ETNA BIOTECH
Zydus Cadila Research Center

H2020 Project: 777123

STRiTUVaD

Doc. Name: Deliverable 4 WP2

Doc. Num: 005/19 Version: A

Report on the creation of the subjects-specific and virtual patients' libraries that will be used in the implementation of the in silico clinical trial

11/09/2019

The purpose of this deliverable is to describe the methodology and the technique used to create a set of subjects-specific models with the aim to reproduce biological diversity of the subjects. The library of subjects created are personalized with a "vector of features" that identifies a specific real patient with the final goal to obtain a library of virtual M. tuberculosis infected patients that will be vaccinated and treated accordingly.

Author: Francesco Pappalardo, Giulia Russo, Marzio Pennisi.

replaces version	n.a.	from	n.a.	Page
valid from date (dd/mm/yyyy)		11/09/2019		1 / 36
Recipient	STriTuVaD Consortium			

Topic

Partner: UniCT

Deliverable: The goal of this deliverable is to describe the methodology used to create a set of subjects-specific models with the aim to reproduce their biological diversity. Subsequently, a "vector of features" that identifies a specific real patient is used to personalize the subject specific models. The target of this deliverable is to have a library of virtual patients infected by M. Tuberculosis that can be used to test in silico the efficacy of new therapies, once a fully validated model of the disease progression, and of the treatment effect on such progression are available. Virtual in silico patient cohorts can be used in a number of ways, including the so-called silico-augmented clinical trials, where a Bayesian adaptive design combines the responses observed in physical patients with those predicted for virtual patients.

Introduction

Partner: UniCT, USFD.

UniCT partner has worked on its pre-existing modeling framework (UISS – Universal Immune System Simulator) in order to allow it to reproduce the immune system dynamics at a large scale. Pre-existing UISS is a computational framework that makes use of a multi-scale, multi-organ, agent-based model (ABM) simulator of the immune system.

UISS takes into account both cellular and molecular entities. Cellular entities can take up a state from a certain set of suitable states and their dynamics is realized by means of state changes. A state change takes place when a cell interacts with another cell or with a molecule or both of them. We considered the relevant lymphocytes, i.e. B lymphocytes, helper, cytotoxic and regulatory T lymphocytes and

natural killer cells (NK). Monocytes are represented as well and we take care of macrophages (M) and dendritic cells (DCs). For what concerns molecules, the model distinguishes between simple small molecules like interleukins (ILs) or signaling molecules in general and more complex molecules like immunoglobulins (Ig) and antigens (Ag), for which we need to represent the specificity.

At the same level of entities, UISS implements immune system activities. They include both interactions and functions. Functions refer to the main immune system tasks. 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 and immune memory.

UISS represents receptors and ligands as bit strings and use a string matching rule to model affinity. This clever idea was introduced by Farmer and Packard[1] as a way to perform calculations for determining molecular complementarity and predicting the optimal size of an epitope. From immunology, we know 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. Binding is modeled by a string matching rule 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 large-scale properties of the immune system. Character strings produced accurate models when benchmarked to experiment, suggesting that the abstraction captures important features of receptor/ligand binding.

[1] J. D. Farmer, N. H. Packard, and A. S. Perelson. The im- mune system, adaptation, and machine learning. Physica D, 22:187–204, 1986.

In particular, specificity is implemented in UISS by a bit-string polyclonal lattice method. Bit-string refers to the way the molecules and the specificity among molecules is represented, polyclonal indicates that more clones of different specificity of lymphocytes are represented and lattice means that we use a discrete lattice to represent the space, that is, the space is discrete. The set of lymphocytes receptors is represented by bit-strings of length "h" which then form the so called "shape space". A clonal set of cells is characterized by the same clonotypic receptor, i.e. by the same bit-string of length l. The potential repertoire of receptors scales as 2^l. The receptor–coreceptor binding among the entities are described in terms of matching between binary strings with fixed directional reading frame. Bit-strings represent the generic binding site between cells (through their receptors) and target molecules (through peptides and epitopes). An interaction between two entities is a complex action which eventually ends with a state change of one or both entities. Specific interactions need a recognition phase between the two entities; recognition is based on Hamming distance and affinity function, and is eventually enhanced by adjuvants. When two entities, which may interact, lie in the same lattice site then they interact with a probabilistic law. All entities which may interact and are in the same site have a positive interaction. Physical proximity is modeled through the concept of lattice-site. All interactions among cells and molecules take place within a lattice-site in a single time step, so that there is no correlation between entities residing on different sites at a fixed time. The simulation space is represented as a L x L lattice, with periodic boundary conditions to the left and right side, while the top and bottom are represented by rigid walls. All entities are allowed to move with uniform probability between neighbouring lattices in the grid with equal diffusion coefficient (Brownian motion).

