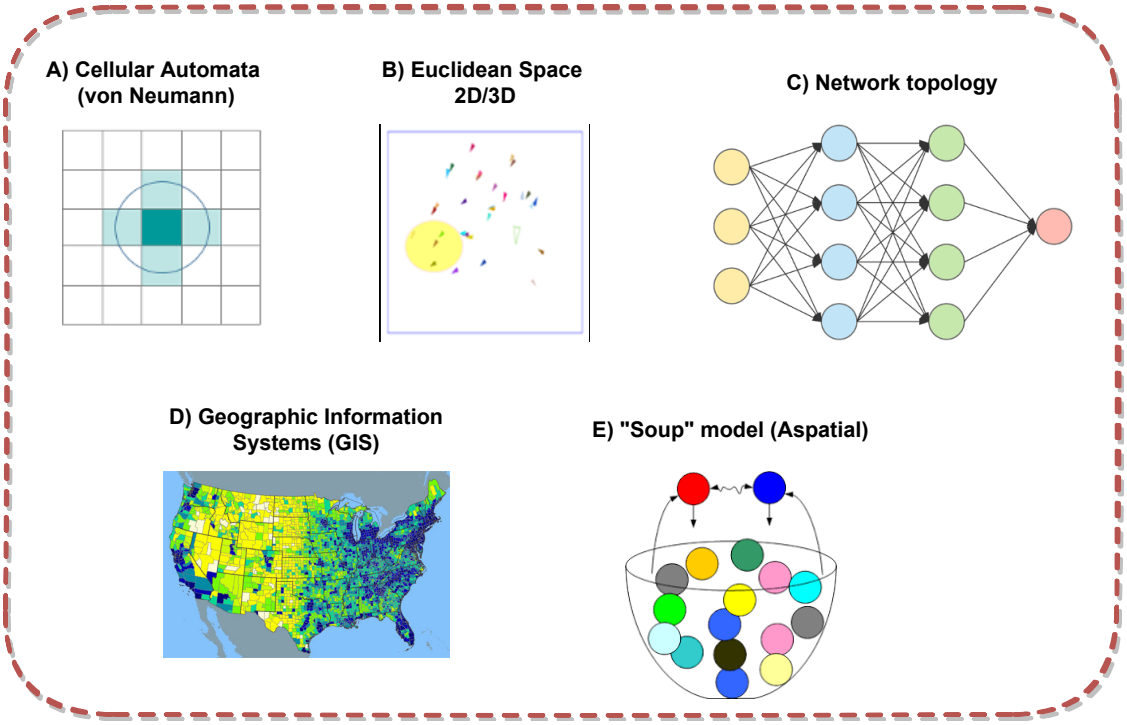


Figure 3.4: Representation of different types of connections between agents.

3.5.3 Environmental representation

In an ABM, agents interact with each other and with the environment. This latter plays a fundamental role in ABMs and can be represented differently, conveying a specific type of scenario.

The environment can be subdivided by a scale that defines the amount of information it possesses, and represented either realistically or artificially. Furthermore, the environment in ABMs can be defined as either cell-based or object-based. Both applications have their limitations and strengths. Cell-based representation is focused on data processing collected directly from the point of interest. The limit of this approach deals with the considerable amount of data to be processed, leading

researchers to use robust software and hardware to process the data collected. According to Bian [44], the cell-based system should be used when the environment owns specific features of marked heterogeneity and complex dynamics.

The representation of the environment through objects requires fewer computational resources, but allows the access to a great deal of information (Figure 3.4 - Panel D). The object-based representation is particularly suitable when many landscape features need to be included within the model. Furthermore, it is indispensable to capture the relationships between the various landscape features [44]. The last distinction in the representation of the environment deals with the fact that the environment could be static or dynamic. A static environment finds its use within ABMs when one aims to investigate the emerging characteristics of a system, due to the agents interactions within a specific environment. While, a dynamic environment is more suitable for studying the effect of change on the space in which the agents are located.

Chapter 4

A framework based on ABM: Universal Immune System Simulator (UISS)

4.1 Introduction

The Universal Immune System Simulator (UISS) is a computational framework that uses a multi-scale, multi-organ, three-dimensional agent-based simulator of the immune system. UISS is entirely written in the standard ANSI C-99 programming language, which allows it to be architecture-independent. The modeling approach used by UISS is the same as the one used by Celada-Seiden (see Chapter 3). Cellular and molecular entities were considered as agents in the model. Such entities can be heterogeneous, have internal properties (life, energy, and other), act and make decisions (move, interact with other agents in their neighborhood, change their internalizing state or die) individually or as a result of interaction with other agents. As mentioned, the cellular entities can take up a state from a specific set of suitable states, and their dynamics are realized through state changes. A state change

occurs when a cell interacts with another cell or with a molecule or both of them. UISS considers the relevant lymphocytes, i.e., B lymphocytes, T helper, cytotoxic and regulatory T lymphocytes, and natural killer cells. Monocytes are represented as well, and UISS also takes care of macrophages and dendritic cells. For what concerns the molecular side, the model distinguishes between simple small molecules like interleukins or signaling molecules in general and more complex molecules like immunoglobulins and antigens, for which their specificity must be represented. At the same level of entities, also immune system activities are implemented. They include both interactions and functions. Functions refer to the primary immune system functions. In particular, UISS takes care of the diversity of specific elements:

- major histocompatibility classes restriction;
- clonal selection by antigen affinity;
- thymus education of T cells;
- antigen processing and presentation (both the cytosolic and endocytic pathways are implemented);
- cell-cell cooperation;
- homeostasis of cells created by the bone marrow;
- hypermutation of antibodies;
- cellular and humoral response;
- immune memory.

In UISS, as in most ABM approaches, time is discrete. This means that all activities in the system are tracked and measured using equidistant time intervals.

An interaction between two entities is a complex action that ends with a change of state of one or both entities. The entities must be "close enough" to interact with each other. Specifically, physical proximity is modeled through the definition of a lattice site. All interactions among cells and molecules occur within a given lattice site at a single time step, so that there is no correlation between entities on different sites at a given time. In UISS, the simulation space can be represented by a $2D$ hexagonal lattice LxL (6 neighbors) or a $3D$ as a cubic lattice $LxLxL$, with periodic boundary conditions or rigid walls on the edges depending on the question of interest.

All entities can move with a uniform probability between neighboring lattices in the grid with an equal diffusion coefficient (Brownian motion). This simulation space is used to represent three anatomical compartments: the thymus, the bone marrow, and a portion of a generic secondary organ. Interactions can be seen as Bernoulli events, so each interaction has a certain probability of occurring. Interactions can be classified as specific or non-specific interactions:

- **Specific interactions**: they involve cells from the adaptive immune system equipped with specific receptors. In this case, the probability p of the interaction will depend on the result of the recognition phase, in which the affinity between the receptors involved plays an important role.

- **Non-specific interactions**: they are those that refer to the use of non-specific receptors. Thus, considering Toll-like receptors (TLRs), they will recognize pathogen-associated molecular patterns (PAMPs) expressed by pathogens with low specificity. These will not be explicitly modeled in UISS; a fixed probability p will be used for all interactions involving the same TLR-PAMP pair.

UISS represents receptors and ligands as bit strings and uses a string-matching rule to model affinity. This clever idea was introduced by Farmer et al., [45] as a way to perform calculations for determining molecular complementarity and predicting the optimal size of an epitope. From immunology, one knows that binding is a threshold effect consisting of two components: the affinity of a single receptor and ligand and the total binding, or avidity of multiple binding pairs. A string-matching rule models the binding by counting the number of positions in the string at which the symbols are complementary (known as Hamming distance). Repertoires are represented in the model as sets of strings. By adopting bit strings, many binding events can be simulated quickly, making it feasible to study the large-scale properties of the immune system. Character strings produce accurate models when benchmarked to experiment, suggesting that the abstraction captures receptor/ligand binding features.

Natural scale and multi-organ ABM simulations, however, require thousands of millions of agents. This represents an issue even for modern CPUs and personal computers. To this end, high-performance computing (HPC) resources are mandatory to reproduce the natural scale behavior of the immune system and related pathologies. Due to the intrinsic nature of the biological and immunological entities that mainly act and interact locally, the simulation of big tissues and/or organs can be split across different CPUs cores. So can leading to separate simulation spaces that can be executed in parallel for most of the time, except for the processes that involve entities migration from an organ to another or occasional movement across adjacent tissues fragments belonging to different simulation spaces. This entitles high degrees of scalability in the function of the number of available CPU cores. The model and its computer implementation are very flexible, and new biological entities and

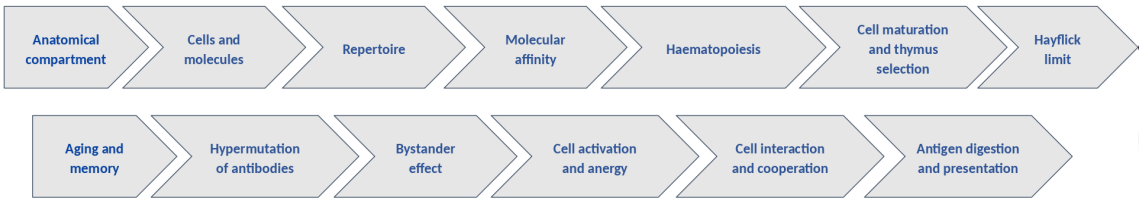


Figure 4.1: Some of the key features implemented in UISS.

interactions can be easily added (Figure 4.1). UISS is then able to reproduce and provide *in silico* predictions about immune system-related pathology.

What is a digital twin (DT) ? A digital twin is a digital representation of a real-world entity or system. Implementing a digital twin is an encapsulated software object or model that mirrors a unique physical object, process, organization, person, or other abstraction. Data from multiple digital twins can be aggregated for a composite view across a number of real-world entities, such as a person and its related processes.

The UISS framework has been used for successfully modeling and simulating multiple immune system-related pathologies. It has initially used to model and simulate the immune system responses to mammary carcinoma tumor cells in naive and vaccinated mice with an immune preventive vaccine [46] (named Triplex), showing the ability to accurately reproduce the experimental results. *in silico* experiments carried out on two large statistical samples of virtual mice underlined that the humoral response is fundamental in controlling the tumor growth and therefore, suggested the selection and timing of experiments for measuring the activity of T cells. The resulting model has been then used in conjunction with two well-known optimization techniques, namely genetic algorithms (GA) and simulated annealing (SA), to suggest vaccination protocols capable of guaranteeing the same survival rates (entitled with the use of a Chronic protocol with the lowest possible number of vaccine

administrations) [47]. GA have been used in HPC environments showing, as a result, a reduction of approximately 50% in the number of vaccinations [48].
Furthermore, SA improved the quality of the suggested solution and decreased the required computational time of two orders of magnitude [48, 49, 50, 51, 52]. The optimized protocol has been tested in vivo [47]. GA techniques have also been applied to optimize the Highly Active Anti-Retroviral Therapies (HAART) protocol to significantly prolong life for people infected by HIV [53]. UISS was also specialized to model the effects of the Triplex vaccine as a therapeutic agent against lung metastases derived by mammary carcinoma [54].
In a further case study, UISS was applied to simulate the main characteristics and dynamics of the immune system activities in digital twins suffering from a relapsing-remitting form of multiple sclerosis (RRMS) [55]. In this case, the simulator revealed that it may potentially assist MS specialists in classifying the severity level of MS at onset and choosing the best treatment strategy [56]. The final goal was to build a comprehensive profile, including also genetics, immunological, and environmental data so to reliably predict the real dynamics of MS at patient level and inform the therapeutic choice at early stages. UISS was also extended to the SARS-CoV-2 scenario. In this disease module, UISS is able to reproduces the fundamental SARS-CoV-2 immune system dynamics. The lung compartment, that represents the main target organ of the virus, along with the generic lymph node, that allows immune system entities to be activated and selected [57] were considered. Moreover, it has been used to predict the outcome of one of the latest suggested approach based on monoclonal antibody.
In the next section, the specific implementation in UISS for the tuberculosis module will be illustrated.

4.2 A specific module for Tuberculosis: UISS-TB

UISS has been extended to include all entities, interactions, and target tissues needed to represent and mimic tuberculosis dynamics and related interaction with the host immune system. It has been applied the methodology described in [58], as it is divided into three different levels. The first level includes the steps involved in the model development:

(i) Identification of the model objectives.

(ii) Collection of the current knowledge about the biological system under investigation.

(iii) Selection of the most appropriate model structure to satisfy the model objectives.

(iv) Translation of goals and knowledge into model hypotheses.

(v) Design and drawing of a conceptual model.

(vi) Identification of the mathematical techniques and development of the formal model.

The second level describes the tuning step, i.e., estimating and fitting model parameters. Finally, the third level deals with analysis and evaluation, i.e., comparing the model results against experimental data sets and analyzing discrepancies.
Our model considers both innate and adaptive immunity (both cellular and humoral) and immune memory. Figure 4.2 depicts all the entities implemented within the simulation framework, especially immune cells, cytokines, and specific biological mechanisms of TB. All the TB disease model entities interact with each other and are

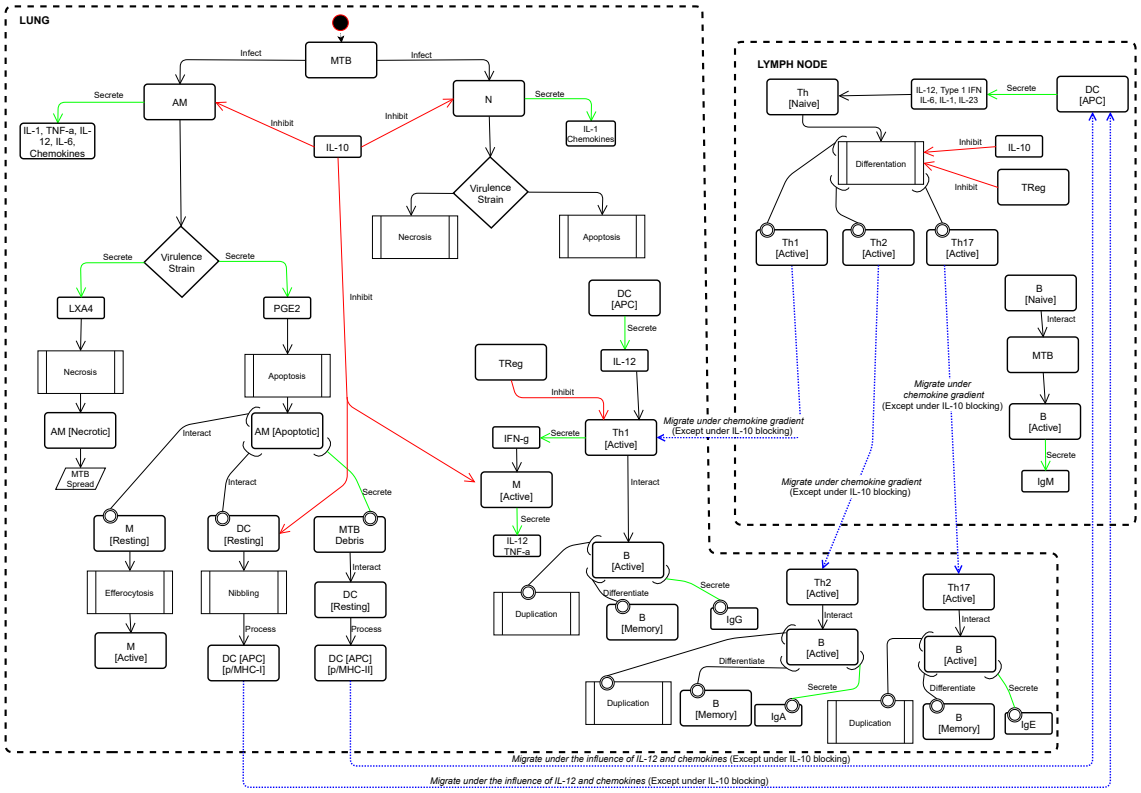


Figure 4.2: Pulmonary tuberculosis – immune system interaction disease model. A conceptual description of the leading entities and interactions of MTB – immune system. The main two compartments are represented: the lung and the peripheral lymph nodes. The representation depicts both cellular and humoral responses when MTB droplets infect alveolar macrophages resident in the lung. The cascade of cytokines and chemokines is also represented with possible different behaviors depending on the virulence of the MTB strain.

appropriately located in two specific compartments: the lung (pulmonary alveoli) and the peripheral lymph nodes. The starting point of the conceptual TB disease model consists of the aerosol droplets of MTB that reach lung alveolar macrophages (AMs) on one side and neutrophils (N) on the other one. The initial challenge of MTB is implemented simulating a virtual injection of MTB bacilli into a lattice point in the lung compartment. Then the bacilli are free to move and disseminate randomly. When an AM becomes infected, it secretes IL-1, TNF-α, IL-12, IL-6, and

chemokines. Depending on MTB strains and their virulence, infected AM can play a different role in determining the downstream pathways leading to the induction of either apoptosis or necrosis and the outcome of the infection. In this context, lipoxin A4 (LXA4) promotes necrosis, while prostaglandin E2 (PGE2) is a proapoptotic factor. When the necrosis process is favored, the AM becomes necrotic and contributes to the MTB spread. Otherwise, when the AM becomes apoptotic, simultaneously, three specific scenarios can occur. Firstly, AM apoptotic can interact with a lung resting macrophage (M) and lead to efferocytosis of macrophages, in other words, an engulfment of AM apoptotic by M, essential for tissue homeostasis and immunity. This means switches from "resting" to "active" status. Secondly, AM apoptotic cells can encounter a lung dendritic cell (DC). AM apoptotic can be taken up by DC that captures antigens (Ag) through a process called nibbling; then, DC will process and present the resulting fragments to antigen-specific T lymphocytes in the context of molecules of the major histocompatibility complex of class I (MHC-I) or related proteins. From this point forward, MTB–antigen processing DCs, migrate to the local lung-draining lymph nodes (by 8–12 days post-infection), driving naïve T cell polarization. This migration is influenced by IL-12 release and other chemokines, except when IL-10 is present and can block this moving. A third interaction dealt with the secretion of MTB debris from AM apoptotic: MTB debris will interact with DCs, in status resting, that will process and present the resulting fragments to antigen-specific T lymphocytes in the context of molecules of the major histocompatibility complex class II (MHC-II) or related proteins. When MTB infects a lung N, N produces and secretes IL-1 and other chemokines. Like AM, also for N, the MTB strain can lead to a different role in the induction of either apoptosis or necrosis and the final outcome of the infection. Both AM and N effector functions can be

negatively modulated by IL-10 induction during MTB infection. Respectively, IL-10 can lead to the inhibition of macrophage and neutrophil effector functions, reducing the bacterial killing and impairing the secretion of cytokines and chemokines. As previously said, IL-10 can also block chemotactic factors that control DC moving to the lung-draining lymph nodes. The scenario inside the lymph node depicts the DC cells in antigen-presenting cell status secreting IL-12, Type 1 IFN, IL-6, and IL-23 and driving naïve T cell differentiation toward a Th1, Th2, or Th17 phenotype. Th cell population differentiation can be negatively modulated by IL-10 and regulatory T cells (Treg).

Protective antigen-specific Th1 cells migrate back to the lungs about 14–17 days after the initial exposure and infection to MTB and under a chemokine gradient (except when IL-10 blocks this process). In the lung, the activated Th1 cell population produces and secrete IFN-γ, causing macrophage activation, relative cytokine production (IL-12 and TNF-α), and bacterial control. It is worth mentioning that in this context, IL-10 can block macrophage activation and consequent cytokine secretion. Also, TReg negatively modulates Th1 population effector functions. For completeness, the Th1 cell population also interacts with B cells leading to three specific processes at the same time. B cells duplicate after a successful interaction, differentiate in memory B, and secrete immunoglobulins type G (IgG). Similarly, Th2 cell migration and interaction with B cells leads to B cells duplication, differentiation in memory B cells, and immunoglobulins type A (IgA). For the Th17 cell population, their migration and interaction with B cells will finally lead to cell duplication and secretion of immunoglobulins type E (IgE). B cells also interact with MTB; after that, B cells become active and secrete immunoglobulins type M (IgM).

4.2.1 Cellular entities

The cellular entities added to the simulator to simulate the dynamics of tuberculosis are:

- **Neutrophils**: Neutrophils are a type of white blood cell. In fact, most of the white blood cells that lead the immune system response are neutrophils. There are four other types of white blood cells. Neutrophils are the most plentiful type within white blood cells, making up 55 to 70 percent of our white blood cells. White blood cells, also called leukocytes, are a key part of our immune system. White blood cells produce chemicals that fight antigens targeted to the source of the infection or inflammation. Neutrophils are not limited to a specific area of circulation. They can move freely through the walls of veins and into the tissues of our body to immediately attack specific antigens.

- **Alveolar macrophages**: Alveolar macrophages (AM), also known as dust cells, are white blood cells. Alveolar macrophages are the first line of defense against invading respiratory pathogens. They reside in pulmonary alveoli and inter-alveolar septum near pneumocytes. The alveoli are the terminal unit of the respiratory system responsible for gaseous exchange. The alveoli are comprised of three different kinds of cells: (1) Type I pneumocytes, (2) Type II pneumocytes, and 3) Alveolar macrophages. (1) Type I pneumocytes build up the structure of the alveolar wall and aid in respiration process. Specifically, they do not replicate. (2) Type II pneumocytes secrete a lipoprotein called surfactant that prevents the collapse of the alveoli even after exhalation. (3) Alveolar macrophages produce various signaling chemicals that interact with

other cells of the immune system to orchestrate a response that maintains immunologic and tissue homeostasis in the body [59, 60].

- **Mycobacterium tuberculosis**: Mycobacteria belong to the Mycobacteriaceae family and the order Actinomycetales. Within the pathogenic species belonging to the M. tuberculosis complex, which comprises eight distinct subgroups, the most common and influential agent of human disease is Mycobacterium tuberculosis [61]. (For further details, see Chapter 1).