UISS was extended to consider immune system dynamics at a large scale. This was achieved extending UISS platform in three critical points: i) the peripheral blood compartment was increased in size in order to deal with a representative portion of peripheral blood and included immune system entities circulating through it; ii) immune system repertoire implementation was extended in order to take into account an immune system potential diversity at human natural scale i.e. about 10^{20} order

magnitude of T and B cells clonotypes; iii) compartments addition to simulate critical organs targeted by tuberculosis i.e., lungs and near lymph nodes.

The human T-cell receptor (TCR) repertoire i.e., the range of different TCRs expressed ranges between 10^{15} to 10^{20} clonotypes by recombination, random insertion, deletion and substitution; the size of the antibody repertoire has been calculated as 10^{15}. UISS immunological repertoire uses binary strings of 12 bits, reaching a diversity of 2^{12}. In order to simulate the immune system at a human scale it is needed that UISS deals with a repertoire of 10^{20} i.e., having bit-strings of about 60 bits in length. To compute the binding probabilities of all interactions among cells and molecules, UISS uses Hamming distance computation i.e., the difference in terms of bits between two different bit strings. The computational effort required to compute the Hamming distance linearly depending on the size of the string. However, even a linear effort may be computational heavy if many computations are required. The best-known method to compute Hamming distance between two bit strings relies on a pre-computed look-up table kept in computer cache memory; it is computationally fast if the cache misses are rare. With this method, the computation of Hamming distance of binary strings of length l requires a look-up table of $O(2^{2l})$ entries. To manage bit strings of length 60 we developed a new algorithm that we implemented in UISS. The algorithm is based on magic numbers. Magic numbers in mathematics and computer science are numbers that show special properties when used in certain computations. In particular, they are successfully used in algorithms involving bit handling. They offer faster versions of algorithms over those that can be developed without their use, typically by a factor of n/logn. Our implementation, which is an extension of the well-known bit count algorithm, implements a branchless computation of the population count. It is based on a O(log(n)) algorithm that successively groups the bits into groups of 2, 4, 8, 16, 32, and 64 (using binary magic numbers) while maintaining a count of the set bits in each group. With this implementation, we were able to deal with binary strings of 64 bits. The method is easily extendable to any bit strings length, but keeping the limit of 12 elementary CPU operations, requires CPUs capable of doing l-bits operations in one time, l>64.

Finally, in UISS it has been set a new compartment i.e., the lung compartment. Lymph nodes compartment is already present and it is extended to take care also of the lymph nodes around lungs. Multi-organ ABM simulations, however, require thousands of millions of agents. To this end, high-performance computing (HPC) and Graphics Processing Units (GPUs) resources are mandatory to reproduce the large scale behavior of the immune system and related pathologies. Due to the intrinsic nature of the biological and immunological entities that mostly act and interact locally, the simulation of big tissues and/or organs has been split across different computing cores in order to have a parallel run 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 clearly entitles high degrees of scalability in function of the number of available computing cores. To this end, UniCT partner has interacted with USFD partner in order to set a FlameGPU software to enable large scale simulations on GPUs.

The UISS model in its entirely has been ported to the FLAME GPU framework. This includes the implementation of various agent-cell and cell-environment interactions. The model has been provided to UniCT via a GitHub repository which is used for tracking any reproducibility issues. Public documentation for the use of the code (including access to TUOS GPU facilities) is available at: https://flame-gpu-uiss.readthedocs.io/en/latest/.

To model the entire dynamics of MTB and interactions with the immune system machinery, we selected all the players that have a role in TB in order to include the disease model into the simulation framework. The starting point was the implementation of all the biological entities involved into the dynamics of MTB, both at cellular and molecular level. To analyze the behavior of a possible therapeutic intervention, we selected the two vaccines we are going to test inside the STriTuVaD project i.e., RUTI and ID93 vaccines. To this end, we implemented inside UISS the main mechanism of action (MoA) of these immunotherapeutic vaccines.

Our model takes into account both innate and adaptive immunity (both cellular and humoral) and immune memory. Figure 1 depicts all the entities implemented within the simulation framework,

especially immune cells, cytokines and specific biological processes along with their peculiar properties in TB dynamics. All the entities of the TB disease model interact each other, and are appropriately located in two specific compartments: the lung (pulmonary alveoli) and the peripheral lymph nodes.