- **Liposome**: A liposome is a spherical vesicle having at least one lipid bilayer [62]. The liposome can be used as a drug delivery vehicle to administer nutrients and pharmaceutical drugs, such as lipid nanoparticles in mRNA vaccines and DNA vaccines [63]. Liposomes can be prepared by disrupting biological membranes. A liposome design may employ surface ligands for attaching to unhealthy tissue.

4.2.2 Molecular entities

The molecular entities added to the simulator to simulate the dynamics of tuberculosis are:

- **Interleukins**: cytokines whose presence at the lattice site promotes or blocks specific interactions by increasing or decreasing their probability of being released [64].

- **Interferons**: cytokines produced by both white blood and tissue cells in response to the presence of specific antigens coming from viruses, bacteria, or tumor cells [65].

- **Vitamin D**: a regulator of calcium metabolism that helps in maintaining normal levels of calcium and phosphorus in the blood [66].
- **Lipoxin A4**: a bioactive molecule that helps in hindering the biochemical process that leads to inflammation, and removing those cells that promote inflammation. It represents a pro-necrotic factor within the cellular compartment during tuberculosis pathogenesis [67].
- **Prostaglandin E2**: a mediator that influences critical physiological events such as blood coagulation, gastric mucous membranes, and is involved in the mechanism of inflammation. It represents a pro-apoptotic factor, i.e., it activates a sort of cell programmed death that can be spontaneous or induced by agents of different origin [68]. In tuberculosis, it favors the apoptosis of MTB infected cells.
- **Isoniazid**: Isoniazid (INH) is the main antibiotic used in the treatment of tuberculosis. It is used alone or in combination with other specific drugs [69].

Molecular entities do not have internal states and are not represented individually, but populations are defined as different concentrations of different molecular entities within the lattice.

4.2.3 Interactions among entities

The main interactions between the entities and/or with the Mycobacterium tuberculosis are the following ones:

- Macrophage (M) - Alveolar Macrophage (AM)
- Alveolar macrophage (AM) - Mycobacterium tuberculosis (MTB)

- Neutrophils (N) - Mycobacterium tuberculosis (MTB)
- Dentritic cells (DC) - Mycobacterium tuberculosis (MTB)
- Cytotoxic T cells (CT) - Alveolar Macrophage (AM)
- Dentritic cells (DC) - Alveolar Macrophage (AM)
- Reactivation of Tuberculosis
- Dentritic Cells (DC) - Liposome (LP)
- Cytotoxic T Cells (TC) - Liposome (LP)
- Isoniazid (INH) - Mycobacterium tuberculosis (MTB)

The interactions have been represented graphically with 2DIs tool [70].

4.2.4 Treatment strategies implemented

At this step, two treatments against tuberculosis were modeled and implemented within the computational infrastructure: INH (Section 4.2.4), the antibiotic administered to those individuals who get infected with the Mycobacterium, and RUTI® vaccine (Section 4.2.4) that has been specifically developed as a therapeutic vaccine for TB. The vaccine is capable to reduce bacillary load when administered after chemotherapy in murine and guinea pig models. It is also immunogenic when given to healthy adults and individuals with latent TB.

Isoniazid antibiotic

INH is one of the most effective anti-TB drugs used for TB treatment. This compound was first synthesized in the early 20th century, and its activity against TB

was first reported in the early 1950s [71, 72]. With the introduction of INH, TB treatment was first considered feasible. At therapeutic levels, INH is bactericidal against actively growing intracellular and extracellular MTB organisms. INH is used in conjunction with other effective anti-tuberculosis agents under multi-drug therapy.

This pro-drug requires activation by the heme enzyme catalase/peroxidase (KatG) of MTB. Johnsson et al. [73] subsequently established that INH is a pro-drug and requires conversion by the mycobacterial catalase-peroxidase (KatG) encoded by the KatG gene. The activation mechanism has not yet been clearly understood as the binding interaction has not been appropriately established [74].

INH works by killing the bacteria that cause tuberculosis disease. Still, it is thought to prevent the tuberculosis mycobacteria from synthesizing substances called mycolic acids [75] needed to form the cell walls of the bacteria. INH also seems to combine with an enzyme that interferes with the cell metabolism of the mycobacteria. As a result of this metabolic alteration, the mycobacteria die. There are two stages in the treatment of tuberculosis. In the first two months, treatment is aimed at killing as many bacteria as possible. Therefore, several anti-TB drugs with different mechanisms of action are used simultaneously, usually rifampicin, pyrazinamide, isoniazid, and ethambutol.

Hence, combined protocol treatment is more likely to be effective than using a single medicine. Controversially, using more drugs together make it less likely that bacteria will develop resistance to the treatment. After this time, some of the medicines are stopped, and the others (usually rifampicin and isoniazid) are continued for four months to kill any remaining bacteria. INH is used in both stages of treatment.

Later on, mutations in the katG gene in Mtb strains were associated with an INH-resistant phenotype [76]. Many of the tuberculosis outbreaks in the world are drug-resistant (DR), along with an increasing threat in certain regions around the world [77, 78]. Most drug-resistant MTB clinical strains are resistant to INH [79].

RUTI® vaccine

Therapeutic vaccines do not aim to prevent tuberculosis infection, but act once the infection, whether latent or active, is established. The main goal is to administer these vaccines in addition to antituberculosis drugs, especially in more complex cases of tuberculosis. The main aims are to:

- reduce the duration of drug treatment;
- promotes a rapid resolution of the clinical picture.

One of the most recent therapeutic vaccines is RUTI® [80, 81] developed by Archivel Farma to integrate the treatment of latent TB by reducing the duration of prophylactic therapy with INH. The vaccine showed a very good safety profile in a phase 2 trial [82] and is expected to start a phase 3 soon. Figure 4.3 shows a summary of the therapeutic effects of RUTI®, focusing on decreasing the probability of tuberculosis reactivation. RUTI® has demonstrated its efficacy in controlling LTBI in experimental models of mice and guinea-pigs after a short period of chemotherapy. These experiments in animals showed the induction of a mixed (Th1, Th2, and Th3), polyantigenic response with no local or systemic toxicity. Local accumulation of specific CD8 T cells and a strong humoral response are characteristic features of RUTI® that explain its protective properties. RUTI® was also designed to trigger a new immunological response against antigens of the latent bacilli, i.e., the so-called

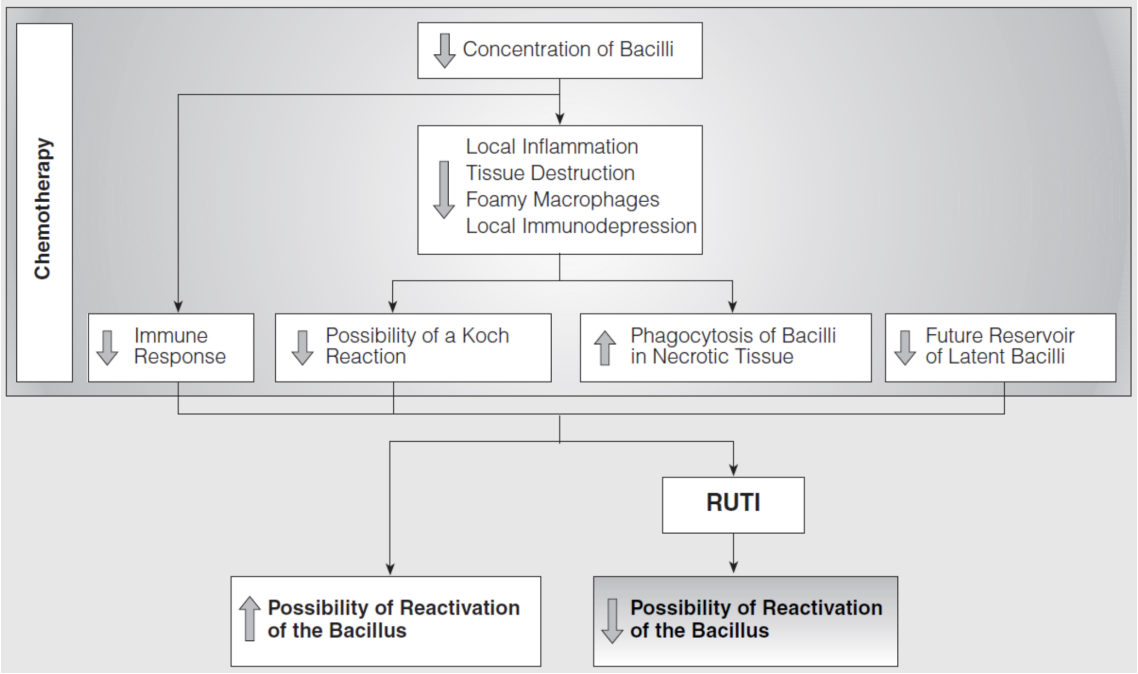


Figure 4.3: Temporal strategy for the use of RUTI®, indicating the effects of short-course chemotherapy and the requirement for subsequent immunotherapy [83]

"structural" antigens [84] and those associated with stress responses. Protective immunity arises against antigens that are actively released by growing bacilli. Likely, the "focus" of the immunological response only on growing bacilli allows non-active bacilli to remain "invisible" to the specific immunity of the host.

RUTI® is made with bacilli grown under the stressful conditions of starvation, low pO2, and low pH. Conditions are achieved gradually by culturing solid media [85, 86]. Bacilli used to make RUTI® are subjected to conditions that are probably found in the granuloma of hosts with active immunity: in other words, a low pO2 in the fibrotic structure of the granuloma and the low pH and starving conditions inside the phagolysosome of the activated macrophage [87].

The choice of fragmentation of the bacilli for RUTI® formulation allowed optimal

presentation of cell wall antigens, as well as the choice of its composition into liposomes [88]. The average diameter of these fragments, which is 0.1 mm, allows the antigens in the cell wall to be effectively presented and thus provide an easier recognition of latent bacilli. Moreover, it is well known that the cell wall of M. tuberculosis has adjuvant properties [89] that ensure the induction of an immunological response without further adjuvant measures. In addition, the induction of a polyantigenic reaction is postulated to help recognize latent bacilli. Another relevant feature of RUTI® is the "detoxification" of the cell fragments obtained.

4.3 Retrospective validation of UISS-TB

To assess UISS-TB credibility, the results obtained from several tuberculosis studies, widely reported in literature have been retrospectively reproduced. In particular, a court of statistically significant digital twins with specific characteristics acquired from the selected studies was generated for each reference study considered [90, 91, 92, 93, 94, 82].

A retrospective study investigates specific outcomes at the beginning of a study by looking backward at data collected from previous patients. Retrospective studies may be either cohort or case-control studies and have four primary purposes:

- to be an audit tool for comparison of the historical data with current or future practice;
- to test a potential hypothesis regarding suspected risk factors concerning an outcome;
- to ascertain the sample size and data required for a prospective study or trial;

In the following paragraphs the *in silico* results, reproducting specific TB in vivo studies will be reported. The main aim is to demonstrate that UISS-TB accurately reproduces the natural mechanisms of pulmonary tuberculosis infection from a phatophysiological point of view, along with the reproduction of the Mechanism of Action (MoA) of isoniazid and RUTI® [91, 95, 96]. Thus, the three levels of the retrospective validation will be described on the basis of three layers: physiological, disease, and treatment layers.

4.3.1 Physiological model

Introduction

When challenged by an exposure to MTB, the immune system of normal healthy adults responds in a fairly standardised way, which is described in detail in the dedicated literature [97, 98]. As a first evidence of credibility, UISS-TB is able, when informed with the input features of an average patient, to reproduce each of these standardised immune responses. This preliminary evidence confirms that UISS-TB is capable of replicating the fundamental features of the human immune system. Actual extension of the UISS simulation platform is able, at the end of the game, to reproduce and simulate two specific MTB scenarios i.e., the one in which the host immune system is able to recognize and clear the infection and the other one in which, instead, MTB establishes a chronic infection with some granulomas formation as a reservoir of MTB infection.

Results

As first evidence of credibility, UISS-TB is able to mirror each of the expected immune responses when informed of the input features of an average patient exposed

to an Mtb challenge. Due to the lack of quantitative data related to the early phase of Mtb infection in humans, a semi-quantitative approach to validate UISS-TB predictions was considered. In particular, for each depicted immune system dynamics in the early response to Mtb, an expert panel of immunologists provided three possible scoring figures: 0 erroneous prediction, 1 partially corrected prediction (in this case, a note explains the reasons why), 2 fully corrected prediction.
Specifically, mean behavior and error bars are shown for each biological entity. A total of 100 digital twins simulations has been randomly run varying HLA repertoire, and one microliter of peripheral blood / 1 cubic millimeter of lung tissue was simulated. It has been carried out several studies which are:

- **The innate early host immune response to Mtb infection by the digital patients**: after Mtb challenge at day 0, an increase was observed in terms of influx of phagocytic cells in the early phase of infection, including recruited and activated neutrophils, macrophages, MHC-I and MHC-II antigen-presenting cells by DCs and primarily resident alveolar macrophages [13].

- **The typical cellular response of CD4+ and CD8+ T cells mounted against Mtb infection**: on average, CD4+ Th17 population is the first prominent response, followed by CD4+ Th1 response, contributing in the attempt to control the infection. It has been observed that the referred timeline is in good agreement with the described typical T cell response [13, 99].

- **The dynamics of the typical CD4+ Th1 cytokines signature**: IL-1 and IL-12 are released almost immediately after the Mtb exposure, while IFN-γ and IL-2 after a few days when the adaptive response is being mounted [13].

- **The dynamics of the typical CD4+ Th17 cytokines signature**: it has been observed that their levels are higher when compared to the Th1 signature cytokines, indicating that the first line of adaptive defense is directed against the extracellular activity of Mtb [13].
- **The dynamics of LXA4 and PGE2**: It has been observed a higher level of the latter one: this is in line with the low virulence of the Mtb simulated strain (such as H37Ra) that promotes PGE2 pro-apoptotic factor. [13].
- **The dynamics of Mtb viable bacilli and specific IgM, IgG, and IgA anti-Mtb**: it has been observed that the immune system can control Mtb spread after the first week post-infection; also this behaviour can be reproduced and detected through UISS-TB and this is in line with literature data. [13].
- **Evaluation of QuantiFERON-TB Gold Plus for predicting incident tuberculosis among recent contacts**: Guptal et al.'s work [100] has been used to assess the capability of UISS-TB in predicting the numerosity of healthy patients that eventually get infected after MTB exposure. UISS-TB predicted 121 QFT-plus positive results against 126 ones reported in the study, showing excellent evidence that mirrors experimental data.

4.3.2 Disease model

Introduction

Today, finding untreated tuberculosis patients data is quite impossible, because it would be unethical to run a clinical trial on active TB patients with an untreated (placebo) arm. However, the literature provides extensive evidences of the epidemiology of active TB when untreated, thanks also to a large amount of data coming

from sanatorium reports of different world zones.

To demonstrate that UISS-TB accurately reproduce the natural history of pulmonary tuberculosis infection, model predictions have been validated in terms of::

1. Rate of mortality over a population of untreated subjects exposed to MTB infection [92];

2. Rate of patients with latent MTB infection that develop the active disease form over a period of time [93].

Results

Ragonnet et al. [92] considered the reports identified in a previous systematic review of different studies from the prechemotherapy era and extracted detailed data on mortality rate over time. They used a Bayesian framework to estimate the rates of TB-induced mortality and self-cure. The inference was performed separately for smear-positive TB (SP-TB) and smear-negative TB (SN-TB). They included 41 cohorts of SP-TB patients and 19 cohorts of pulmonary SN-TB patients in the analysis. The median estimates of the TB-specific mortality rates were 0.389 year^{-1} (95% credible interval [CrI], 0.335–0.449) and 0.025 year^{-1} (95% CrI, 0.017–0.035) for SP-TB and SN-TB patients, respectively.

To demonstrate that UISS-TB is able to realistically reproduce these in vivo results, five SP-TB and two SN-TB cohorts with 1500 and 513 digital twins have been generated respectively. It has been obtained that the median estimates of the TB-specific mortality rates were 0.33±0.23 and 0.02±0.02 for SP-TB and SN-TB patients (Figure 4.4). The results obtained are very close to the ones observed n vivo. This demonstrates that UISS-TB is able to accurately reproduce the mortality rate caused by the disease.

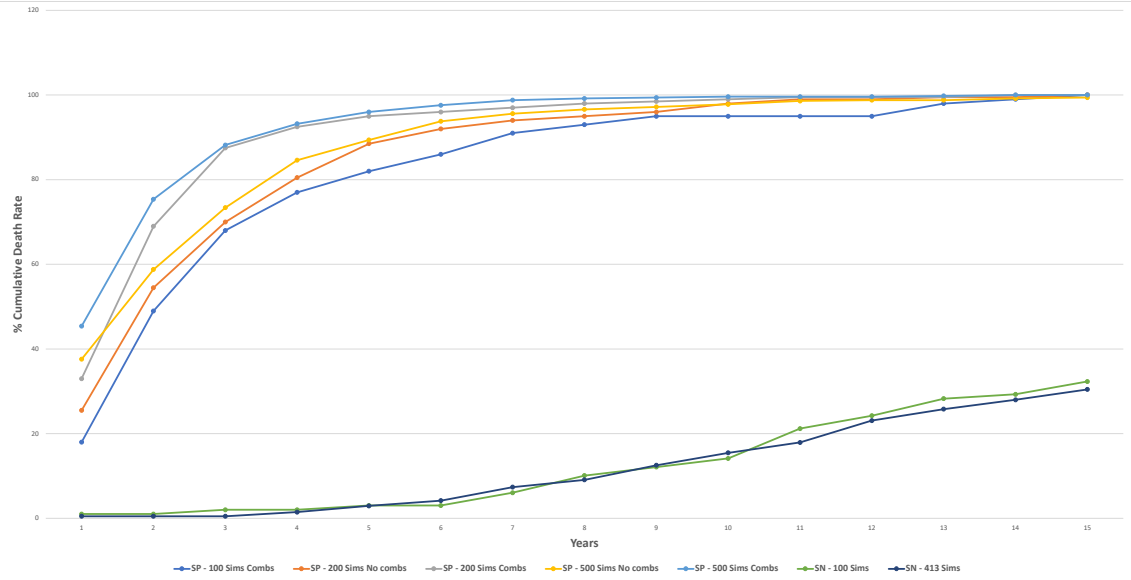


Figure 4.4: The figure shows respectively the cumulative percentage of death (Smear-positive and Smear-negative) of the digital patients generated through the observed data from literature [92].

Shea et al. [93] estimated the rate of reactivation tuberculosis (TB) in the United States, from an overall and by population subgroup point of view, considering specific TB cases. The rate of reactivation of TB was defined as the number of non-genotypically clustered TB cases divided by the number of person-years at risk for reactivation and due to prevalent latent TB infection (LTBI). Of the 39,920 TB cases reported during 2006–2008, 79.7% of them was attributed to reactivation.

To reproduce *in silico* these observed results, a cohort of 1000 digital twins was generated through a normal (or Gauss) distribution, using mean age and its standard deviation provided by the reference study. The simulation was set to for five years to establish LTBI and then TB reactivation. In the time range of 2-4 years, it has been observed that 79.7% of cases of patients progressed to an active form of TB. Finally, UISS-TB was able to estimate the rate of disease reactivation almost perfectly in comparison with the literature.

4.3.3 Treatment model

Introduction

UISS accurately simulates tuberculosis dynamics and its interaction within the immune system, and how it predicts the efficacy of the mechanism of action of isoniazid [94] and RUTI® vaccine [82] in a specific digital population cohort. Specifically, two groups of digital twins have been simulated. The first group was only treated with isoniazid, while the second one was treated with the combination of RUTI® vaccine and isoniazid, according to the dosage strategy described in the clinical trial design [91]. UISS-TB shows to be in good agreement with clinical trial results suggesting that RUTI® vaccine may favor a partial recover of infected lung tissue.

Results

In Katiyar et al.'s study [94], they evaluated and assessed the effectiveness of a high-dose of isoniazid (INH) (16–18 mg/kg) as adjuvant to second-line therapy in documented cases of MDR-TB. This was achieved through a double-blind, randomized controlled trial with three treatment arms (high-dose INH, normal-dose INH, and placebo) in addition to second-line drugs. The study primary outcomes are represented by time to sputum culture conversion and proportion with sputum culture negative six months after treatment initiation. To reproduce *in silico* the above mentioned results, three cohorts of approximately 300 digital twins with the dosages defined in [94] have been generated. Kaplan-Meier was used to represent the conversion to SN-TB for some patients as performed in the reference study. In particular, the Kaplan-Meier procedure estimates time-to-event models in the presence of censored cases. The Kaplan-Meier model is based on estimating conditional probabilities at each time point when an event occurs, and taking the product limit

of those probabilities to estimate the survival rate at each point in time.

Nell et al. [82] evaluated the safety, tolerability, and immunogenicity of three different doses (5, 25, and 50 mg) of the novel antituberculosis vaccine RUTI® compared to placebo in subjects with latent tuberculosis infection through a double-blind, randomized, placebo-controlled Phase II clinical trial (95 patients randomized). Three different RUTI® doses and placebo were tested, randomized both in HIV-positive (n = 47) and HIV-negative subjects (n = 48), after completion of one-month isoniazid (INH) pre-vaccination. Each subject received two vaccine administrations, 28 days apart. Overall, a polyantigenic response was observed, which differed by HIV status. The best polyantigenic response was obtained when administering 25 mg RUTI®, especially in HIV-positive subjects, which did not increase after the second vaccination.

To reproduce *in silico* the in vivo results, three cohorts (about 300 digital twins with HIV negative), one for each dose (5, 25, and 50 mg) will be generated. Each cohort will be simulated with the same treatment period according to the literature. Finally, it should be analyzed each population and verified that 25 mg dosage is the best one as observed in vivo.

In summary, *in silico* trials, innovations represent a powerful pipeline for predicting the effects of specific therapeutic strategies and related clinical outcomes. Here, it has been shown that the simulated mechanism of action INH is in good alignment with the results coming from past clinical trials, and UISS-TB in silico platform will be soon validated retrospectively.

4.4 UISS-TB Graphical User Interface

As UISS is written in C language, a Graphic User Interface (GUI) and a web server is needed to provide a user-friendly interface. The GUI can be accessed at https://combine.dmi.unict.it/UISS-TB/. In a previous work, it was presented a web interface developed in Flask micro-server [96].

Flask [101] is a web application framework written in Python [102]. It was developed by Armin Ronacher [103] and is based on the Werkzeg WSGI toolkit [104] and the Jinja2 template engine [105] (both projects were developed by the same team).

Today, the performance of the web platform has been improved by replacing the Flask framework with Django [106]. Django is a high-level Python web framework that enables the rapid development of secure and maintainable websites. It is free and open-source, owns a thriving and active community, great documentation, and many free and paid-for support options. These enhancements allow the launch of the simulations separately from the main thread and in a more efficient way. Figure 4.5 shows a screenshot of the web Graphic User Interface (GUI) in which it is possible to set different combinations of cellular and non-cellular parameters.

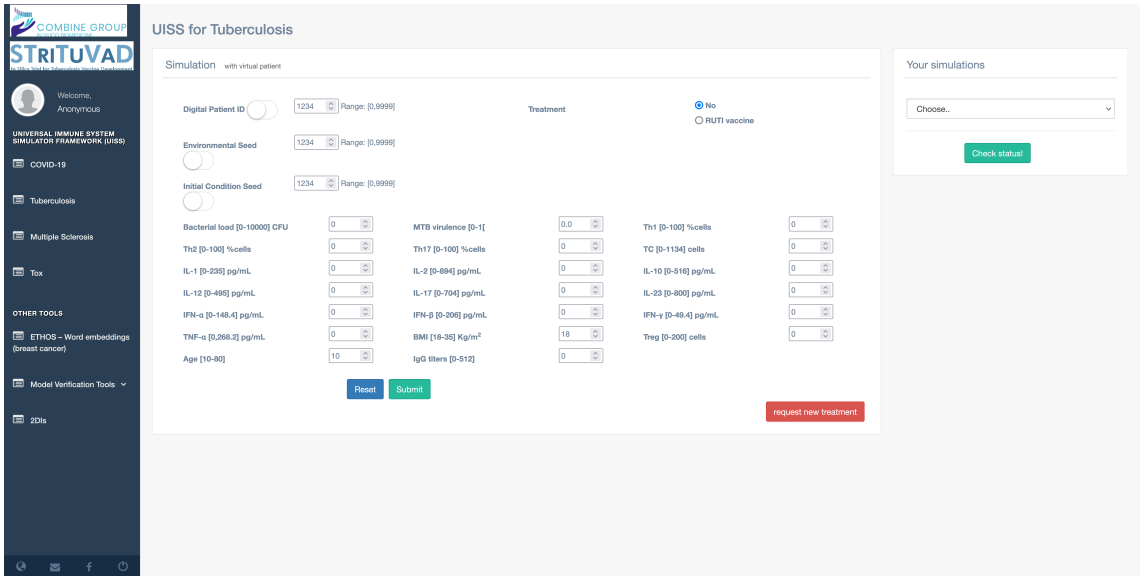


Figure 4.5: Web Graphic User Interface of UISS-TB. This figure depicts the GUI of UISS that allows the run of the simulations. The "Simulation's Parameters" zone, on the left side of the figure, represents the vector of features for the personalization of the digital patients. The "Your Simulations" box, on the right side of the figure, shows the list of all the simulations launched by the user. The simulations are classified in "running" or in "completed" status.

On the right side, one can see a box called "Your simulation" containing a list of the simulations, sorted by their creation date and classified in "running" or in "completed" status. On the left side, one can see a box named "Simulation Parameters" containing a set of the biological and physiopathological parameters that compose the vector of features created for the customization of TB patients. To better represent the biological diversity of TB patients, it was enriched the composition of the vector of features used for the generation of digital patients libraries. In particular, the "vector of features" that defines a specific TB patient is composed by the following parameters: 1) MTB virulence; 2) MTB Sputum; 3) CD4-Th1; 4) CD4-Th2; 5) IgG; 6) TC; 7) IL-23; 8) IL-12; 9) IL17-A; 10) IL-2; 11) IL-1; 12) IL-10; 13) IFN1A; 14) IFN1B; 15) IFNG; 16) TNF; 17) Treg; 18) LXA4; 19) PGE2; 20)

Vitamin D; 21) Age; 22) BMI. The digital patients were generated according to the steps explained in [107]. For each entity, it specified the unit of measurement and the range of values used to perform the simulations.

The simulation will be performed asynchronously. This allows launching the process separately from the main thread, notifying the user when the simulation is finished, and performing multiple simulations simultaneously.

In detail, after the user connects to the UISS-TB web interface, she/he can select the Tuberculosis disease model. After that, the general GUI panel appears. The user finds in the vector of features some parameters already filled in default values. The user can vary these values according to the ranges shown within brackets near the selected parameter. After that, one can click on the Submit button, and an unique identification simulation number will be assigned. After selecting the simulation id, the user can check the simulation status by clicking on the check status button. When the simulation is completed, the user can visualize the results of the immune system dynamics, choosing the one she/he would like to analyze (Figure 4.6).

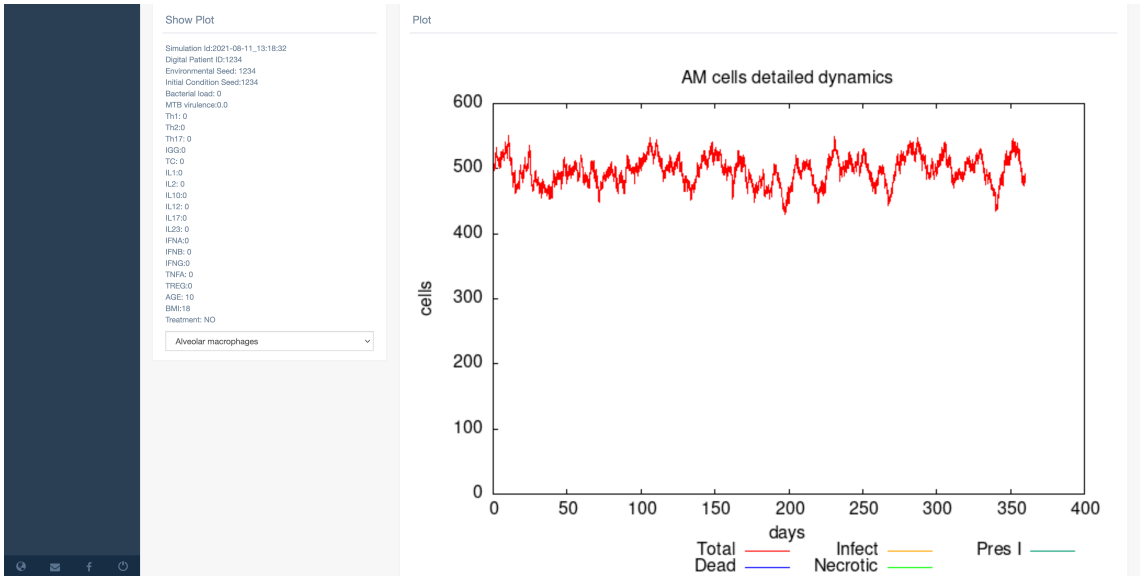


Figure 4.6: The figure shows on the left all the parameters selected for the simulation and the related simulation plot on the right. On the bottom right different types of graphs can be selected.

In collaboration with the USFD partner (flameGPU development team) [108, 109, 110], a version of the GUI based on the client/server model is currently under development. The client-side is managed by UniCT, while USFD develops the server-side. The main objective is to maintain all the advantages described above and run the UISS-TB simulator remotely on the University of Sheffield's HPC machines using GPUs [111, 112]. This would result in lower simulation times, but simultaneously, can allow to run many more simulations from different users.

Chapter 5

A statistical analysis tool for UISS: MetricUISS

UISS general framework, at each completed simulation, generates different types of files describing the trend [56, 95], i.e., the amount of cellular and molecular entities at each time step. For this reason, a tool called MetricUISS was developed to have simultaneously multiple simulations and perform statistical analysis.

MetricUISS is a tool developed in Python3 [102], used in the biomedical field, particularly for the statistical analysis of large amounts of data generated by UISS and/or for the generation of a cohort of digital patients [57]. It is divided into specific modules. Thanks to this, it is easier to develop new features or to extend existing ones. Therefore, MetricUISS can be adapted to any version of UISS, i.e., with a specific disease module or the physiological one. It uses the Pandas library [113] to manage all these data and perform statistical functionality. In Figure 5.1, one can observe how the class is defined and all the variables useful for the operation. In this context, the "$\_\_init()\_\_$" function is always executed when a class is created; it is used to assign specific values to the properties of the object created and to performs additional functions based on these values. The tool is executed with a

```python
class Metric:
    def __init__(self, args):
        self.num_step = 0
        self.nsims_not_completed = 0

        self.mL = args.mL
        self.timeplot = args.timeplot
        self.steperror = args.steperror
        self.formatplot = args.formatplot
        self.cpu = args.cpu
        if args.command == 'analysis':
            self.datafile = args.datafile
            self.nsims = args.nsims
            self.rseed1 = args.rseed1
            self.rseed2 = args.rseed2
            self.rseed3 = args.rseed3
            self.mL = args.mL

            self.clear_data()
            self.analysis_mode()
        elif args.command == 'plot':
            self.datafile = None
            self.nsims = len(glob.glob(os.path.join(os.getcwd(), 'metric_results', 'simulations', '*')))
            self.rseed1 = None
            self.rseed2 = None
            self.rseed3 = None

        self.plot_mode()

        self.quantitative_analysis()

        if self.formatplot == 'pdf':
            self.create_pdf()
        else:
            print(f'Non è possibile creare il PDF con il formato {self.formatplot}')

        # print(self.datafile, self.nsims, self.rseed1, self.rseed2,
        #       self.rseed3, self.timeplot, self.steperror, self.cpu)

```

Figure 5.1: Python 'Metric' class uses the argument 'args.command' to define its behavior based on the input value, i.e., analysis or plot.

single command line from the shell containing all parameters to be set, as it does not own any graphical interface, as well as the following command:

python3 metric.py $< mode >$

The "argparse" library [114] is used to input selected arguments to the tool before it starts. This library makes writing command-line interfaces easier by making it more user-friendly. Argparse automatically scans the input arguments and converts them into a python object called a "dictionary" for use within the source code [115].

The argparse module also automatically generates help and usage messages, with the –help argument, which can also be shortened to -h, and issues errors when users give the program invalid arguments.

In particular, MetricUISS requires a first argument, "mode", which defines a set of instructions to be executed based on the assigned value. There are two available approaches implemented within the "mode" attribute:

1. **Analysis**: allows "n" simulations to be carried out and all the statistical analyses set up (in this case for Tuberculosis) to be performed. As can be seen in Figure 5.2 - Panel A, several arguments can be set:

   - **-h**: displays all available parameters for this "mode" selected;
   - **-d**: defines the setting of data-input to be performed for UISS;
   - **-s**: indicates the number of simulations to generate;
   - **-rseed**: indicates the type of seed assigned to every simulation; it can be random, random "retrievable", or static;
   - **-ml**: defines the number of microliters;
   - **-t**: indicates the time to be plotted on the x-axis (i.e., seconds, minutes, days, months, weeks, years);
   - **-e**: defines at how many time-steps the error bars should be plotted; by default, it is set to 100 time-steps;
   - **-f**: indicates the type of extension to use for the generated plots; by default, it uses the PDF extension;
   - **-c**: sets the number of CPUs assigned to run the simulations and the analysis; if not specified, the tool will automatically use all CPUs installed on the machine.

Digital twins	1 CPU	2 CPUs	4 CPUs	8 CPUs
25	256 s	138 s	82 s	55 s
50	481 s	261 s	143 s	97 s
75	736 s	382 s	214 s	136 s
100	963 s	505 s	285 s	177 s

Table 5.1: Computational efforts expressed in seconds required to complete the different number of simulations varying the numbers of assigned processors.

2. **Plot**: using previously generated simulations, all plots can be created also producing any variants or new ones. As one can seen in Figure 5.2 - Panel B, several arguments can be set:

   - **-h**: displays all available parameters for this "mode" selected;
   - **-ml**: defines the number of microliters;
   - **-t**: indicates the time will be plotted on the x-axis (i.e., seconds, minutes, days, months, weeks, years);
   - **-e**: defines at how many time-steps an error bar should be plotted; by default, it is set to 100 time-steps;
   - **-f**: indicates the type of extension to use for the generated plots; by default, it uses the PDF extension;
   - **-c**: sets the number of CPUs assigned to run the simulations and the analysis; if not specified, the tool will automatically use all CPUs installed on the machine.

Hence, as previously mentioned, MetricUISS can be run in multiprocessor mode [116] through the joblib library [117], with the advantage to speed up the execution, due to the number of CPUs assigned.

A

```
usage: metric.py analysis [-h] -d DATAFILE -s NSIMS [-r1 RSEED1] [-r2 RSEED2]
                          [-r3 RSEED3] [-ml ML] [-t TIMEPLOT] [-e STEPERROR]
                          [-f FORMATPLOT] [-c CPU]

optional arguments:
  -h, --help            show this help message and exit
  -d DATAFILE, --datafile DATAFILE
                        Indica il nome del datafile
  -s NSIMS, --nsims NSIMS
                        Indica il numero di simulazioni totali
  -r1 RSEED1, --rseed1 RSEED1
                        Indica il valore delle RS1 - [-1=valore random
                        recuperabile, 0=valore random, [1-10000]= valore
                        statico] - Valore di default = -1
  -r2 RSEED2, --rseed2 RSEED2
                        Indica il valore delle RS2- [-1=valore random
                        recuperabile, 0=valore random, [1-10000]= valore
                        statico] - Valore di default = 342
  -r3 RSEED3, --rseed3 RSEED3
                        Indica il valore delle RS3- [-1=valore random
                        recuperabile, 0=valore random, [1-10000]= valore
                        statico] - Valore di default = 1234
  -ml ML, --ml ML       Indica il valore dei mL - Valore di default = 1
  -t TIMEPLOT, --timeplot TIMEPLOT
                        Definisce il tempo dell'asse delle x - [0=seconds,
                        1=minutes, 2=hours, 3=days, 4=weeks, 5=months,
                        6=years] - Valore di default = 3
  -e STEPERROR, --steperror STEPERROR
                        Definisce lo step_error - Valore di default = 100
  -f FORMATPLOT, --formatplot FORMATPLOT
                        Definisce il formato del plot - Valore di default =
                        pdf
  -c CPU, --cpu CPU     Massimo numero di cpu assegnate all'analisi
```

B

```
usage: metric.py plot [-h] [-ml ML] [-t TIMEPLOT] [-e STEPERROR]
                      [-f FORMATPLOT] [-c CPU]

optional arguments:
  -h, --help            show this help message and exit
  -ml ML, --ml ML       Indica il valore dei mL - Valore di default = 1
  -t TIMEPLOT, --timeplot TIMEPLOT
                        Definisce il tempo dell'asse delle x - [0=seconds,
                        1=minutes, 2=hours, 3=days, 4=weeks, 5=months,
                        6=years] - Valore di default = 3
  -e STEPERROR, --steperror STEPERROR
                        Definisce lo step_error - Valore di default = 100
  -f FORMATPLOT, --formatplot FORMATPLOT
                        Definisce il formato del plot - Valore di default =
                        pdf
  -c CPU, --cpu CPU     Massimo numero di cpu assegnate all'analisi
```

Figure 5.2: Panels A and B show the available topics for the "analysis" and "plot" approaches respectively.

MetricUISS tests were run on a server equipped with 8 Intel Xeon E5 CPUs at 2.40 GHz and 128 GB of RAM with different combinations of simulations ranging from 25 to 100, using 1, 2, 4, and 8 CPUs. Table 5.1 shows the total time expressed in seconds to complete the n simulations classified by the number of CPUs and simulation time (3 months - 90 days). Figure 5.3 shows the multiprocessor scalability of MetricUISS in terms of computational cost, as described above. As one can observe, a clear improvement can be obtained when using many processors (8 in this case) and a large number of simulations.

In Figure 5.4, the logic behind the operation of MetricUISS is shown and it can be divided in four basic steps: i) parameter setting, ii) running UISS, iii) analysis, and iv) representation of results. The current functionality implemented within the tool includes:

- descriptive statistics: [118]

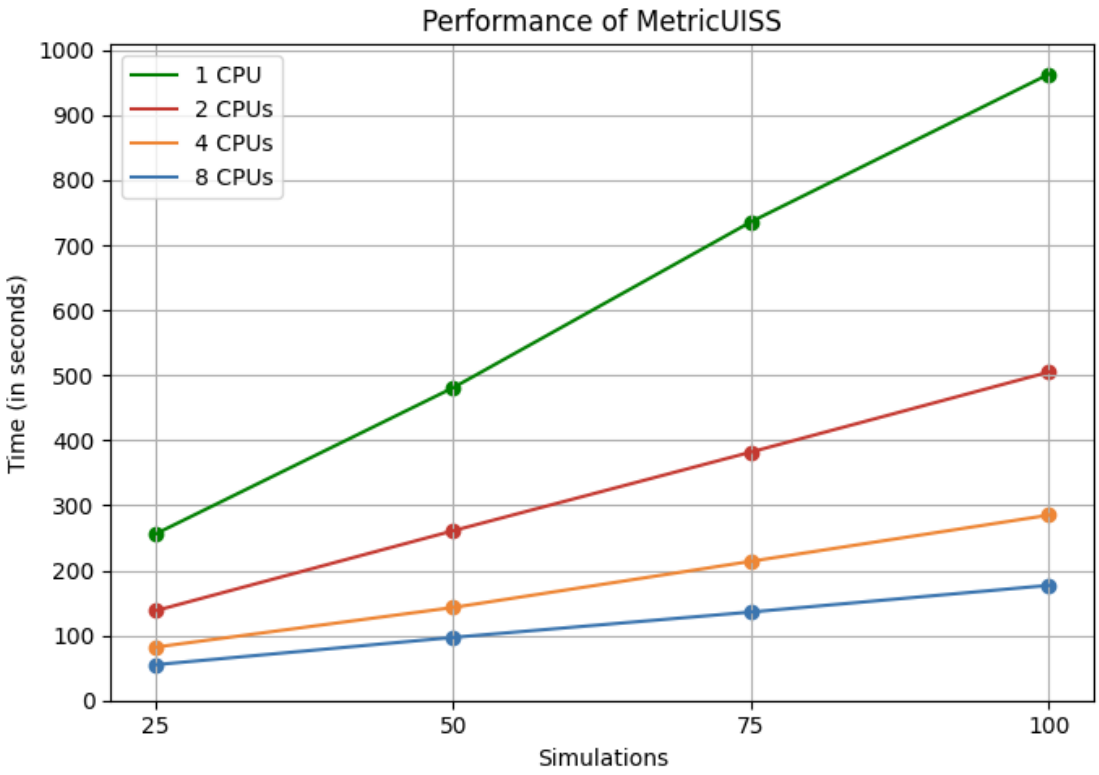


Figure 5.3: The figure shows the time taken into account (performance analysis) by MetricUISS to complete the simulations (25, 50, 75, 100) using a different number of CPUs (1, 2, 4, 8).

  - Calculation of the mean value
  - Measurement of the variance
  - Analysis of the minimum value
  - Calculation of the maximum value

- Extraction of a list of statistically relevant digital patients (useful to study a single individual) [119, 120].
- Filtering of those digital patients in which the bacterial load reaches high levels [121].

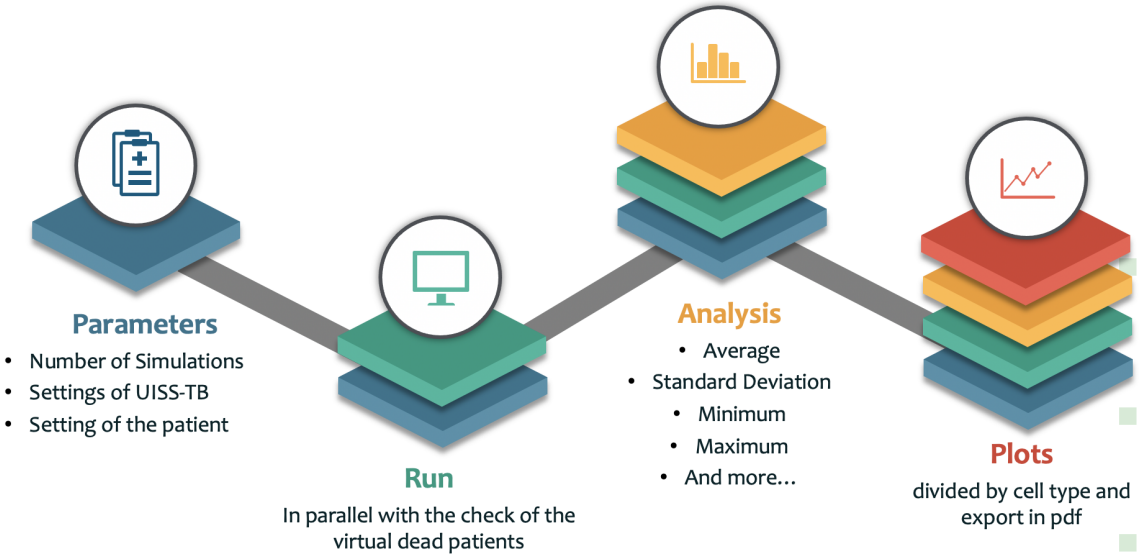


Figure 5.4: The basic steps of MetricUISS.

- Creation of specific tables subdivided by cell types (useful to study the total quantity in a specific time-step).
- Creation of plots by cell type based on mean and variance values.
- Generation or import of digital tuberculosis patients (with active or latent tuberculosis) with specific characteristics from third-party software [95].
- Prediction of antibiotic and/or vaccine effects or their combination [57].
- Export of plots in pdf format.