The starting point of our TB disease model consists on the aerosol droplets of MTB that reach lung alveolar macrophages (AMs) on one side and neutrophils (N) on the other one. When an AM becomes infected, it secretes IL-1, TNF-alpha, 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 to the final outcome of the infection. In this context, lipoxin A4 (LXA4) promotes necrosis, while prostaglandin E2 (PGE2) is a proapoptotic factor. When 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. a) In the first-case scenario, AM apoptotic can interact with a lung resting macrophage (M) and lead to the 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. b) In the second scenario, AM apoptotic cells can encounter a lung DC. AM apoptotic can be taken up by DC that capture antigen 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-2 release and other chemokines, except when IL-10 is present and is able to block this moving. c) The last scenario considers 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. Just like AM, also for N the MTB strain can lead to a different role in the induction of either apoptosis or necrosis and to 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, with reduction of bacterial killing and impairing of 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, activated Th1 cell population, produce and secrete IFN-γ, causing macrophage activation, relative cytokine production (IL-12 and TNF-alpha) and bacterial control. It is worth to mention that in this context, IL-10 can block macrophage activation and consequent cytokine secretion. Also, TReg negatively modulate Th1 population effector functions.

For the sake of completeness, the Th1 cell population interacts also with B cells leading to three specific processes at the same time: after a successful interaction, B cells duplicate, 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 secretion of immunoglobulins type A (IgA). For what concern Th17 cell population, their migration and interaction with B cells will finally lead to cell duplication and secretion of immunoglobulins type E (IgE). Inside the lymph node, B cells interact also with MTB; after that B cells become active and secrete immunoglobulins type M (IgM).