MetricUISS makes use of the matplotlib library [122] for plot generation. Matplotlib is a plotting library for the Python programming language and its numerical mathematical extension NumPy [123]. In particular, three different categories of plots can be distinguished:

1. **Quantitative plots for each cellular and molecular entity and related status**. More explanatorily, Figure 5.5 shows alveolar macrophages in the

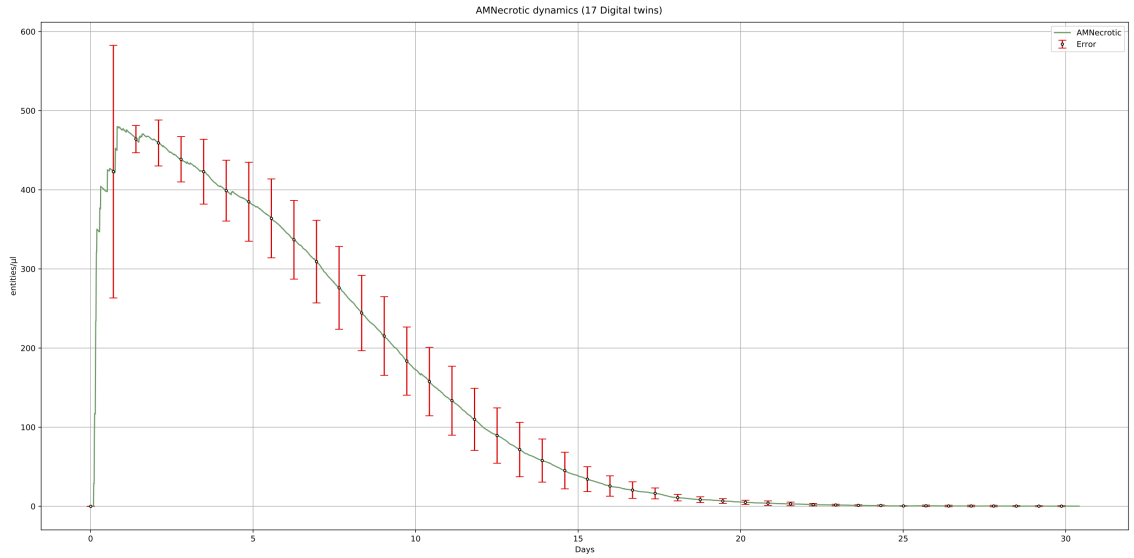


Figure 5.5: The figure represents the dynamics of alveolar macrophages in a scenario simulated for one month (30 days) with a cohort of 20 digital twins. The green line shows the average trend of the cell populations, while the error bars in red represents the standard deviation.

necrotic status over a total of 20 digital twins. It can be seen that 3 digital twins reached lethal bacterial load levels during the simulation.

2. **Plots with different counts for entities divided by compartment**. Figure 5.6 depicts the average percentage of cells (neutrophils, macrophages, epithelial cells, and lymphocytes) divided between the lung compartment and MTB sputum [124].

3. **Customized plots**. Specifically, MetricUISS allows a function to create combinations of plots (Figure 5.7) based on a specific study purpose (e.g., writing a scientific paper, presentation, or other).

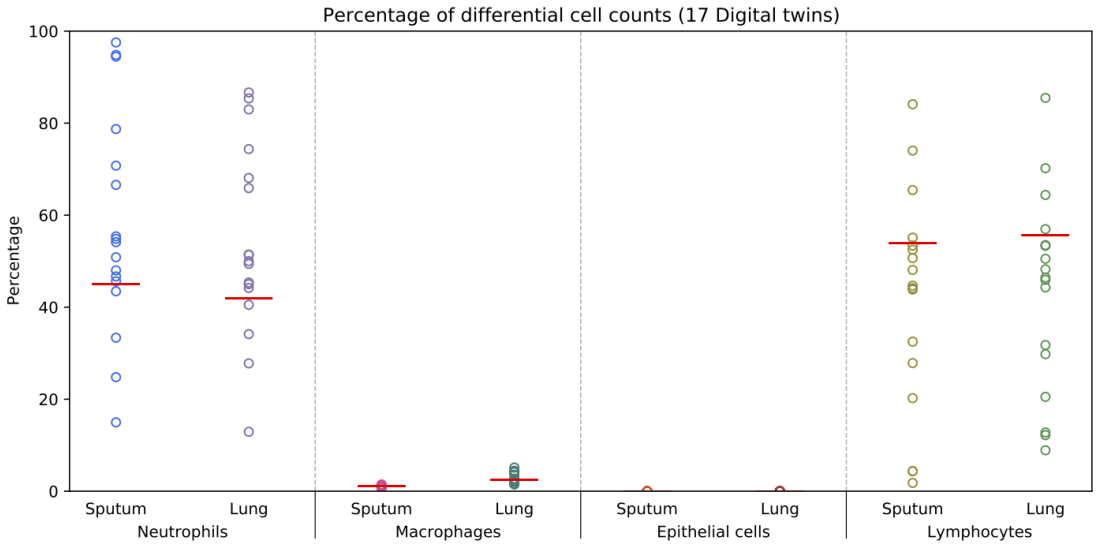


Figure 5.6: The figure depicts the differential cell counts in lung and sputum of the neutrophils, macrophages, epithelial cells, and lymphocytes. Horizontal bars in red indicate the mean value. The bullets represent the entity in each simulation divided for the compartment and expressed in percentage.

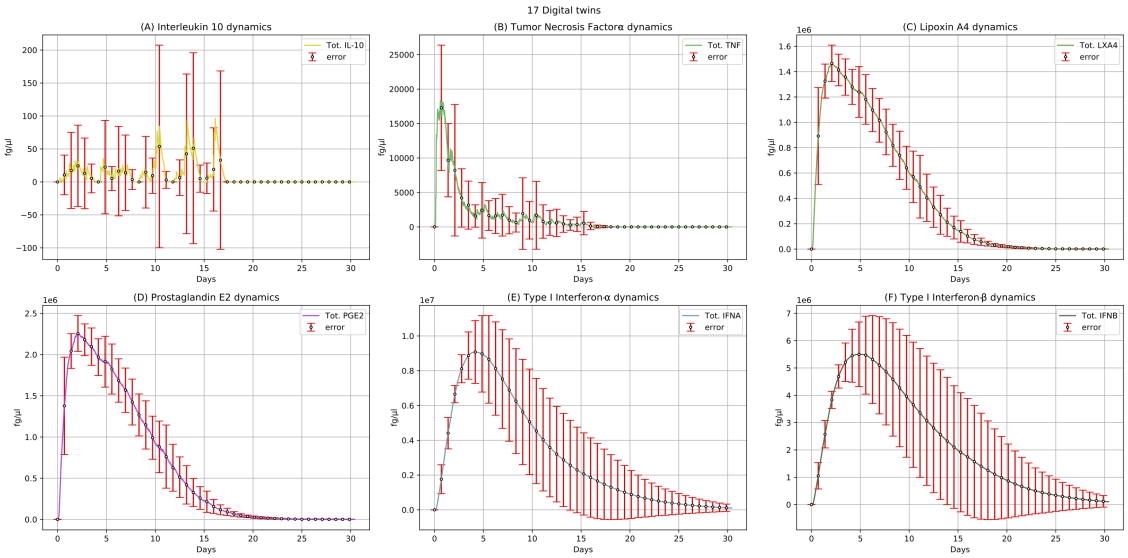


Figure 5.7: The figure represents a multiplot composed by: A) Interleukin 10, B) Tumor Necrosis Factor-alpha, C) Lipoxin A4, D) Prostaglandin E2, E) Type I Interferons: Interferon-alpha and F) Type I Interferons: Interferon-Beta.

Chapter 6

A pathway analysis tool for UISS: ParallEl paThways AnaLyzer (PETAL)

ParallEl paThways AnaLyzer (PETAL) [125, 126] is a tool developed in Python3 language [102] (now compatible with the latest version of Python 3.9.x) that provides better use and integration of Kyoto Encyclopedia of Genes and Genomes (KEGG) [127]. It represents a new pathway analysis tool that automatically explores and scans significant nodes of a pathway within KEGG, using new specially developed open-source tools and Python libraries, such as Pandas [128] and Joblib [117].
Specifically, it allows users to find hidden interactions between significant proteins belonging to the same pathway and other proteins within possible related pathways. Along with this, PETAL allows to search deeply for ancestor and descendant nodes of a specific target gene, making this task faster in terms of performance.
PETAL is potentially able to find pathways that are distant from those containing the target of interest. In addition, the core is scalable and parallelized with the ability to easily add new functionality and specific modules to simplify and automate

the discovery of new therapeutic targets or biomarkers.

The following paragraphs describe that PETAL exploits a data structure called Dataframe [128], containing specific information between two interlinked genes. For each relationship found, information about the two genes, such as the gene name and its identifier under the Homo sapiens (hsa) field, the isoforms involved, the pathways of origin, the protein interaction with its biological function, and the number of occurrences is stored in the data structure. Furthermore, during the analysis, the number of occurrences is calculated for each connection.

Finally, PETAL provides a graphical interface that allows the user to navigate through an interactive tree containing all the information found during the search. The tree is constructed from the data contained within the Dataframe and formatted in JSON. The main advantages of the tool can be summarised in three points:

1. a better performing of Breadth-First Search (BFS) [129];
2. the calculation of occurrences during analysis (i.e., the number of times a given path was found in all the pathways retrieved by KEGG during the search process between the initial and final gene);
3. a user-friendly interface with more information displayed.

PETAL output is made available through the d3.js library [130]. D3.js is a JavaScript library developed to display data dynamically and interactively, starting from some organized numeric data that combine HTML5 [131], Scalable Vector Graphics (SVG) [132], and Cascading Style Sheets (CSS) [133]. It is worth mentioning that PETAL GUI was not developed entirely from scratch but through the 'Radial Tree' project developed by Wm Leler [134]. It represents a viewer of interactive radial trees in which users can navigate and explore all the available information obtained during

the search. To this aim, the GUI includes specific commands that allow users to manage the output tree. Kyoto Encyclopedia of Genes and Genomes (KEGG) collects databases distributed according to their specific functionality, specifically in genomes, biological pathways, diseases, drugs, and chemicals.

The KEGG database project was initiated in 1995 by Minoru Kanehisa, predicting the need for a computer resource that could be used for biological interpretation of genome sequence data, so he began to develop the KEGG PATHWAY database [135, 136].

It contains manually drawn pathway maps representing experimental knowledge about metabolism and several other cell and organism functions. In particular, each pathway includes a range of information, such as a network of molecular interactions and reactions and the links between genes and gene products (mainly proteins) within the genome (Figures 6.1 and (6.2)).

KEGG is widely used in bioinformatics, particularly for data analysis in genomics, metagenomics, modeling and simulation in systems biology, and translational research in drug development.

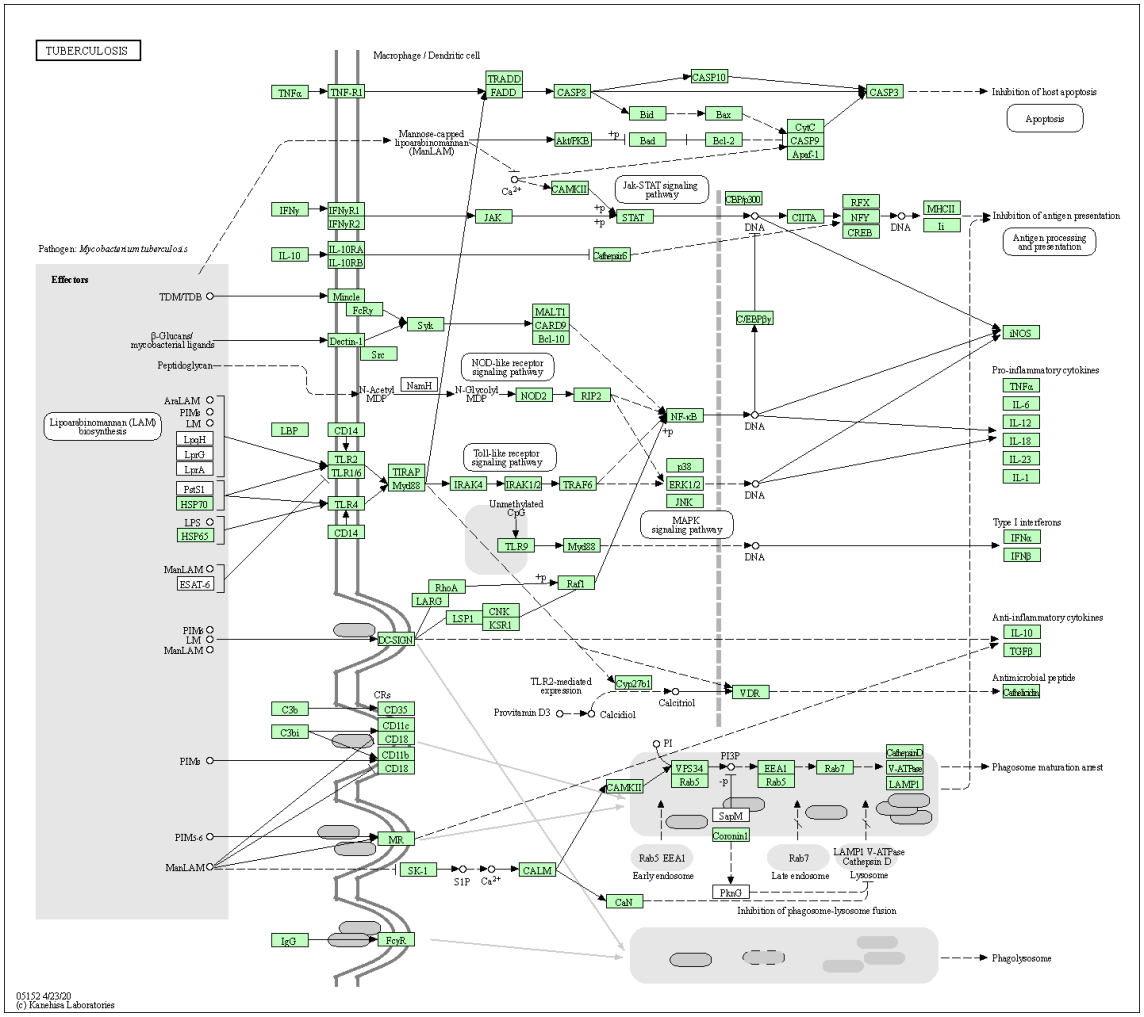


Figure 6.1: Tuberculosis pathway with identifier hsa05152 [137]. The map shows all the interactions and relationships involved in this specific disease.

KEGG **PATHWAY: hsa05152** Help

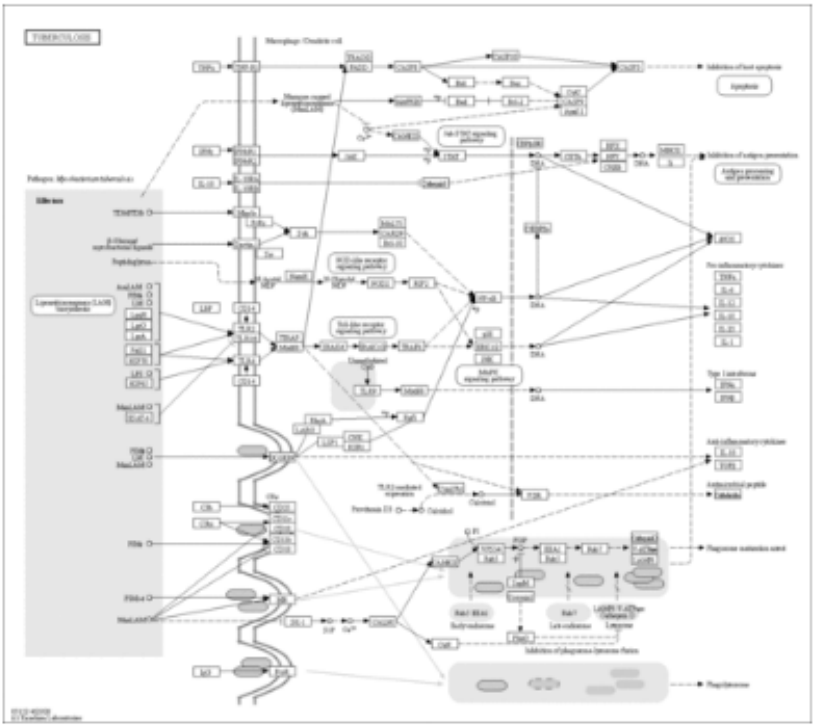
Entry	hsa05152 Pathway
Name	Tuberculosis - Homo sapiens (human)
Description	Tuberculosis, or TB, is an infectious disease caused by Mycobacterium tuberculosis. One third of the world's population is thought to be infected with TB. About 90% of those infected result in latent infections, and about 10% of latent infections develop active diseases when their immune system is impaired due to the age, other diseases such as AIDS or exposure to immunosuppressive drugs. TB is transmitted through the air and primarily attacks the lungs, then it can spread by the circulatory system to other parts of body. Once TB bacilli have entered the host by the respiratory route and infected macrophages in the lungs, they interfere with phagosomal maturation, antigen presentation, apoptosis and host immune system to establish persistent or latent infection.
Class	Human Diseases; Infectious disease: bacterial BRITE hierarchy
Pathway map	hsa05152 Tuberculosis  Ortholog table
Disease	H00342 Tuberculosis
Organism	Homo sapiens (human) [GN:hsa]
Gene	7124 TNF; tumor necrosis factor [KO:K03156] 7132 TNFRSF1A; TNF receptor superfamily member 1A [KO:K03158]

Figure 6.2: It shows all descriptive and non-descriptive information related to the selected pathway [138], such as: i) name, ii) description, iii) map, iv) organism, v) list of genes involved and many other specific information. This figure shows only a sketch of the hugeness of the potential retrievable information.

KEGG is divided into 4 macro-databases and various sub-databases with reference to the following functionality:

1. **System Information**
    - **Pathway**: pathway maps for cellular and organismal functions – https://www.kegg.jp/kegg/pathway.html
    - **Module**: modules or functional units of genes – https://www.kegg.jp/kegg/module.html
    - **Brite**: hierarchical classifications of biological entities – https://www.kegg.jp/kegg/brite.html
2. **Genomic information**
    - **Genome**: complete genomes – https://www.kegg.jp/kegg/genome.html
    - **Genes**: genes and proteins in the complete genomes – https://www.kegg.jp/kegg/genome.html
    - **Orthology**: ortholog groups of genes in the complete genomes – https://www.kegg.jp/kegg/ko.html
3. **Chemical information**
    - **Compound, Glycan**: chemical compounds and glycans – https://www.kegg.jp/kegg/compound/, https://www.kegg.jp/kegg/glycan/
    - **Reaction, Rpair, Rclass**: chemical reactions – https://www.kegg.jp/kegg/reaction/

- **Enzyme**: enzyme nomenclature – https://www.kegg.jp/kegg/annotation/enzyme.html

4. **Health information**

- **Disease**: human diseases – https://www.kegg.jp/kegg/disease/
- **Drug**: approved drugs – https://www.kegg.jp/kegg/drug/

PETAL uses the BFS search logic to carry out the breadth analysis, which starts with the initial retrieving of data, particularly the biological pathways present in KEGG.

The breadth search is an uninformed search in which the nodes of the logical tree are analyzed following order of proximity to the root node (Figure 6.3). The nodes closest to the root node are expanded first, e.g., from left to right, and then all other successor nodes are extended to a depth d. Each successor node is added to the First Input First Output (FIFO) queue of nodes yet to be analysed as soon as it is expanded. For demonstrative simplicity, in the following state-space, the order of exploration of the nodes is described. Root node 1 is the first to be explored, successor node 2 is the second, successor node 3 is the third, and so on. For each exploration, child nodes 4 and 5 are identified and added to the queue in the open list of nodes still to be explored. The pseudo-code is described in detail in Figure 6.4.

The breadth search is optimal when the step cost is equal for each vertex and is a non-decreasing function of the depth of the search tree. Its complexity is an exponential function of the node depth (d), branching factor (b), and the maximum number of expanded nodes is O (bd). The complexity is defined as follows:

$$O(bd) = 1 + b^1 + b^2 + \ldots + b^{d-1} \tag{6.1}$$

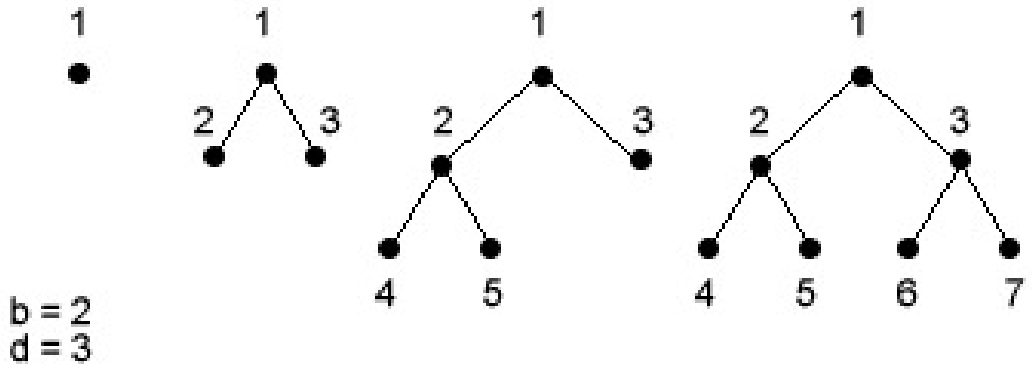


Figure 6.3: The logic of breadth-first search.

```
1  procedure BFS(G, root) is
2      let Q be a queue
3      label root as discovered
4      Q.enqueue(root)
5      while Q is not empty do
6          v := Q.dequeue()
7          if v is the goal then
8              return v
9          for all edges from v to w in G.adjacentEdges(v) do
10             if w is not labeled as discovered then
11                 label w as discovered
12                 Q.enqueue(w)
```

Figure 6.4: BFS pseudo-code.