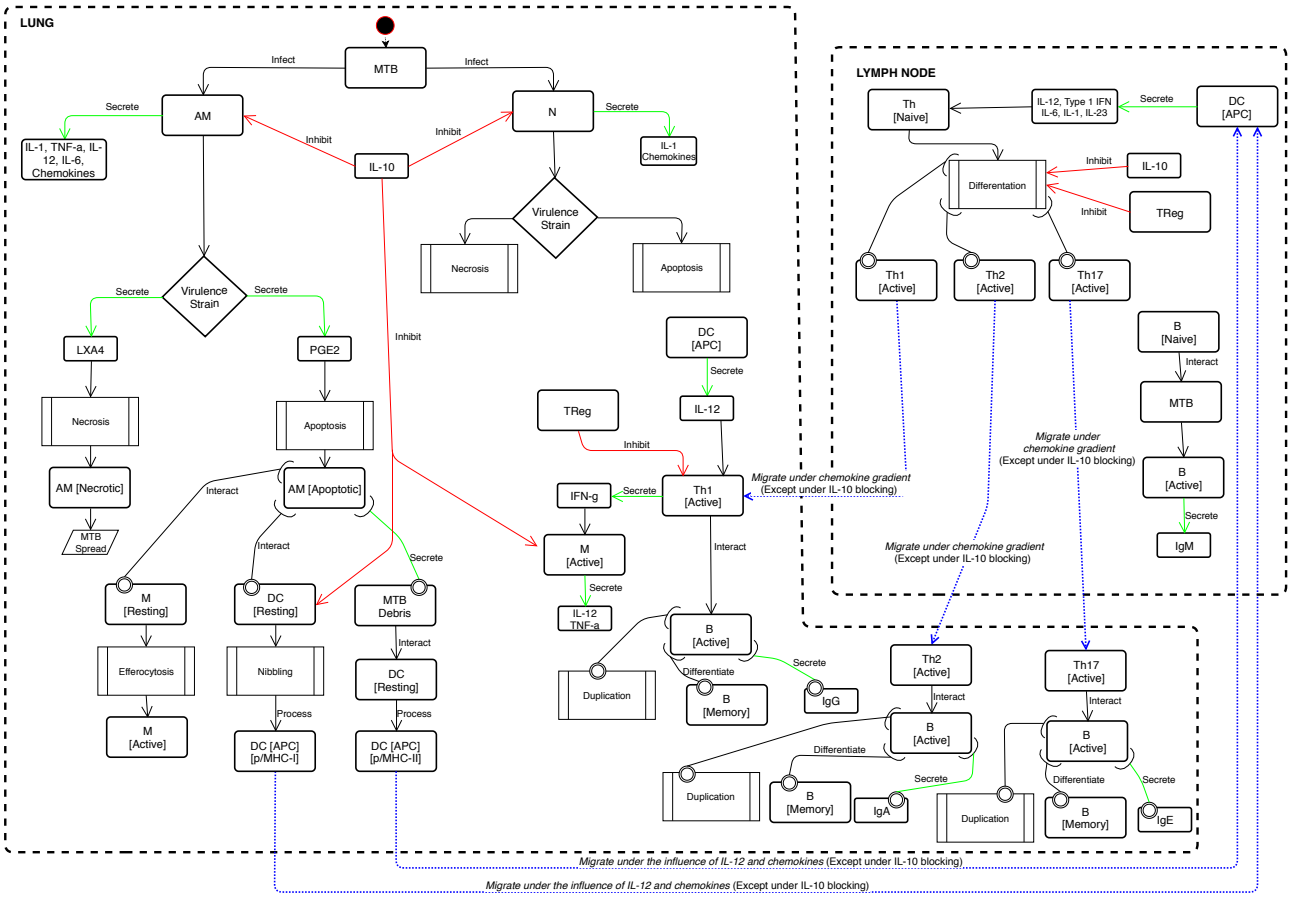

Figure 1. Conceptual description of the main entities and interactions of UISS-TB. The main two compartments are represented: the lung and the peripheral lymph nodes. The representation depicts both cellular and humoral response when MTB droplets infect alveolar macrophages resident in the lung. The cascade of cytokines and chemokines is also represented with possible different behaviors depending on the virulence of the MTB strain.

Subsequently, UISS was extended to implement the needed features to allow UISS modeling framework to simulate the artificial immunity induced by two vaccines i.e., RUTI vaccine and ID93 vaccine, provided by AF partner and IDRI partner, respectively. To do so, a detailed knowledge about the components that are used in RUTI and ID93 vaccines was required. Vaccines use several mechanisms to be able to induce a proper immune response and each of them has been added to the simulation platform. Antigen structure, liposome recognition by the immune system, adjuvants stimulation of the innate immunity and the delivery dynamics of each vaccine have been implemented

accordingly. Moreover, as the clinical trial we set in the project for the validation of the in silico clinical trial will make use of antibiotic strategy based on isoniazid, we further extended the UISS framework to take into account the M. tuberculosis – isoniazid interaction. Both bactericidal and bacteriostatic mechanisms were implemented in the UISS model. Bactericidal effect has been implemented getting from the literature the mechanism of action as well as the pharmacodynamics and pharmacokinetics of the antibiotics. Bacteriostatic effect has been implemented decreasing the rate of duplication of bacteria agents while in presence of antibiotics concentration. The model is able to distinguish the rapidly dividing mycobacteria from the slow one as UISS follows the single entity dynamics.

To validate what we implemented into the UISS computational framework, we started to evaluate some scenarios that could arise during the MTB – immune system – vaccines interactions and antibiotics. We analysed three simulation scenarios. The first one is represented by the establishment of MTB latent chronic infection with some typical granuloma formation as a reservoir of MTB infection. The second one deals with a reactivation phase during latent chronic infection where a possible breakdown of the granuloma may lead to the spread of the bacilli and to the reactivation of the disease, with an increased necrotic burden. The third consists on the latent chronic disease infection scenario during RUTI or ID93 vaccine administration: here we show the evolution in time of M and CD4 T cell populations, in conjunction with the temporal evolution of IFN-γ that owns an important role as a biomarker of protection. D2.3 shows the result obtained in these simulations.

To reproduce biological diversity of the subjects that have to be simulated, an appropriate strategy for the generation of libraries of virtual in silico patients has to be developed. Here, this goal has been obtained through the combination of three different approaches. The first is the creation of the initial immune system repertoire, generated in a way that simulates the DNA fragments assembly accounting for the inherent stochasticity of the process, but also on the presence of the Class I or Class II HLA patterns.

The second is related to the simulation of immune system interactions. These take place in a different order, that we simulated accordingly. This imitates, to some extent, the randomness of the immune response in the initial phases, where innate immunity takes place (consider, for example, the case that a fully matching CD4 T cell is present at the specific lymph node where a DC is presenting the processed antigen). The last approach is represented by the identification of a "vector of features" that combines both biological and pathophysiological parameters that personalize the virtual in silico patient to reproduce the physiology and the pathophysiology of the subject. In particular, the virtual in silico patient model defines a specific patient through a vector of 26 features: Drug Sensitive (DS)/Multi-drug resistant (MDR); Bacteria Load (BL) in sputum; MTB strain; CD4 Th1; CD4 Th2; IgG titers; CD8 T cells; IL-1; IL-2; IL-10; IL-12; IL-17; IL-23; IFN Type I: IFNγ; TNFα; TGFβ; LXA4; PGE2; Chemokines; Vitamin D; HLA-1; HLA-2; FoxP3; Age; BMI.

In the next section, we describe in detail the methodology we developed to generate the library of virtual in silico patients.

Methodology

The UISS implementation targeting tuberculosis (here referred to as UISS-TB) requires 26 different information on each individual patent to predict the progression of the disease in that specific patient, and how such progression would change when that specific patient receives a given treatment. In this perspective, a patient is defined in this regard by 26 specific values, and thus any set of such 26 specific values defines a patient, physical or virtual. In physical patients, the 26 values are measured/observed, while in virtual patients the plausible sets can be generated if a detailed description of the values, ranges, and correlation structure of these data are present in physical cohorts. If we have such information, we can develop procedure that generate cohorts of virtual patients that matches the characteristic of physical patients' populations. We propose a sequential procedure for simulating a virtual cohort consistent with some predetermined population characteristics.

In order to create a virtual in silico patient, one needs to provide a single value for each one of the 26 features. These values can be taken from individual physical patients, fed to UISS-TB and use the output for validation; i.e. assess the prediction accuracy and precision of the simulator predictions, against the actual observed outcome.

However, if a cohort of virtual patients is to be produced, we should have a mechanism for producing as many different input vectors as needed that are biological/physiological plausible. Formally, this requires the characterisation of the joint distribution of the inputs in the population. To this end, partner UniCT compiled for each input the range of values observed in physical cohorts, as well as average and standard deviation of such values. If we assume each input is normally distributed over the physical population, we could generate virtual patients by sampling the distribution of each value. However, this approach would neglect the biological correlations between inputs and thus would not guarantee a physiologically plausible input vector. Hence, we must take into account these correlations. Given that we have 25 numerical input variables (DS/MDR is a factor), we should specify $25 \times 24 / 2 = 300$ correlations. An experts' panel at partner UNICT and USFD qualified these correlations. It was found that all correlations are positive, except for TReg and IL-10, which correlate negatively with the rest of the features.

In theory, one could elicit the joint distribution of the 25 features, i.e. describe mathematically how each feature is related to each other in a space of 25 dimensions; but this would be not only extremely difficult, but also time consuming and data demanding. Our approach is to rely on current mathematical biology consensus and use a Gaussian to represent the population distribution.

To systematically define the joint distribution of the 25 features we would need to have a huge number of independent experimental observations: assuming we need between 100 and 1000 observations for correlation, this would mean to have a retrospective clinical study where all 25 features were measured in 30,000-300,000 patients. Since such a huge retrospective study does not exist, we need to explore alternative approaches. One such approaches rely on the assumption that the values of all 25 features over the patients' population are assumed to be normally distributed. While this

assumption is not true in general, for the 25 features of interest here all the available retrospective data reported in the literature[2,3,4,5,6,7] confirm the validity of this assumption; thus, we will assume it true and develop the virtual cohort relying on it Moreover, we also got preliminary data from our clinical partners i.e., IDRI and Archivel Farma and AIIMS that provided us a selection of values coming from a set of patients affected by pulmonary TB. As the prospective clinical trials planned in the project produce their observations, it will be necessary to confirm on those data the validity of this assumption.

Table 1 depicts the ranges, for every singular feature, we extrapolated from the retrospective data cited above.

[2] Saini D, Hopkins GW, Seay SA, et al. Ultra-low dose of Mycobacterium tuberculosis aerosol creates partial infection in mice. Tuberculosis (Edinb). 2012;92(2):160–165. doi:10.1016/j.tube.2011.11.007

[3] Gabriele Pedruzzi, Phonindra Nath Das, Kanury Vs Rao, Samrat Chatterjee. Understanding PGE2, LXA4 and LTB4 balance during Mycobacterium tuberculosis infection through mathematical model. J Theor Biol. 2016 Jan 21; 389: 159–170. doi: 10.1016/j.jtbi.2015.10.025

[4] Abbas A, Lichtman AH, Pillai S. Cellular and molecular immunology 9th Edition, Philadelphia, PA : Elsevier, 2018

[5] Stenken JA, Poschenrieder AJ. Bioanalytical chemistry of cytokines--a review. Anal Chim Acta. 2015;853:95–115. doi:10.1016/j.aca.2014.10.009