Figure 6.5 shows the schematic logic of PETAL. It can be seen that before starting the analysis, PETAL requires the user to provide specific details such as the pathway, gene, and maximum depth. In this way, less common pathways can be discovered, and the number of times they appear in other biological pathways can be shown. The advantage is that the results are obtained automatically and quickly, saving time and reducing complexity. This process would have taken much longer if conducted manually. In addition, a graphical interface allows the exploration of the

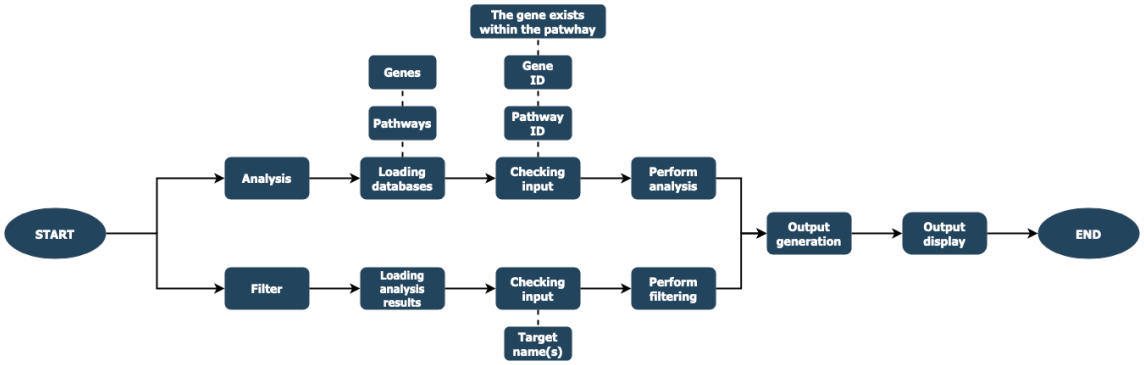


Figure 6.5: PETAL functional scheme.

output tree, showing all connections between genes in different depths, the number of occurrences, and other information about the pathway.

The initial parameters are passed through batch mode and read with Python's argparse library, and two distinct behaviors are implemented:

1. **Analysis**

   - **Load (optional – boolean)**: if true, this parameter allows the last analysis previously completed to be reloaded and extended to a greater depth.
   - **Pathway (required – string)**: this parameter represents the biological pathway (in hsa format) from which the analysis will start (e.g., hsa05152);
   - **Gene (required – string)**: it represents the starting gene present in the selected pathway. If the inserted gene is not detected in the biological pathway, the analysis will fail (e.g., hsa:3458);
   - **Depth (required – integer)**: it indicates the maximum search depth of the analysis;

- **#CPU (optional – integer)**: it indicates the maximum number of CPUs used during the analysis. If it is set to 0 or its value exceeds the number of installed CPUs, the analysis will be automatically charged to the maximum number.

2. **Filter**

- **Target (required – string)**: it represents the name(s) of the target genes to be searched for in the CSV files.
- **#CPU (optional – integer)**: indicates the maximum number of CPUs used during the analysis. If it is set to 0 or its value exceeds the number of installed CPUs, the analysis will be automatically charged to the maximum number.

In the current version of PETAL (v2.0), the database download has been removed and integrated within the source code using a single CSV file. This feature has ensured an increase in performance, i.e., the time to complete analysis has been drastically reduced.

PETAL, at this step, will load two dataframes into the RAM, one for pathways and one for genes, respectively. During the analysis, all the queries implemented are applied to these two dataframes to complete the running successfully. If an update to the database is released, PETAL automatically downloads it and immediately proceeds with the analysis. If there are no updates, it proceeds to the next step.

PETAL tool checks the user's input data in 3 different steps. Firstly, it verifies that the pathway id is in the "hsa" format (i.e., hsa05152 - Tuberculosis pathway) [137, 138] and exists within the previously loaded pathway database. The second step

deals with the checking of the gene id in its "hsa" format (i.e., hsa:3458 - Interferon-gamma) and its presence within the gene database. Lastly, PETAL checks whether the selected gene is within the pathway and owns at least one relationship. If these three checks fail (execution stops), an error message is returned specifying the reason of this stop.

In this step, the global DataFrame structure is created. It contains all the connections found during the analysis, grouped by their depth. In figure 6.6, each row shows the following data collection and information:

- **depth**: shows the depth of the connection between the two genes (e.g., 1);
- **starting_gene_id**: specifies the name of the initial gene in hsa format (e.g., hsa:3458);
- **starting_gene_name**: indicates the name of the initial gene (e.g., IFN-gamma);
- **ending_gene_id**: specifies the name of the ending gene in hsa format (e.g., hsa:3459);
- **ending_gene_name**: indicates the name of the ending gene (e.g., IFNGR1);
- **relation**: shows the protein interaction between the two genes (e.g., PPrel);
- **subtype**: represents the nature of the biochemical interaction (e.g., inhibition, phosphorylation);
- **reference_pathway**: deals with the pathway in which the connection between the two genes was found (e.g., hsa05230);
- **fullpath**: specifies the complete hierarchy from the initial gene to the ending gene found by the analysis (e.g., MAPK1/MYC/LDHAL6A);

- **occurrences**: shows the number of occurrences of a specific gene (e.g., 6).

The analysis is performed in parallel through the Joblib library. The analysis deals with a depth equal or superior to 1. In the former case, the analysis takes care of the initial gene, while a list of genes is considered in the latter one. Specifically, the analysis consists of the following sub-phases:

- Depth 1:
  - Reading of the KGML file of the selected pathway.
  - Providing the genes directly relate to the starting gene, including their biological relationships.
  - All equal rows (duplicates) are removed except those per relation. The number of replications per relation will be kept as the number of occurrences found. For example, if relation A-B is found ten times (i.e., ten rows in the dataframe), the nine rows are removed, and in the one remaining row, the number of occurrences becomes 10.
  - Appending the results obtained to the global dataframe.
  - Exporting these results to a CSV file.
- Depth $2, \ldots, n$:
  - Extraction of the descendent genes previously found and saving them in a list.
  - Parallel processing of each gene in the list through all steps reported in the depth sub-step 1.

A JSON file must be created to view the tree by inserting all the links saved in the global data frame after the analysis. Figure 6.7 shows a screenshot of PETAL

	depth	starting_gene_id	starting_gene_name	starting_isoforms_id	starting_isoforms_name	ending_gene_id	ending_gene_name	ending_isoforms_id	ending_isoforms_name	relation	subtype	reference_pathway	fullpath	occurrences
2	1	hsa:3458	IFNG			hsa:3459	IFNGR1			PPrel	activation	hsa05152	IFNG/IFNGR1	1
3	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:6772	STAT1			PPrel	activation;indirect effect	hsa04380	IFNG/IFNGR1/STAT1	1
4	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:355	FAS			PPrel	indirect effect	hsa04650	IFNG/IFNGR1/FAS	1
5	2	hsa:3454	IFNAR1	hsa:3455;hsa:3459;hsa:3460	IFNAR2;IFNGR1;IFNGR2	hsa:8743	TNFSF10			PPrel	indirect effect	hsa04650	IFNG/IFNGR1/TNFSF10	1
6	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:6772	STAT1			PPrel	activation	hsa04659	IFNG/IFNGR1/STAT1	1
7	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:4843	NOS2			PPrel	activation;indirect effect	hsa05142	IFNG/IFNGR1/NOS2	1
8	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:4790	NFKB1	hsa:5970	RELA	PPrel	activation;indirect effect	hsa04066	IFNG/IFNGR1/NFKB1	1
9	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:3716	JAK1	hsa:3717	JAK2	PPrel	binding/association	hsa05164	IFNG/IFNGR1/JAK1	1
10	2	hsa:1439	CSF2RB	hsa:149233;hsa:2057;hsa:3454;hsa:3455;hsa	IL23R;EPOR;IFNAR1;IFNAR2;IFN	hsa:3716	JAK1	hsa:3717;hsa:3718	JAK2;JAK3	PPrel	activation	hsa05200	IFNG/IFNGR1/JAK1	1
11	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:3265	HRAS	hsa:3845;hsa:4893	KRAS;NRAS	PPrel	activation	hsa05235	IFNG/IFNGR1/HRAS	1
12	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:5290	PIK3CA	hsa:5291;hsa:5293;hsa:5295	PIK3CB;PIK3CD;PIK3R1;PIK3R	PPrel	activation	hsa05235	IFNG/IFNGR1/PIK3CA	1
13	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:3716	JAK1	hsa:3717	JAK2	PPrel	activation	hsa05235	IFNG/IFNGR1/JAK1	1
14	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:6772	STAT1			PPrel	activation	hsa05321	IFNG/IFNGR1/STAT1	1
15	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:3716	JAK1	hsa:3717	JAK2	PPrel	binding/association	hsa05145	IFNG/IFNGR1/JAK1	1
16	2	hsa:116379	IL22RA2	hsa:1271;hsa:1438;hsa:1439;hsa:1441;hsa:14	CNTFR;CSF2RA;CSF2RB;CSF3R;I	hsa:3716	JAK1	hsa:3717;hsa:3718;hsa:7297	JAK2;JAK3;TYK2	PPrel	binding/association	hsa04630	IFNG/IFNGR1/JAK1	1
17	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:3716	JAK1	hsa:3717	JAK2	PPrel	binding/association	hsa05168	IFNG/IFNGR1/JAK1	1
18	2	hsa:3454	IFNAR1	hsa:3455;hsa:3459;hsa:3460	IFNAR2;IFNGR1;IFNGR2	hsa:3716	JAK1	hsa:3717;hsa:3718;hsa:7297	JAK2;JAK3;TYK2	PPrel	activation	hsa04217	IFNG/IFNGR1/JAK1	1
19	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:6772	STAT1			PPrel	activation	hsa04658	IFNG/IFNGR1/STAT1	1
20	2	hsa:3459	IFNGR1	hsa:3460	IFNGR2	hsa:3716	JAK1	hsa:3717	JAK2	PPrel	binding/association	hsa04658	IFNG/IFNGR1/JAK1	1

Figure 6.6: A sketch of the results obtained from the in-depth analysis.

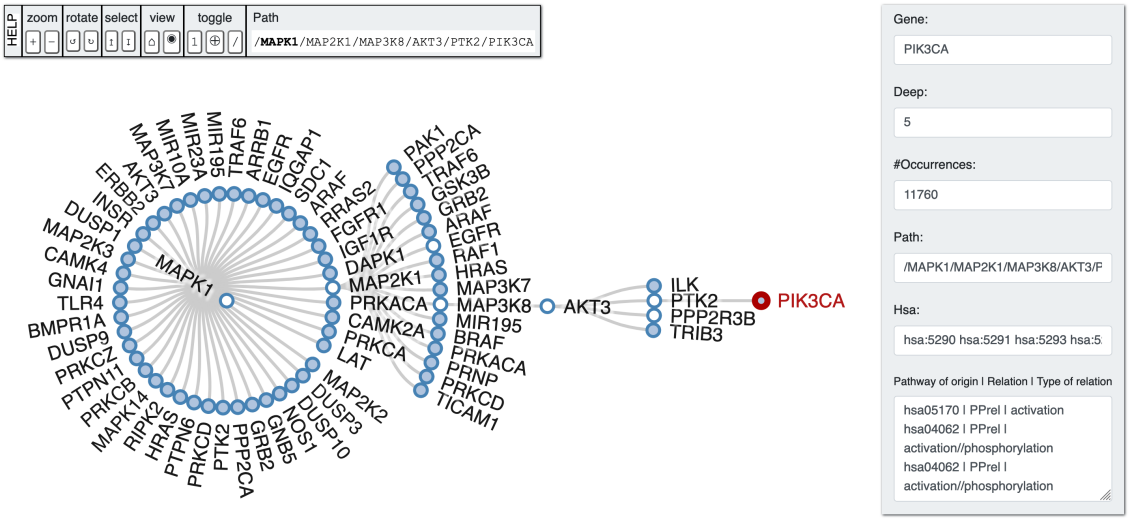


Figure 6.7: An output tree example (MAPK1 pathway) was obtained by running the algorithm with a depth equal to 5. The figure depicts an interactive instance of the output tree generated by analyzing the biological pathway 'MAPK'. The starting gene is represented by MAPK1 (alias ERK), and the depth search level is equal to five. As one can see, all the genes in output are classified by their depth; the gene in red (PIK3CA) represents the ending gene level reached by the analysis. The box on the right shows specific information related to the pathway understudy for every single gene. The upper panel represents the graphical user interface control buttons.

GUI: on the left, one can see the connection tree and the command toolbox. On the right, one can see the information of a specific gene divided by: i) whole path, ii) depth, iii) number of occurrences, and iv) related organism code (e.g., "hsa"), which identifies the gene accordingly to KEGG nomenclature. At the moment, a parser is being developed to allow UISS to correctly interpret the results obtained from PETAL. PETAL predictive capability, considering as test cases three specific

cancer scenarios coming from the recent literature. In particular, it has been selected three studies that suggest particular genes as potential driver of drug resistance to EGFR tyrosine kinases inhibitors [139], used in routine cancer clinical practice. The cancer scenarios, it was considered are chronic myelogenous leukemia (CML), non-small cell lung cancer (NSCLC), and head and neck squamous cell carcinoma (HNSCC), as reported in [140], [141], and [142]. For each of them, the authors respectively suggested SNCA [143], BCL-6 [144], and YAP-1 [145] as potential driver genes of drug resistance to EGFR tyrosine kinases inhibitors. The main factors provoking drug resistance to EGFR tyrosine kinases inhibitors (TKIs) [146] are still poorly understood, even though several mechanisms of EGFR-TKI-resistance have been elucidated by analyzing tumor samples obtained at different stages of tumor progression [147, 148]. Here, it has been used PETAL to test its capability to identify these potential target genes involved in tumor progression and EGFR inhibitors resistance. PETAL detected the same targets investigated in the recent literature for the ancestor and descendent nodes of SNCA, BCL-6, and YAP- 1, across the EGFR pathway [149]. PETAL, to validate it, has been carried out the analysis searching for the three potential driver genes of drug resistance to EGFR tyrosine kinases inhibitors, starting from EGFR gene protein (hsa:956) (Figure 6.8) and related pathway of origin (EGFR tyrosine kinase inhibitor resistance, hsa01521, `https://www.kegg.jp/kegg-bin/show_pathway?hsa01521`), according to KEGG nomenclature. The three selected genes are:

- BCL6 with identifier "hsa:604" related to NSCLC disease;
- SNCA with the identifier "hsa:6622" related to CML disease;
- YAP1 with identifier "hsa:10413" related to HNSCC disease.

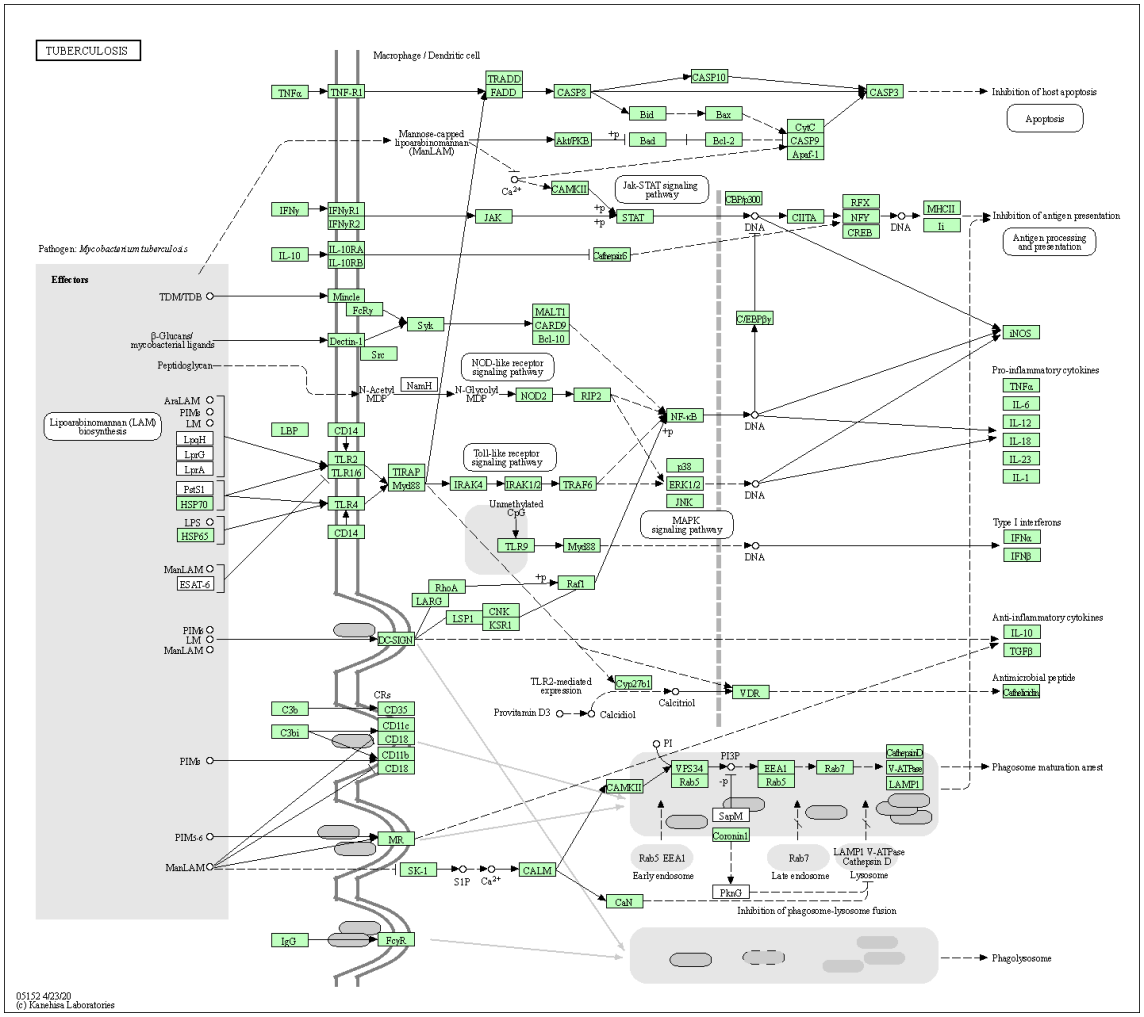


Figure 6.8: EGFR tyrosine kinase inhibitor resistance pathway: the starting gene EGFR in all its isoforms is highlighted in red.

Case test with BLC6 target gene, PETAL discovered 13 different paths to reach BLC6 from EGFR (Figure 6.9). It is worth mentioning that the entire data table contains more than five entries. It has been observed that, independently of the depth level of the analysis, BCL6 is always connected with the FOXO6 gene [150] through direct interaction. Furthermore, the "Pathway of origin" column highlights the involvement of the FOXO signaling pathway (hsa04068) and the related gene expression (GErel) between BCL6 and FOXO6. These relationships highlight the

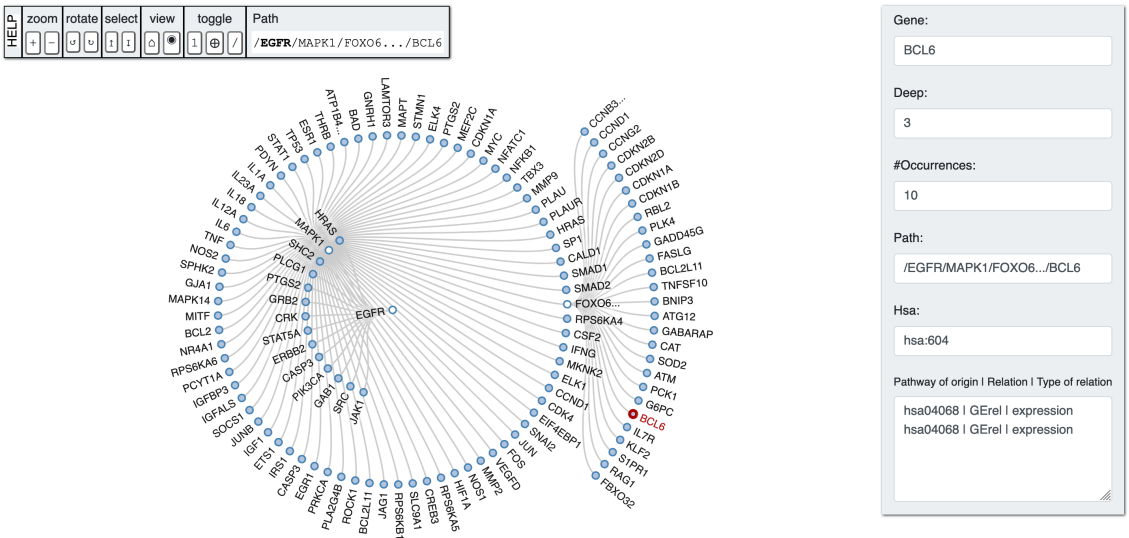


Figure 6.9: An output tree example (starting from the EGFR pathway) was obtained by running the algorithm with a depth equal to 15. BCL6 target gene is highlighted in red.

already prominent role of FOXO-group proteins associated with cell cycle progression and DNA-repair [151, 152].

For what concerns SNCA, PETAL discovered 12 different paths to reach SNCA from EGFR. Through the filtering of data, it is observed that SNCA is always linked to PRKN protein [153] through a protein-protein relation (PPrel), independently of the depth level of the analysis and path. Even though PRKN is linked to SNCA through the Parkinson's disease pathway (hsa05012), this protein gene may also act as a tumor suppressor protein, according to the literature [154]. Figure 6.10 shows a screenshot of PETAL GUI after the in-depth search analysis of the SNCA gene.

Case test with YAP1 target gene, PETAL discovered 52 different paths to reach YAP1 (Figure 6.11), starting from EGFR and EGFR tyrosine kinase inhibitor resistance pathway [155]. YAP1 is detected within a depth level equal to 6. In this case, AMOT protein represents the closer gene connected to YAP1 and is involved in the modulation of YAP1 at that depth level. Moreover, YAP1 was also detected,

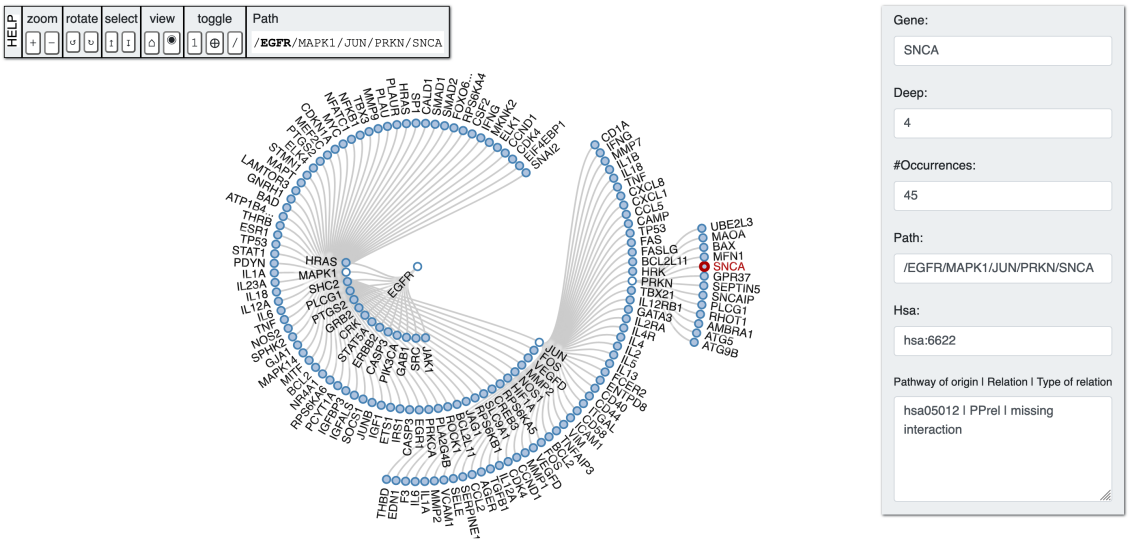


Figure 6.10: An output tree example (starting from EGFR pathway) was obtained by running the algorithm with a depth equal to 15. SNCA target gene is highlighted in red.

launching the algorithm with a depth level equal to 15. In this case, the closest gene connected to YAP1 is the YWHAQ protein [156]. It is worth mentioning that Hippo signaling pathway [157] is always present in this analysis, independently of the depth level of the analysis and starting gene. To our best knowledge, many of the genes involved in Hippo signaling are recognized as tumor suppressor genes, and they are mutated in many human cancers [158]. In this context, YAP1 is considered an oncogene and has already been detected in high concentrations in some human cancers [159].

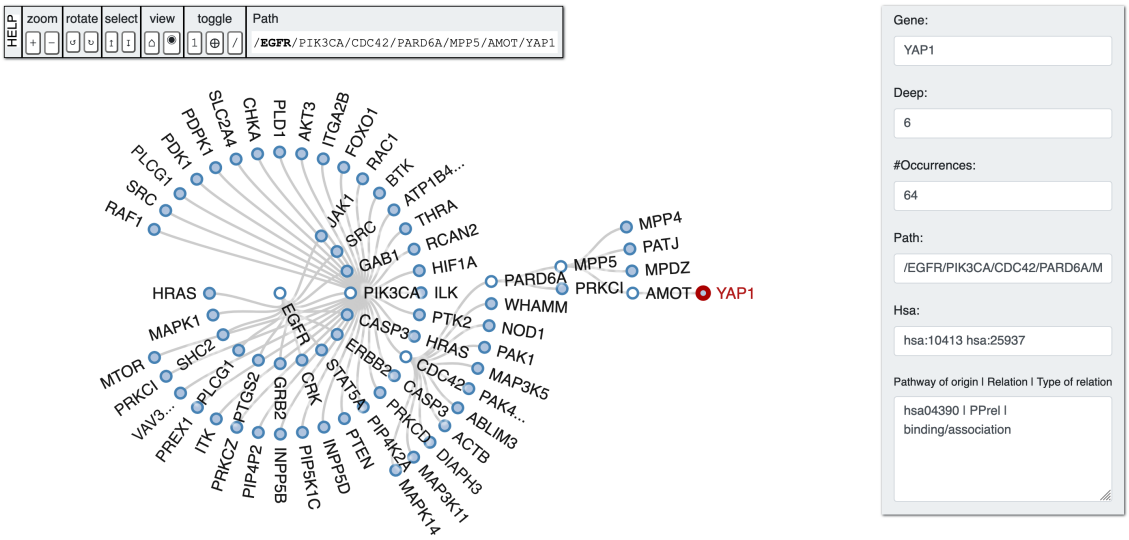


Figure 6.11: An output tree example (starting from EGFR pathway) was obtained by running the algorithm with a depth equal to 15. YAP1 target gene is highlighted in red.

Chapter 7

A pattern-recognition tool: Pattern rEcognition frAmeworK (PEAK)

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus that causes coronavirus disease 2019 (COVID-19), a contagious disease plaguing the world for over two years. Despite the great multitude of critical situations and difficulties, this pandemic also made the possibility to highlight the importance of specific tools based on artificial intelligence to predict the spread of COVID-19 and overcome the related conseqeunces [160, 161]. The main open issue of these tools deals with the requirement of advanced knowledge of regression and classification algorithm theory. Hence, researchers are often looking for tools and software solutions able to satisfy this important aspect of data analysis. Pattern rEcognition frAmeworK (PEAK), a newly developed tool for data analysis, through the use of specially designed open-source tools and Python libraries [102], such as Pandas, Joblib, Matplotlib, and Scikit, could be used to achieve this objective.

PEAK includes a scalable, parallelized engine to easily add new features and specific modules to simplify and automate data analysis. During my Ph.D. programme, PEAK usability and efficiency were tested and verified by performing a particular case study using the dataset of the 10.000 DT samples obtained running Universal Immune System Simulator for SARS-CoV-2 (UISS-SARS-CoV-2) [57]. As previously mentioned, UISS can be extended to reproduce different diseases and related treatments [91, 162, 163, 164, 52, 165, 95]. In this specific case, UISS-SARS-CoV-2 is an extension of UISS capable to reflects the dynamics of COVID-19 infection.

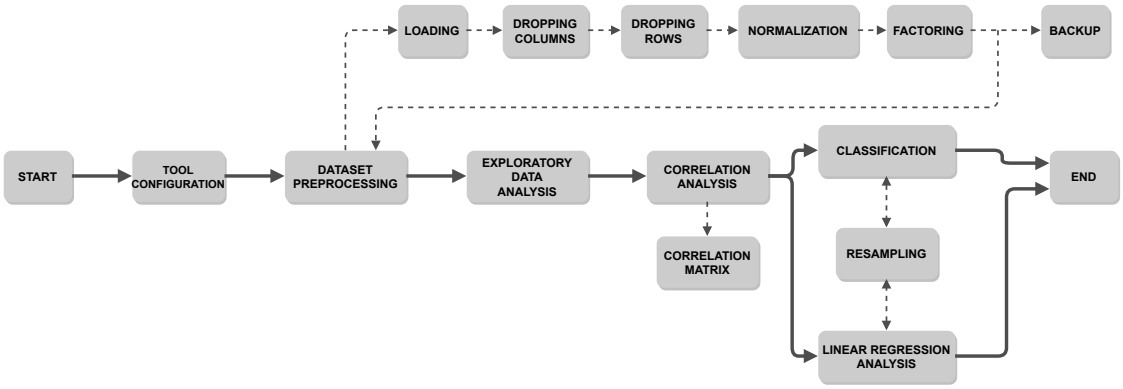


Figure 7.1: PEAK functional scheme. The tool consists of different steps: dataset import, data processing, data exploration, correlation analysis, regression analysis, and/or classification.

PEAK workflow is straightforward to carry out different analyses, as depicted in Figure 7.1. First, it consists of the input dataset processing followed by the regression or classification analysis according to the question of interest. Specifically, PEAK requires four parameters section to get started: i) settings, ii) dataset, iii) regression, and iv) classification. Each section requires additional sub-parameters. The advantage of using this tool lies in getting the results automatically and quickly, saving time, and decreasing complexity. This process would have taken much longer if conducted manually.

As a case study, the tool was used on a well-defined dataset retrieved from UISS-SARS-CoV-2 [57]. Specifically, a cohort of 10,000 digital twins with different immunological characteristics was generated. For each dataset entry, the mean was calculated over the columns of our interest, which concern:

- Cytotoxic T cell (TC)
- IgG antibodies (IgG)
- Interferon gamma (IFN-γ)
- Lung epithelial cells (LEP)
- Interleukin 6 (IL-6)
- Interleukin 12 (IL-12)

Pearson correlation coefficient [166] has been calculated for each column of the dataset. Correlation matrix (Figure 7.2) is then generated using the same indices abovementioned. Observed correlations are not related by a cause-and-effect relationship but they represent the capability of one variable to change as a function of the other considered variable. According to the COVID19 dataset, the correlation between TC and LEP is negative -0.73. This reflects the immunological behavior generally observed when TC becomes active, migrating into the site of infection, recognizing and killing infected LEP to attempt the eradication of the reservoir of infection. The correlation between IL-6 and LEP instead is strongly negative at -0.96. This reflects the well-observed scenario in which IL-6 is correlated with a fatal prognosis. Moreover, the correlation between TH1 and IFN-γ is strongly positive at 0.97. This reflects the well-known behavior that links TH1 with the IFN-γ

secretion. Finally, the correlation between TC and IL-6 is positive at 0.77. This reflects the inflammatory response that, in effective immune response, can contribute to eradicate any infection reservoir.

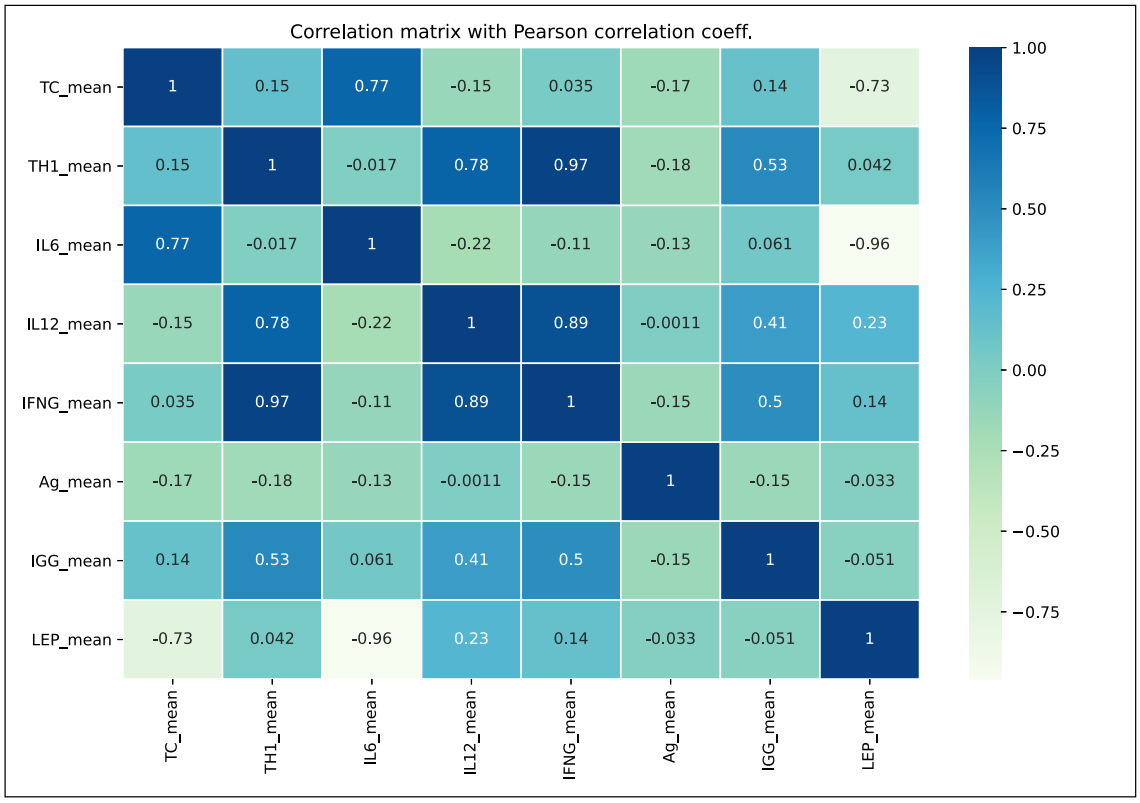


Figure 7.2: Correlation matrix. The figure depict the correlation matrix of Pearson. Each row shows the correlation coefficients be- tween the variables, and each cell shows the correlation between two variables.

PEAK is an open-source tool that offers a wide range of features for data analysis through regression or classification of the provided dataset in a fully automated way. Users can find the source code freely available on the GitHub repository. An intuitive graphical interface will be added to PEAK to allow the user to perform analysis. New analyses (e.g., Bag-of-word) will be added, while existing analyses will be improved in due course.

Chapter 8

Conclusions

Computational models are useful for the understanding of biological systems. The pharmaceutical companies suggest that computational biology can play an excellent role in this field. *In silico* models can offer specific answers to the general behavior of the immune system, analysis of cellular and molecular interactions, effects of treatments, and progression of the disease.

Inside the STriTuVaD project, UISS-TB has contributed in a significant way in providing a level 3 simulation platform (i.e., each individual of the reference population is simulated and represented using biological and physiopathological data coming from real patients). UISS was able to achieve the necessary statistical power (eventually integrated by a number of digital patients) to offer an effective way to estimate time to inactivation of M. tuberculosis with a standard phase II clinical trial, and also to obtain an *in silico* prediction of the effect of any recurrences. In this context, UISS in silico platform is addressing the main goal of the STriTuVaD project providing a computational infrastructure able to simulate the relevant individual human physiology and physiopathology in patients affected by M. tuberculosis and to predict the general outcome of a specific vaccination strategy against the disease. UISS-TB is a state-of-the-art agent-based model able to track the dynamics of TB

infection in humans. Individual digital patients are defined by a vector of features, known to be fundamental in TB infection dynamics and commonly clinically measured.

UISS-TB disease model was developed to simulate the dynamics and the interactions of the MTB. Moreover, the main hallmarks of MTB infection (latent and active phases) have been identified. Current results show UISS capability to simulate the intrinsic immune system behavior that elicits the complete clearance of the infection or, eventually, the chronic establishment of MTB reservoir inside the host. However, UISS-TB has been extended to take care also of the reactivation of TB latent infection and specific mechanisms of action of both antibiotics treatments and vaccination strategies were added.

Further steps deal with the creation of a set of digital subjects to reproduce the biological diversity of the simulated individuals.

Through a "vector of features" that combines both biological and pathophysiological parameters, the digital patient was personalized to reproduce its subject's physiology and pathology.

To successfully carry out the qualification of UISS-TB simulator, more accurate retrospective and prospective analyses will be conducted.

PETAL will be extended with new databases other than KEGG (e.g., Reactome and WikiPathways). PEAK will be enabled to perform further statistical analyses (e.g., Bag-of-word) and modules related to artificial intelligence (e.g., Word2Vec).

In conclusion, in the vaccine development pipeline, UISS general platform with its modules can be considered one of the leading in silico trial solutions along with specific tools (e.g., PEAK and PETAL) that could further extend its functionalities.

Bibliography

[1] A. Koch and V. Mizrahi. "Mycobacterium tuberculosis". In: *Trends Microbiol* 26.6 (June 2018), pp. 555–556.

[2] J. Furin, H. Cox, and M. Pai. "Tuberculosis". In: *Lancet* 393.10181 (Apr. 2019), pp. 1642–1656.

[3] A. Pawlowski, M. Jansson, M. Sköld, M. E. Rottenberg, and G. Källenius. "Tuberculosis and HIV Co-Infection". In: *PLOS Pathogens* 8.2 (Feb. 2012), pp. 1–7. DOI: 10.1371/journal.ppat.1002464. URL: https://doi.org/10.1371/journal.ppat.1002464.

[4] W. H. Organization. *Global tuberculosis report 2018*. World Health Organization, 2018, 231 p.

[5] R. Verma, P. Khanna, and B. Mehta. "Revised national tuberculosis control program in India: the need to strengthen". In: *Int J Prev Med* 4.1 (2013), pp. 1–5.

[6] WHO. *Roadmap to implement the tuberculosis action plan for the WHO European Region 2016–2020*. https://www.euro.who.int/__data/assets/pdf_file/0020/318233/50148-WHO-TB-Plan_May17_web.pdf. Online.

[7] IlSole24Ore. *Tubercolosi, immigrazione e Italia. Tutti i numeri*. https://www.infodata.ilsole24ore.com/2018/04/03/tubercolosi-immigrazione-italia-tutti-numeri/. Online.

[8] P. R. Donald, B. J. Marais, and C. E. Barry. "Age and the epidemiology and pathogenesis of tuberculosis". In: *Lancet* 375.9729 (2010), pp. 1852–1854.

[9] K. DHEDA, S. K. SCHWANDER, B. ZHU, R. N. Van ZYL-SMIT, and Y. ZHANG. "The immunology of tuberculosis: From bench to bedside". In: *Respirology* 15.3 (2010), pp. 433–450. DOI: https://doi.org/10.1111/j.1440-1843.2010.01739.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1843.2010.01739.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1440-1843.2010.01739.x.

[10] N. Fogel. "Tuberculosis: a disease without boundaries". In: *Tuberculosis (Edinb)* 95.5 (2015), pp. 527–531.

[11] P. Narasimhan, J. Wood, C. R. Macintyre, and D. Mathai. "Risk factors for tuberculosis". In: *Pulm Med* 2013 (2013), p. 828939.

[12] K. Lönnroth, E. Jaramillo, B. G. Williams, C. Dye, and M. Raviglione. "Drivers of tuberculosis epidemics: the role of risk factors and social determinants". In: *Soc Sci Med* 68.12 (2009), pp. 2240–2246.

[13] A. O'Garra, P. S. Redford, F. W. McNab, C. I. Bloom, R. J. Wilkinson, and M. P. Berry. "The immune response in tuberculosis". In: *Annu Rev Immunol* 31 (2013), pp. 475–527.

[14] P. L. Lin and J. L. Flynn. "Understanding latent tuberculosis: a moving target". In: *J Immunol* 185.1 (2010), pp. 15–22.

[15] J. M. Tufariello, J. Chan, and J. L. Flynn. "Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection". In: *Lancet Infect Dis* 3.9 (2003), pp. 578–590.

[16] M. Cailleaux-Cezar, D. de A Melo, G. M. Xavier, C. L. de Salles, F. C. de Mello, A. Ruffino-Netto, J. E. Golub, A. Efron, R. E. Chaisson, and M. B. Conde. "Tuberculosis incidence among contacts of active pulmonary tuberculosis". In: *Int J Tuberc Lung Dis* 13.2 (2009), pp. 190–195.

[17] E. Nemes, H. Geldenhuys, V. Rozot, K. T. Rutkowski, F. Ratangee, N. Bilek, S. Mabwe, L. Makhethe, M. Erasmus, A. Toefy, H. Mulenga, W. A. Hanekom, S. G. Self, L.-G. Bekker, R. Ryall, S. Gurunathan, C. A. DiazGranados, P. Andersen, I. Kromann, T. Evans, R. D. Ellis, B. Landry, D. A. Hokey, R. Hopkins, A. M. Ginsberg, T. J. Scriba, and M. Hatherill. "Prevention of M. tuberculosis Infection with H4:IC31 Vaccine or BCG Revaccination". In: *New England Journal of Medicine* 379.2 (2018). PMID: 29996082, pp. 138–149. DOI: 10.1056/NEJMoa1714021. eprint: https://doi.org/10.1056/NEJMoa1714021. URL: https://doi.org/10.1056/NEJMoa1714021.

[18] R. Rowland and H. McShane. "Tuberculosis vaccines in clinical trials". In: *Expert Rev Vaccines* 10.5 (2011), pp. 645–658.

[19] A. M. Ginsberg, M. Ruhwald, H. Mearns, and H. McShane. "TB vaccines in clinical development". In: *Tuberculosis* 99 (2016). Supplement Issue: The 4th Global Forum on TB Vaccines: Shanghai, China. 21 – 24 April 2015, S16–S20. ISSN: 1472-9792. DOI: https://doi.org/10.1016/j.tube.2016.05.013. URL: https://www.sciencedirect.com/science/article/pii/S1472979216301913.

[20] S. Tiberi, A. Scardigli, R. Centis, L. D'Ambrosio, M. Muñoz-Torrico, M. Salazar-Lezama, A. Spanevello, D. Visca, A. Zumla, G. B. Migliori, and J. A.

Caminero Luna. "Classifying new anti-tuberculosis drugs: rationale and future perspectives". In: *Int J Infect Dis* 56 (2017), pp. 181–184.

[21] S. Tiberi, M. Muñoz-Torrico, R. Duarte, M. Dalcolmo, L. D'Ambrosio, and G.-B. Migliori. "New drugs and perspectives for new anti-tuberculosis regimens". In: *Pulmonology* 24.2 (2018). World TB day 2018: targeting tuberculosis, pp. 86–98. ISSN: 2531-0437. DOI: https://doi.org/10.1016/j.rppnen.2017.10.009. URL: https://www.sciencedirect.com/science/article/pii/S2173511517301690.

[22] A. A. White. "The essence of chaos by Edward N. Lorenz. UCL Press. 1993. 227 pp. Hardback £16.95. ISBN 1 85728 187 X". In: *Meteorological Applications* 1.3 (1994), pp. 289–290. DOI: https://doi.org/10.1002/met.5060010310. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.5060010310. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.5060010310.

[23] J. Gleick. *Chaos: Making a new science*. Penguin Books, 1988.

[24] K. T. Alligood, T. D. Sauer, and J. A. Yorke. *Chaos: An introduction to dynamical systems*. Springer, 1997.

[25] H. Holland. *Emergence: From chaos to order*. Basic Books, 1998.

[26] D. C. Wynn and P. J. Clarkson. "Process models in design and development". In: *Research in Engineering Design* 29.2 (2018), pp. 161–202. ISSN: 1435-6066. DOI: 10.1007/s00163-017-0262-7. URL: https://doi.org/10.1007/s00163-017-0262-7.

[27] L. Preziosi. *Cancer modelling and simulation*. CRC Press, 2003.

[28] U. Behn, F. Celada, and P. E. Seiden. *Modelli matematici in immunologia*. Frontiere della Vita, 1998. URL: https://www.treccani.it/enciclopedia/modelli-matematici-in-immunologia_%28Frontiere-della-Vita%29/.

[29] J. Kari. *Cellular Automata*. 2013. URL: https://www.cs.tau.ac.il/~nachumd/models/CA.pdf.

[30] F. Chiacchio, M. Pennisi, G. Russo, S. Motta, and F. Pappalardo. "Agent-based modeling of the immune system: NetLogo, a promising framework". In: *BioMed research international* 2014 (2014).

[31] J. Kari. "Universal pattern generation by cellular automata". In: *Theoretical Computer Science* 429 (2012). Magic in Science, pp. 180–184. ISSN: 0304-3975. DOI: https://doi.org/10.1016/j.tcs.2011.12.037. URL: https://www.sciencedirect.com/science/article/pii/S0304397511010036.

[32] A. Ballier, P. Guillon, and J. Kari. "Limit Sets of Stable and Unstable Cellular Automata". In: *Fundam. Inf.* 110.1–4 (Jan. 2011), 45–57. ISSN: 0169-2968.

[33] J. Kari. "Linear Algebra Based Bounds for One-Dimensional Cellular Automata". In: *Descriptional Complexity of Formal Systems*. Ed. by M. Holzer, M. Kutrib, and G. Pighizzini. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1–7. ISBN: 978-3-642-22600-7.

[34] P. E. Seiden and F. Celada. "A model for simulating cognate recognition and response in the immune system". In: *J Theor Biol* 158.3 (1992), pp. 329–357.

[35] F. Celada and P. E. Seiden. "A computer model of cellular interactions in the immune system". In: *Immunol Today* 13.2 (1992), pp. 56–62.

[36] R. Puzone, B. Kohler, P. Seiden, and F. Celada. "IMMSIM, a flexible model for in machina experiments on immune system responses". In: *Future Generation Computer Systems* 18.7 (2002). Selected papers from CA2000 (6th Int. Workshop on Cellular Automata of IFIP working group 1.5, Osaka, Japan, Aug. 21-22, 2000) and ACRI2000 (4th Int. Conf. on Cellular Automata in Research and Industry, Karlsruhe, Germany, Oct. 4-6, 2000), pp. 961–972. ISSN: 0167-739X. DOI: https://doi.org/10.1016/S0167-739X(02)00075-4. URL: https://www.sciencedirect.com/science/article/pii/S0167739X02000754.

[37] M. Niazi and A. Hussain. "Agent-based computing from multi-agent systems to agent-based models: a visual survey". In: *Scientometrics* 89.2 (2011), pp. 479–499. DOI: 10.1007/s11192-011-0468-9. URL: https://ideas.repec.org/a/spr/scient/v89y2011i2d10.1007_s11192-011-0468-9.html.

[38] L. Gustafsson and M. Sternad. "Consistent micro, macro and state-based population modelling". In: *Math Biosci* 225.2 (2010), pp. 94–107.

[39] E. Bonabeau. "Agent-based modeling: methods and techniques for simulating human systems". In: *Proc Natl Acad Sci U S A* 99 Suppl 3 (2002), pp. 7280–7287.

[40] J. Ferber. "Multi-agent systems - an introduction to distributed artificial intelligence". In: 1999.

[41] C. M. Macal. "Agent Based Modeling and Artificial Life". In: *Encyclopedia of Complexity and Systems Science*. Ed. by R. A. Meyers. New York, NY: Springer New York, 2009, pp. 112–131. ISBN: 978-0-387-30440-3. DOI: 10.

1007/978-0-387-30440-3_7. URL: https://doi.org/10.1007/978-0-387-30440-3_7.

[42] C. M. Macal and M. J. North. "Tutorial on agent-based modelling and simulation". In: *Journal of Simulation* 4.3 (2010), pp. 151–162. DOI: 10.1057/jos.2010.3. eprint: https://doi.org/10.1057/jos.2010.3. URL: https://doi.org/10.1057/jos.2010.3.

[43] J. L. Casti. *Would-be worlds : how simulation is changing the frontiers of science / John L. Casti.* eng. New York: J. Wiley, 1997. ISBN: 0471123080.

[44] L. Bian. "The representation of the environment in the context of individual-based modeling". In: *Ecological Modelling* 159.2 (2003), pp. 279–296. ISSN: 0304-3800. DOI: https://doi.org/10.1016/S0304-3800(02)00298-3. URL: https://www.sciencedirect.com/science/article/pii/S0304380002002983.

[45] J. Farmer, N. H. Packard, and A. S. Perelson. "The immune system, adaptation, and machine learning". In: *Physica D: Nonlinear Phenomena* 22.1 (1986). Proceedings of the Fifth Annual International Conference, pp. 187–204. ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(86)90240-X. URL: https://www.sciencedirect.com/science/article/pii/016727898690240X.

[46] F. Pappalardo, P.-L. Lollini, F. Castiglione, and S. Motta. "Modeling and simulation of cancer immunoprevention vaccine". In: *Bioinformatics* 21.12 (Apr. 2005), pp. 2891–2897. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bti426. eprint: https://academic.oup.com/bioinformatics/article-

pdf/21/12/2891/745834/bti426.pdf. URL: https://doi.org/10.1093/bioinformatics/bti426.

[47] A. Palladini, G. Nicoletti, F. Pappalardo, A. Murgo, V. Grosso, V. Stivani, M. L. Ianzano, A. Antognoli, S. Croci, L. Landuzzi, C. De Giovanni, P. Nanni, S. Motta, and P.-L. Lollini. "In silico Modeling and In vivo Efficacy of Cancer-Preventive Vaccinations". In: *Cancer Research* 70.20 (2010), pp. 7755–7763. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-10-0701. eprint: https://cancerres.aacrjournals.org/content/70/20/7755.full.pdf. URL: https://cancerres.aacrjournals.org/content/70/20/7755.

[48] P.-L. Lollini, S. Motta, and F. Pappalardo. "Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator". In: *BMC Bioinformatics* 7.1 (2006), p. 352. ISSN: 1471-2105. DOI: 10.1186/1471-2105-7-352. URL: https://doi.org/10.1186/1471-2105-7-352.

[49] F. Pappalardo, M. Pennisi, F. Castiglione, and S. Motta. "Vaccine protocols optimization: In silico experiences". In: *Biotechnology Advances* 28.1 (2010), pp. 82–93. ISSN: 0734-9750. DOI: https://doi.org/10.1016/j.biotechadv.2009.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0734975009001797.

[50] M. A. Pennisi, F. Pappalardo, P. Zhang, and S. Motta. "Searching of optimal vaccination schedules". In: *IEEE Engineering in Medicine and Biology Magazine* 28.4 (2009), pp. 67–72. DOI: 10.1109/MEMB.2009.932919.

[51] M. Pennisi, R. Catanuto, F. Pappalardo, and S. Motta. "Optimal vaccination schedules using simulated annealing". In: *Bioinformatics* 24.15 (June 2008), pp. 1740–1742. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn260.

eprint: https://academic.oup.com/bioinformatics/article-pdf/24/15/1740/736104/btn260.pdf. URL: https://doi.org/10.1093/bioinformatics/btn260.

[52] F. Pappalardo, S. Motta, P.-L. Lollini, and E. Mastriani. "Analysis of vaccine's schedules using models". In: *Cellular Immunology* 244.2 (2006). International Conference on Immunogenomics and Immunomics, Budapest, Hungary, October 8-12, 2006, pp. 137–140. ISSN: 0008-8749. DOI: https://doi.org/10.1016/j.cellimm.2007.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0008874907000561.

[53] F. Castiglione, F. Pappalardo, M. Bernaschi, and S. Motta. "Optimization of HAART with genetic algorithms and agent-based models of HIV infection". In: *Bioinformatics* 23.24 (Oct. 2007), pp. 3350–3355. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btm408. eprint: https://academic.oup.com/bioinformatics/article-pdf/23/24/3350/8691474/btm408.pdf. URL: https://doi.org/10.1093/bioinformatics/btm408.

[54] M. Pennisi, F. Pappalardo, and S. Motta. "Agent Based Modeling of Lung Metastasis-Immune System Competition". In: *Artificial Immune Systems*. Ed. by P. S. Andrews, J. Timmis, N. D. L. Owens, U. Aickelin, E. Hart, A. Hone, and A. M. Tyrrell. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–3. ISBN: 978-3-642-03246-2.

[55] F. Pappalardo, G. Russo, M. Pennisi, G. Sgroi, G. A. Parasiliti Palumbo, S. Motta, D. Maimone, and F. Chiacchio. "Agent based modeling of relapsing multiple sclerosis: a possible approach to predict treatment outcome".

In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2018, pp. 1380–1385. DOI: 10.1109/BIBM.2018.8621109.

[56] F. Pappalardo, G. Russo, M. Pennisi, G. A. Parasiliti Palumbo, G. Sgroi, S. Motta, and D. Maimone. "The Potential of Computational Modeling to Predict Disease Course and Treatment Response in Patients with Relapsing Multiple Sclerosis". In: *Cells* 9.3 (2020). ISSN: 2073-4409. DOI: 10.3390/cells9030586. URL: https://www.mdpi.com/2073-4409/9/3/586.

[57] G. Russo, M. Pennisi, E. Fichera, S. Motta, G. Raciti, M. Viceconti, and F. Pappalardo. "In silico trial to test COVID-19 candidate vaccines: a case study with UISS platform". In: *BMC Bioinformatics* 21.17 (2020), p. 527. ISSN: 1471-2105. DOI: 10.1186/s12859-020-03872-0. URL: https://doi.org/10.1186/s12859-020-03872-0.

[58] S. Motta and F. Pappalardo. "Mathematical modeling of biological systems". In: *Briefings in Bioinformatics* 14.4 (Oct. 2012), pp. 411–422. ISSN: 1467-5463. DOI: 10.1093/bib/bbs061. eprint: https://academic.oup.com/bib/article-pdf/14/4/411/479256/bbs061.pdf. URL: https://doi.org/10.1093/bib/bbs061.

[59] B. N. Lambrecht. "Alveolar macrophage in the driver's seat". In: *Immunity* 24.4 (2006), pp. 366–368.

[60] A. M. Guth, W. J. Janssen, C. M. Bosio, E. C. Crouch, P. M. Henson, and S. W. Dow. "Lung environment determines unique phenotype of alveolar macrophages". In: *American Journal of Physiology-Lung Cellular and Molecular Physiology* 296.6 (2009). PMID: 19304907, pp. L936–L946. DOI: 10.1152/ajplung.90625.2008. eprint: https://doi.org/10.1152/

ajplung.90625.2008. URL: https://doi.org/10.1152/ajplung.90625.2008.

[61] I. Smith. "¡i¿Mycobacterium tuberculosis¡/i¿ Pathogenesis and Molecular Determinants of Virulence". In: *Clinical Microbiology Reviews* 16.3 (2003), pp. 463–496. DOI: 10.1128/CMR.16.3.463-496.2003. eprint: https://journals.asm.org/doi/pdf/10.1128/CMR.16.3.463-496.2003. URL: https://journals.asm.org/doi/abs/10.1128/CMR.16.3.463-496.2003.

[62] A. Akbarzadeh, R. Rezaei-Sadabady, S. Davaran, S. W. Joo, N. Zarghami, Y. Hanifehpour, M. Samiei, M. Kouhi, and K. Nejati-Koshki. "Liposome: classification, preparation, and applications". In: *Nanoscale Research Letters* 8.1 (2013), p. 102. ISSN: 1556-276X. DOI: 10.1186/1556-276X-8-102. URL: https://doi.org/10.1186/1556-276X-8-102.

[63] V. P. Torchilin. "Multifunctional nanocarriers". In: *Advanced Drug Delivery Reviews* 58.14 (2006). Particulate Nanomedicines, pp. 1532–1555. ISSN: 0169-409X. DOI: https://doi.org/10.1016/j.addr.2006.09.009. URL: https://www.sciencedirect.com/science/article/pii/S0169409X06001785.

[64] S. Sharma and M. Bose. "Role of cytokines in immune response to pulmonary tuberculosis". In: *Asian Pac J Allergy Immunol* 19.3 (2001), pp. 213–219.

[65] T. Khan, H. Mazhar, S. Saleha, H. N. Tipu, N. Muhammad, and M. N. Abbas. "Interferon-Gamma Improves Macrophages Function against M. tuberculosis in Multidrug-Resistant Tuberculosis Patients". In: *Chemotherapy Research and Practice* 2016 (2016).

[66] M. D. Kearns and V. Tangpricha. "The role of vitamin D in tuberculosis". In: *Journal of Clinical & Translational Endocrinology* 1.4 (2014), pp. 167–

169. ISSN: 2214-6237. DOI: https://doi.org/10.1016/j.jcte.2014.08.002. URL: https://www.sciencedirect.com/science/article/pii/S2214623714000313.

[67] A. Bafica, C. A. Scanga, C. Serhan, F. Machado, S. White, A. Sher, and J. Aliberti. "Host control of Mycobacterium tuberculosis is regulated by 5-lipoxygenase–dependent lipoxin production". In: *The Journal of Clinical Investigation* 115.6 (June 2005), pp. 1601–1606. DOI: 10.1172/JCI23949. URL: https://doi.org/10.1172/JCI23949.

[68] J. Rangel Moreno, I. Estrada García, M. De La Luz García Hernández, D. Aguilar Leon, R. Marquez, and R. Hernández Pando. "The role of prostaglandin E2 in the immunopathogenesis of experimental pulmonary tuberculosis". In: *Immunology* 106.2 (2002), pp. 257–266. DOI: https://doi.org/10.1046/j.1365-2567.2002.01403.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2567.2002.01403.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2567.2002.01403.x.

[69] C. Vilchèze and W. R. Jacobs. "The Isoniazid Paradigm of Killing, Resistance, and Persistence in Mycobacterium tuberculosis". In: *Journal of Molecular Biology* 431.18 (2019). The molecular basis of antibiotic action and resistance, pp. 3450–3461. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2019.02.016. URL: https://www.sciencedirect.com/science/article/pii/S0022283619300932.

[70] M. Pennisi, G. Russo, G. Sgroi, G. Parasiliti, and F. Pappalardo. "2DIs: A SBML Compliant Web Platform for the Design and Modeling of Immune

System Interactions". In: *Intelligent Computing Theories and Application*. Ed. by D.-S. Huang, K.-H. Jo, and J. C. Figueroa-García. Cham: Springer International Publishing, 2017, pp. 145–154. ISBN: 978-3-319-63312-1.

[71] H. L. Rieder. "Fourth-generation fluoroquinolones in tuberculosis". In: *Lancet* 373.9670 (2009), pp. 1148–1149.

[72] H. H. FOX. "The chemical approach to the control of tuberculosis". In: *Science* 116.3006 (1952), pp. 129–134.

[73] K. Johnsson and P. G. Schultz. "Mechanistic Studies of the Oxidation of Isoniazid by the Catalase Peroxidase from Mycobacterium tuberculosis". In: *Journal of the American Chemical Society* 116.16 (1994), pp. 7425–7426. DOI: 10.1021/ja00095a063. eprint: https://doi.org/10.1021/ja00095a063. URL: https://doi.org/10.1021/ja00095a063.

[74] C. Metcalfe, I. K. Macdonald, E. J. Murphy, K. A. Brown, E. L. Raven, and P. C. Moody. "The tuberculosis prodrug isoniazid bound to activating peroxidases". In: *J Biol Chem* 283.10 (2008), pp. 6193–6200.

[75] K Takayama, L Wang, and H. David. "Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of Mycobacterium tuberculosis". In: *Antimicrobial agents and chemotherapy* 2.1 (1972), 29—35. ISSN: 0066-4804. DOI: 10.1128/aac.2.1.29. URL: https://europepmc.org/articles/PMC444261.

[76] Y. Zhang, B. Heym, B. Allen, D. Young, and S. Cole. "The catalase—peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis". In: *Nature* 358.6387 (1992), pp. 591–593. ISSN: 1476-4687. DOI: 10.1038/358591a0. URL: https://doi.org/10.1038/358591a0.

[77] J. B. Nachega and R. E. Chaisson. "Tuberculosis Drug Resistance: A Global Threat". In: *Clinical Infectious Diseases* 36.Supplement$_1$ (Jan. 2003), S24–S30. ISSN: 1058-4838. DOI: 10.1086/344657. eprint: https://academic.oup.com/cid/article-pdf/36/Supplement_1/S24/20904795/36-Supplement_1-S24.pdf. URL: https://doi.org/10.1086/344657.

[78] N. S. Shah, A. Wright, G. H. Bai, L. Barrera, F. Boulahbal, N. Martín-Casabona, F. Drobniewski, C. Gilpin, M. Havelková, R. Lepe, R. Lumb, B. Metchock, F. Portaels, M. F. Rodrigues, S. Rüsch-Gerdes, A. Van Deun, V. Vincent, K. Laserson, C. Wells, and J. P. Cegielski. "Worldwide emergence of extensively drug-resistant tuberculosis". In: *Emerg Infect Dis* 13.3 (2007), pp. 380–387.

[79] H. Marrakchi, G. Lanéelle, and A. Quémard. "InhA, a target of the anti-tuberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II". In: *Microbiology* 146.2 (2000), pp. 289–296. ISSN: 1465-2080. DOI: https://doi.org/10.1099/00221287-146-2-289. URL: https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-146-2-289.

[80] P.-J. Cardona. "RUTI: A new chance to shorten the treatment of latent tuberculosis infection". In: *Tuberculosis* 86.3 (2006). International Workshop on Tuberculosis Vaccines, pp. 273–289. ISSN: 1472-9792. DOI: https://doi.org/10.1016/j.tube.2006.01.024. URL: https://www.sciencedirect.com/science/article/pii/S1472979206000436.

[81] S. A. Prabowo, H. Painter, A. Zelmer, S. G. Smith, K. Seifert, M. Amat, P.-J. Cardona, and H. A. Fletcher. "RUTI Vaccination Enhances Inhibition

of Mycobacterial Growth ex vivo and Induces a Shift of Monocyte Phenotype in Mice". In: *Frontiers in Immunology* 10 (2019), p. 894. ISSN: 1664-3224. DOI: 10.3389/fimmu.2019.00894. URL: https://www.frontiersin.org/article/10.3389/fimmu.2019.00894.

[82] A. S. Nell, E. D'lom, P. Bouic, M. Sabaté, R. Bosser, J. Picas, M. Amat, G. Churchyard, and P.-J. Cardona. "Safety, Tolerability, and Immunogenicity of the Novel Antituberculous Vaccine RUTI: Randomized, Placebo-Controlled Phase II Clinical Trial in Patients with Latent Tuberculosis Infection". In: *PLOS ONE* 9.2 (Feb. 2014), pp. 1–10. DOI: 10.1371/journal.pone.0089612. URL: https://doi.org/10.1371/journal.pone.0089612.

[83] P. Cardona and I. Amat. "Origen y desarrollo de RUTI, una nueva vacuna terapéutica contra la infección por Mycobacterium tuberculosis". In: *Archivos de Bronconeumología* 42.1 (2006), pp. 25–32. ISSN: 0300-2896. DOI: https://doi.org/10.1157/13083277. URL: https://www.sciencedirect.com/science/article/pii/S0300289606705932.

[84] P.-J. Cardona, I. Amat, S. Gordillo, V. Arcos, E. Guirado, J. Díaz, C. Vilaplana, G. Tapia, and V. Ausina. "Immunotherapy with fragmented Mycobacterium tuberculosis cells increases the effectiveness of chemotherapy against a chronical infection in a murine model of tuberculosis". In: *Vaccine* 23.11 (2005), pp. 1393–1398. ISSN: 0264-410X. DOI: https://doi.org/10.1016/j.vaccine.2004.09.008. URL: https://www.sciencedirect.com/science/article/pii/S0264410X04006796.

[85] A. C. Peters, J. W. T. Wimpenny, and J. P. Coombs. "Oxygen Profiles in, and in the Agar Beneath, Colonies of Bacillus Cereus, Staphylococcus Albus

and Escherichia Coli". In: *Microbiology* 133.5 (1987), pp. 1257–1263. ISSN: 1465-2080. DOI: https://doi.org/10.1099/00221287-133-5-1257. URL: https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-133-5-1257.

[86] T. P. Robinson, J. W. T. Wimpenny, and R. G. Earnshaw. "pH gradients through colonies of Bacillus cereus and the surrounding agar". In: *Microbiology* 137.12 (1991), pp. 2885–2889. ISSN: 1465-2080. DOI: https://doi.org/10.1099/00221287-137-12-2885. URL: https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-137-12-2885.

[87] J. G. WALLACE. "The heat resistance of tubercle bacilli in the lungs of infected mice". In: *Am Rev Respir Dis* 83 (1961), pp. 866–871.

[88] L. Leserman. "Liposomes as Protein Carriers in Immunology". In: *Journal of Liposome Research* 14.3-4 (2004), pp. 175–189. DOI: 10.1081/LPR-200039198. eprint: https://doi.org/10.1081/LPR-200039198. URL: https://doi.org/10.1081/LPR-200039198.

[89] C. L. LARSON, J. F. BELL, R. H. LIST, E. RIBI, and W. C. WICHT. "SYMPOSIUM ON RELATIONSHIP OF STRUCTURE OF MICROORGANISMS TO THEIR IMMUNOLOGICAL PROPERTIES. II. HOST-REACTIVE PROPERTIES OF CELL WALLS AND PROTOPLASM FROM MYCOBACTERIA". In: *Bacteriol Rev* 27 (1963), pp. 341–351.

[90] F. Pappalardo, G. Russo, M. Pennisi, G. Sgroi, G. A. Parasiliti Palumbo, S. Motta, and E. Fichera. "An agent based modeling approach for the analysis of tuberculosis – immune system dynamics". In: *2018 IEEE International*

Conference on Bioinformatics and Biomedicine (BIBM). 2018, pp. 1386–1392. DOI: 10.1109/BIBM.2018.8621355.

[91] G. Russo, G. Sgroi, G. A. Parasiliti Palumbo, M. Pennisi, M. A. Juarez, P.-J. Cardona, S. Motta, K. B. Walker, E. Fichera, M. Viceconti, and F. Pappalardo. "Moving forward through the in silico modeling of tuberculosis: a further step with UISS-TB". In: *BMC Bioinformatics* 21.17 (2020), p. 458. ISSN: 1471-2105. DOI: 10.1186/s12859-020-03762-5. URL: https://doi.org/10.1186/s12859-020-03762-5.

[92] R. Ragonnet, J. A. Flegg, S. L. Brilleman, E. W. Tiemersma, Y. A. Melsew, E. S. McBryde, and J. M. Trauer. "Revisiting the Natural History of Pulmonary Tuberculosis: A Bayesian Estimation of Natural Recovery and Mortality Rates". In: *Clinical Infectious Diseases* 73.1 (Aug. 2020), e88–e96. ISSN: 1058-4838. DOI: 10.1093/cid/ciaa602. eprint: https://academic.oup.com/cid/article-pdf/73/1/e88/38846725/ciaa602.pdf. URL: https://doi.org/10.1093/cid/ciaa602.

[93] K. M. Shea, J. S. Kammerer, C. A. Winston, T. R. Navin, and J. Horsburgh C. Robert. "Estimated Rate of Reactivation of Latent Tuberculosis Infection in the United States, Overall and by Population Subgroup". In: *American Journal of Epidemiology* 179.2 (Oct. 2013), pp. 216–225. ISSN: 0002-9262. DOI: 10.1093/aje/kwt246. eprint: https://academic.oup.com/aje/article-pdf/179/2/216/288239/kwt246.pdf. URL: https://doi.org/10.1093/aje/kwt246.

[94] S. K. Katiyar, S. Bihari, S. Prakash, M. Mamtani, and H. Kulkarni. "A randomised controlled trial of high-dose isoniazid adjuvant therapy for multidrug-resistant tuberculosis". In: *The International Journal of Tuberculosis and Lung Disease* 12.2 (2008), pp. 139–145. ISSN: 1027-3719. URL: https://www.ingentaconnect.com/content/iuatld/ijtld/2008/00000012/00000002/art00005.

[95] M. Pennisi, G. Russo, G. Sgroi, A. Bonaccorso, G. A. Parasiliti Palumbo, E. Fichera, D. K. Mitra, K. B. Walker, P.-J. Cardona, M. Amat, M. Viceconti, and F. Pappalardo. "Predicting the artificial immunity induced by RUTI® vaccine against tuberculosis using universal immune system simulator (UISS)". In: *BMC Bioinformatics* 20.6 (2019), p. 504. ISSN: 1471-2105. DOI: 10.1186/s12859-019-3045-5. URL: https://doi.org/10.1186/s12859-019-3045-5.

[96] G. Russo, F. Pappalardo, M. A. Juarez, M. Pennisi, P. J. Cardona, R. Coler, E. Fichera, and M. Viceconti. "Evaluation of the efficacy of RUTI and ID93/GLA-SE vaccines in tuberculosis treatment: in silico trial through UISS-TB simulator". In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2019, pp. 2197–2201. DOI: 10.1109/BIBM47256.2019.8983060.

[97] K. D. Mayer-Barber and D. L. Barber. "Innate and Adaptive Cellular Immune Responses to Mycobacterium tuberculosis Infection". In: *Cold Spring Harbor Perspectives in Medicine* 5.12 (2015). DOI: 10.1101/cshperspect.a018424. eprint: http://perspectivesinmedicine.cshlp.org/content/5/12/a018424.full.pdf+html. URL: http://perspectivesinmedicine.cshlp.org/content/5/12/a018424.abstract.

[98] A. O'Garra, P. S. Redford, F. W. McNab, C. I. Bloom, R. J. Wilkinson, and M. P. Berry. "The Immune Response in Tuberculosis". In: *Annual Review of Immunology* 31.1 (2013). PMID: 23516984, pp. 475–527. DOI: 10.1146/annurev-immunol-032712-095939. eprint: https://doi.org/10.1146/annurev-immunol-032712-095939. URL: https://doi.org/10.1146/annurev-immunol-032712-095939.

[99] E. Petruccioli, T. Chiacchio, I. Pepponi, V. Vanini, R. Urso, G. Cuzzi, L. Barcellini, D. M. Cirillo, F. Palmieri, G. Ippolito, and D. Goletti. "First characterization of the CD4 and CD8 T-cell responses to QuantiFERON-TB Plus". In: *Journal of Infection* 73.6 (2016), pp. 588–597. ISSN: 0163-4453. DOI: https://doi.org/10.1016/j.jinf.2016.09.008. URL: https://www.sciencedirect.com/science/article/pii/S0163445316302493.

[100] R. K. Gupta, H. Kunst, M. Lipman, M. Noursadeghi, C. Jackson, J. Southern, A. Imran, S. Lozewicz, and I. Abubakar. "Evaluation of QuantiFERON-TB Gold Plus for Predicting Incident Tuberculosis among Recent Contacts: A Prospective Cohort Study". In: *Ann Am Thorac Soc* 17.5 (May 2020), pp. 646–650.

[101] M. Grinberg. *Flask web development: developing web applications with python.* " O'Reilly Media, Inc.", 2018.

[102] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697.

[103] Armin Ronacher. *Software developer and Open Source nut. Creator of the Flask framework.* https://github.com/mitsuhiko. Online.

[104] Armin Ronacher. *Werkzeug is a comprehensive WSGI web application library.* https://werkzeug.palletsprojects.com/en/2.0.x/. Online.

[105] Armin Ronacher. *Jinja is a fast, expressive, extensible templating engine.* https://jinja.palletsprojects.com/en/3.0.x/. Online.

[106] Django Software Foundation. *Django*. Version 2.2. May 5, 2019. URL: https://djangoproject.com.

[107] M. Pennisi, M. A. Juárez, G. Russo, M. Viceconti, and F. Pappalardo. "Generation of digital patients for the simulation of tuberculosis with UISS-TB". In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2019, pp. 2163–2167. DOI: 10.1109/BIBM47256.2019.8983100.

[108] P. Richmond, D. Walker, S. Coakley, and D. Romano. "High performance cellular level agent-based simulation with FLAME for the GPU". In: *Briefings in Bioinformatics* 11.3 (Feb. 2010), pp. 334–347. ISSN: 1467-5463. DOI: 10.1093/bib/bbp073. eprint: https://academic.oup.com/bib/article-pdf/11/3/334/606876/bbp073.pdf. URL: https://doi.org/10.1093/bib/bbp073.

[109] M. Kiran, P. Richmond, M. Holcombe, L. S. Chin, D. Worth, and C. Greenough. "FLAME: Simulating Large Populations of Agents on Parallel Hardware Architectures". In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*. AAMAS '10. Toronto, Canada: International Foundation for Autonomous Agents and Multiagent Systems, 2010, 1633–1636. ISBN: 9780982657119.

[110] M. Kabiri Chimeh, P. Heywood, M. Pennisi, F. Pappalardo, and P. Richmond. "Parallelisation strategies for agent based simulation of immune systems". In: *BMC Bioinformatics* 20.6 (2019), p. 579. ISSN: 1471-2105. DOI: 10.1186/s12859-019-3181-y. URL: https://doi.org/10.1186/s12859-019-3181-y.

[111] FlameGPU. *The Flexible Large Scale Agent Modelling Environment for the Graphics Processing Unit (GPU)*. https://flamegpu.com/. Online.

[112] FlameGPU-UISS-TB. *The Flexible Large Scale Agent Modelling Environment for the Graphics Processing Unit (GPU)*. https://flame-gpu-uiss.readthedocs.io/en/latest/model-and-implementation.html. Online.

[113] W. McKinney. "Data Structures for Statistical Computing in Python". In: 2010.

[114] Python3. *Argparse – Parser for command-line options, arguments and sub-commands*. https://docs.python.org/3.9/library/argparse.html. Online.

[115] Python3. *sys – System-specific parameters and functions*. https://docs.python.org/3.9/library/sys.html. Online.

[116] Anton Malakhov, David Liu, Anton Gorshkov, and Terry Wilmarth. "Composable Multi-Threading and Multi-Processing for Numeric Libraries". In: *Proceedings of the 17th Python in Science Conference*. Ed. by Fatih Akici, David Lippa, Dillon Niederhut, and M. Pacer. 2018, pp. 18 –24. DOI: 10.25080/Majora-4af1f417-003.

[117] Python3. *Joblib: running Python functions as pipeline jobs*. https://joblib.readthedocs.io/en/latest/. Online.

[118] C. F. de Sousa Rodrigues, F. J. C. de Lima, and F. T. Barbosa. "Importance of using basic statistics adequately in clinical research". In: *Brazilian Journal of Anesthesiology (English Edition)* 67.6 (2017), pp. 619–625. ISSN: 0104-0014. DOI: https://doi.org/10.1016/j.bjane.2017.01.011. URL: https://www.sciencedirect.com/science/article/pii/S0104001417300167.

[119] B. Kirkwood and J. Sterne. "Essential Medical Statistics". In: 2003.

[120] N. Pandis. "The sampling distribution". In: *American Journal of Orthodontics and Dentofacial Orthopedics* 147.4 (2015), pp. 517–519. ISSN: 0889-5406. DOI: https://doi.org/10.1016/j.ajodo.2015.01.009. URL: https://www.sciencedirect.com/science/article/pii/S0889540615000190.

[121] M. A. Juárez, M. Pennisi, G. Russo, D. Kiagias, C. Curreli, M. Viceconti, and F. Pappalardo. "Generation of digital patients for the simulation of tuberculosis with UISS-TB". In: *BMC Bioinformatics* 21.17 (2020), p. 449. ISSN: 1471-2105. DOI: 10.1186/s12859-020-03776-z. URL: https://doi.org/10.1186/s12859-020-03776-z.

[122] J. D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[123] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. "Array programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362. ISSN: 1476-4687. DOI:

10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

[124] S. Y. Eum, J. H. Kong, M. S. Hong, Y. J. Lee, J. H. Kim, S. H. Hwang, S. N. Cho, L. E. Via, and C. E. Barry. "Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB". In: *Chest* 137.1 (2010), pp. 122–128.

[125] G. Sgroi, G. Russo, and F. Pappalardo. "PETAL: a Python tool for deep analysis of biological pathways". In: *Bioinformatics* 36.22-23 (Dec. 2020), pp. 5553–5555. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btaa1032. eprint: https://academic.oup.com/bioinformatics/article-pdf/36/22-23/5553/36855788/btaa1032.pdf. URL: https://doi.org/10.1093/bioinformatics/btaa1032.

[126] G. Sgroi, M. Pennisi, G. Russo, F. Pappalardo, and G. A. P. Palumbo. "Evaluation of the predictive capability of PETAL tool: a retrospective study on potential tyrosine kinases drug resistance targets". In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2020, pp. 1275–1280. DOI: 10.1109/BIBM49941.2020.9312986.

[127] M. Kanehisa and S. Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 27–30. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.27. eprint: https://academic.oup.com/nar/article-pdf/28/1/27/9895154/280027.pdf. URL: https://doi.org/10.1093/nar/28.1.27.

[128] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van

der Walt and Jarrod Millman. 2010, pp. 56 –61. DOI: 10.25080/Majora-92bf1922-00a.

[129] D. C. Kozen. "Depth-First and Breadth-First Search". In: *The Design and Analysis of Algorithms*. New York, NY: Springer New York, 1992, pp. 19–24. ISBN: 978-1-4612-4400-4. DOI: 10.1007/978-1-4612-4400-4_4. URL: https://doi.org/10.1007/978-1-4612-4400-4_4.

[130] M. Bostock, V. Ogievetsky, and J. Heer. "D^3 Data-Driven Documents". In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2301–2309. DOI: 10.1109/TVCG.2011.185.

[131] B. Lawson and R. Sharp. *Introducing HTML5*. Voices that matter. New Riders, 2011. ISBN: 9780321687296. URL: https://books.google.it/books?id=a2BVmAEACAAJ.

[132] D. Jackson. "Scalable Vector Graphics (SVG): The World Wide Web Consortium's Recommendation for High Quality Web Graphics". In: *ACM SIGGRAPH 2002 Conference Abstracts and Applications*. SIGGRAPH '02. San Antonio, Texas: Association for Computing Machinery, 2002, p. 319. ISBN: 1581135254. DOI: 10.1145/1242073.1242327. URL: https://doi.org/10.1145/1242073.1242327.

[133] Lie. *Cascading Style Sheets*. Pearson Addison Wesley, 2004. ISBN: 0321193121.

[134] D3.js. *Radial Tree with UI*. https://gist.github.com/wmleler/a734fb2bb3319a2cb386. Online.

[135] M. Kanehisa. "A database for post-genome analysis". In: *Trends Genet* 13.9 (1997), pp. 375–376.

[136] KEGG. *KEGG Pathway Database*. https://www.genome.jp/kegg/pathway.html. Online.

[137] KEGG. *KEGG Pathway: Tuberculosis (hsa05152)*. https://www.kegg.jp/kegg-bin/show_pathway?hsa05152. Online.

[138] KEGG. *Tuberculosis pathway description*. https://www.kegg.jp/dbget-bin/www_bget?pathway+hsa05152. Online.

[139] L. Huang and L. Fu. "Mechanisms of resistance to EGFR tyrosine kinase inhibitors". In: *Acta Pharm Sin B* 5.5 (2015), pp. 390–401.

[140] B. M. Noel, S. B. Ouellette, L. Marholz, D. Dickey, C. Navis, T. Y. Yang, V. Nguyen, S. J. Parker, D. Bernlohr, Z. Sachs, and L. L. Parker. "Multi-omic Profiling of Tyrosine Kinase Inhibitor-Resistant K562 Cells Suggests Metabolic Reprogramming To Promote Cell Survival". In: *J Proteome Res* 18.4 (Apr. 2019), pp. 1842–1856.

[141] Y. Zhou Tran, R. Minozada, X. Cao, H. J. Johansson, R. M. Branca, B. Seashore-Ludlow, and L. M. Orre. "Immediate Adaptation Analysis Implicates BCL6 as an EGFR-TKI Combination Therapy Target in NSCLC". In: *Mol Cell Proteomics* 19.6 (June 2020), pp. 928–943.

[142] A. P. Jain, K. Patel, S. Pinto, A. Radhakrishnan, V. Nanjappa, M. Kumar, R. Raja, A. H. Patil, A. Kumari, M. Manoharan, C. Karunakaran, S. Murugan, T. S. Keshava Prasad, X. Chang, P. P. Mathur, P. Kumar, R. Gupta, R. Gupta, A. Khanna-Gupta, D. Sidransky, A. Chatterjee, and H. Gowda. "MAP2K1 is a potential therapeutic target in erlotinib resistant head and neck squamous cell carcinoma". In: *Scientific Reports* 9.1 (2019),

p. 18793. ISSN: 2045-2322. DOI: 10.1038/s41598-019-55208-5. URL: https://doi.org/10.1038/s41598-019-55208-5.

[143] J. W. Touchman, A. Dehejia, O. Chiba-Falek, D. E. Cabin, J. R. Schwartz, B. M. Orrison, M. H. Polymeropoulos, and R. L. Nussbaum. "Human and mouse alpha-synuclein genes: comparative genomic sequence analysis and identification of a novel gene regulatory element". In: *Genome Res* 11.1 (2001), pp. 78–86.

[144] R. J. Leeman-Neill and G. Bhagat. "BCL6 as a therapeutic target for lymphoma". In: *Expert Opin Ther Targets* 22.2 (Feb. 2018), pp. 143–152.

[145] M. Shibata, K. Ham, and M. O. Hoque. "A time for YAP1: Tumorigenesis, immunosuppression and targeted therapy". In: *Int J Cancer* 143.9 (Nov. 2018), pp. 2133–2144.

[146] D. Westover, J. Zugazagoitia, B. C. Cho, C. M. Lovly, and L. Paz-Ares. "Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors". In: *Ann Oncol* 29.suppl$_1$ (Jan. 2018), pp. i10–i19.

[147] C. Pottier, M. Fresnais, M. Gilon, G. Jérusalem, R. Longuespée, and N. E. Sounni. "Tyrosine Kinase Inhibitors in Cancer: Breakthrough and Challenges of Targeted Therapy". In: *Cancers* 12.3 (2020). ISSN: 2072-6694. DOI: 10.3390/cancers12030731. URL: https://www.mdpi.com/2072-6694/12/3/731.

[148] F. Gianì, G. Russo, M. Pennisi, L. Sciacca, F. Frasca, and F. Pappalardo. "Computational modeling reveals MAP3K8 as mediator of resistance to vemurafenib in thyroid cancer stem cells". In: *Bioinformatics* 35.13 (Nov. 2018), pp. 2267–2275. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty969.

eprint: https://academic.oup.com/bioinformatics/article-pdf/35/13/2267/28878383/bty969.pdf. URL: https://doi.org/10.1093/bioinformatics/bty969.

[149] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. "A comprehensive pathway map of epidermal growth factor receptor signaling". In: *Mol Syst Biol* 1 (2005), p. 2005.0010.

[150] F. M. Jacobs, L. P. van der Heide, P. J. Wijchers, J. P. Burbach, M. F. Hoekman, and M. P. Smidt. "FoxO6, a novel member of the FoxO class of transcription factors with distinct shuttling dynamics". In: *J Biol Chem* 278.38 (2003), pp. 35959–35967.

[151] K. K. Ho, S. S. Myatt, and E. W. Lam. "Many forks in the path: cycling with FoxO". In: *Oncogene* 27.16 (2008), pp. 2300–2311.

[152] H. Tran, A. Brunet, J. M. Grenier, S. R. Datta, A. J. Fornace, P. S. DiStefano, L. W. Chiang, and M. E. Greenberg. "DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein". In: *Science* 296.5567 (2002), pp. 530–534.

[153] C. Alves da Costa, E. Duplan, L. Rouland, and F. Checler. "The Transcription Factor Function of Parkin: Breaking the Dogma". In: *Front Neurosci* 12 (2018), p. 965.

[154] A. Gupta, S. Anjomani-Virmouni, N. Koundouros, and G. Poulogiannis. "loss promotes cancer progression via redox-mediated inactivation of PTEN". In: *Mol Cell Oncol* 4.6 (2017), e1329692.

[155] T. Huang, Y. Zhou, J. Zhang, A. S. L. Cheng, J. Yu, K. F. To, and W. Kang. "The physiological role of Motin family and its dysregulation in tumorigenesis". In: *Journal of Translational Medicine* 16.1 (2018), p. 98. ISSN: 1479-5876. DOI: 10.1186/s12967-018-1466-y. URL: https://doi.org/10.1186/s12967-018-1466-y.

[156] H. Liu, Y. Tang, X. Liu, Q. Zhou, X. Xiao, F. Lan, X. Li, R. Hu, Y. Xiong, and T. Peng. "14-3-3 tau (YWHAQ) gene promoter hypermethylation in human placenta of preeclampsia". In: *Placenta* 35.12 (2014), pp. 981–988.

[157] D. Pan. "The hippo signaling pathway in development and cancer". In: *Dev Cell* 19.4 (2010), pp. 491–505.

[158] M. Yin and L. Zhang. "Hippo signaling: a hub of growth control, tumor suppression and pluripotency maintenance". In: *J Genet Genomics* 38.10 (2011), pp. 471–481.

[159] T. Moroishi, C. G. Hansen, and K. L. Guan. "The emerging roles of YAP and TAZ in cancer". In: *Nat Rev Cancer* 15.2 (2015), pp. 73–79.

[160] Z. Li, S. Yang, and J. Wu. "The Prediction of the Spread of COVID-19 using Regression Models". In: *2020 International Conference on Public Health and Data Science (ICPHDS)*. 2020, pp. 247–252. DOI: 10.1109/ICPHDS51617.2020.00055.

[161] H. B. Syeda, M. Syed, K. W. Sexton, S. Syed, S. Begum, F. Syed, F. Prior, and F. Yu. "Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review". In: *JMIR Med Inform* 9.1 (2021), e23811.

[162] M. Pennisi, G. Russo, G. Sgroi, G. A. P. Palumbo, and F. Pappalardo. "In Silico Evaluation of Daclizumab and Vitamin D Effects in Multiple Sclerosis Using Agent Based Models". In: *Computational Intelligence Methods for Bioinformatics and Biostatistics*. Ed. by P. Cazzaniga, D. Besozzi, I. Merelli, and L. Manzoni. Cham: Springer International Publishing, 2020, pp. 285–298. ISBN: 978-3-030-63061-4.

[163] F. Pappalardo, E. Fichera, N. Paparone, A. Lombardo, M. Pennisi, G. Russo, M. Leotta, F. Pappalardo, A. Pedretti, F. De Fiore, and S. Motta. "A computational model to predict the immune system activation by citrus-derived vaccine adjuvants". In: *Bioinformatics* 32.17 (May 2016), pp. 2672–2680. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw293. eprint: https://academic.oup.com/bioinformatics/article-pdf/32/17/2672/7374205/btw293.pdf. URL: https://doi.org/10.1093/bioinformatics/btw293.

[164] F. Pappalardo, D. Flower, G. Russo, M. Pennisi, and S. Motta. "Computational modelling approaches to vaccinology". In: *Pharmacological Research* 92 (2015). Vaccines: present and future challenges, pp. 40–45. ISSN: 1043-6618. DOI: https://doi.org/10.1016/j.phrs.2014.08.006. URL: https://www.sciencedirect.com/science/article/pii/S1043661814001431.

[165] M. Viceconti, M. A. Juárez, C. Curreli, M. Pennisi, G. Russo, and F. Pappalardo. "Credibility of ¡italic¿In Silico¡/italic¿ Trial Technologies—A Theoretical Framing". In: *IEEE Journal of Biomedical and Health Informatics* 24.1 (2020), pp. 4–13. DOI: 10.1109/JBHI.2019.2949888.

[166] "Pearson's Correlation Coefficient". In: *Encyclopedia of Public Health*. Ed. by W. Kirch. Dordrecht: Springer Netherlands, 2008, pp. 1090–1091. ISBN:

978-1-4020-5614-7. DOI: 10.1007/978-1-4020-5614-7_2569. URL: https://doi.org/10.1007/978-1-4020-5614-7_2569.