[6] Hannan NJ, Bambang K, Kaitu'u-Lino TJ, Konje JC, Tong S. A bioplex analysis of cytokines and chemokines in first trimester maternal plasma to screen for predictors of miscarriage. PLoS One. 2014;9(4):e93320. Published 2014 Apr 3. doi:10.1371/journal.pone.0093320

[7] Mayer-Barber KD, Andrade BB, Oland SD, et al. Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. Nature. 2014;511(7507):99–103. doi:10.1038/nature13489

Model Input Description	Parameter name	Type	UISS Unit of Measurement	Continuos (C) or Discrete (D)?	If C, range of possibile values:	If D, possibile values:
Bacterial load present in the	NumAg	Real	CFU	D		[50-400]
MTB virulence	strain	Real	Adimensional	C	[0-1]	
CD4 T cell type 1	Th1	Real	cells per microL	C	mean count of CD4 is about 1370 cells per microL	
CD4 T cell type 2	Th2	Real	cells per microL	C	mean count of CD4 is about 1370 cells per microL	
Specific IgG titers	IgG	Real	IgG titer	C	[2-8]	
CD8 T cell	TC	Real	cells per microL	C	mean count of CD8 is about 560 cells per microL	
Interleukin 1(*)	IL1	Real	pg/mL	C	[0-10000]	
Interleukin 2(*)	IL2	Real	pg/mL	C	[50-1000]	
Interleukin 10(*)	IL10	Real	pg/mL	C	[5-16]	
Interleukin 12(*)	IL12	Real	pg/mL	C	[3-300]	
Interleukin 17(*)	IL17	Real	pg/mL	C	[0-1000]	
Interleukin 23(*)	IL23	Real	pg/mL	C	[0-1000]	
Type 1 IFN(*)	IFN1	Real	pg/mL	C	[0-11000]	
IFN-γ(*)	IFNG	Real	pg/mL	C	[6-19]	
TNF-α(*)	TNF	Real	pg/mL	C	[4-40]	
TGF-β(*)	TGFB	Real	ng/mL	C	[2-8]	
LXA4	LXA4	Real	ng/mL	C	[0-3]	
PGE2	PGE2	Real	ng/mL	C	[0-2.2]	
General chemokine(*)	ChemGQTY	Real	ng/ml	C	[0-20]	
Vitamin D	VitaminD	Real	ng/ml	C	[0-100]	
HLA-class 1	MHC1	Natural	Adimensional	D		[0-2^NBITSTR-1]
HLA-class 2	MHC2	Natural	Adimensional	D		[0-2^NBITSTR-1]
FoxP3	Treg	Real	cells per microL	C	mean count of FoxP3 is 60 cells per microL	
Age	Age	Natural	Adimensional	D		[0-90]
Body Mass Index	BMI	Real	Kg/m^2	C	[16-41]	

Table 1. Vector of features for virtual cohort generation.

The additional advantage of using this approach will be discussed in the next section. Formally, we say that the vector $x = \{x_1, ..., x_d\}$ follows a d-variate Gaussian distribution if their joint probability density function (pdf) can be written as

$$N_d(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

where $\mu = \{\mu_1, ..., \mu_d\}$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_d^2 \end{pmatrix}$$

is a symmetric matrix, i.e. $\sigma_{ij} = \sigma_{ji}$. These are the covariance between entries of x,

$$\mathrm{Cov}(x_i,\, x_j) = \sigma_{ij}$$

which are related to the correlations by

$$\mathrm{Cor}(x_i,\, x_j) = \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$$

So, if we are able to elicit a measure of correlation between two inputs, we can calculate their covariance.

The elements in the diagonal, σ_i^2 are the marginal variances of each element, x_i, and μ_i the corresponding marginal mean. As mentioned above, UniCT has already compiled a list with these values, so we have elicited values for μ and the diagonal elements of Σ, σ_i^2. With this information, and with the knowledge that all correlations are positive, except those for IL-10, it is possible to build a first estimate of the covariance matrix Σ, and the first virtual cohort can thus be generated.

Meanwhile we have involved the whole consortium in the compilation of a systematic review of the relevant literature, in order to refine covariance matrix Σ with additional correlation information between any two of the 25 features, as observed experimentally in retrospective studies. We expect this refined estimate of the covariance matrix Σ to be available in the next six months. As the results of our own clinical trials are available, we will critically revise the results of this systematic review, and eventually generate a third update of the covariance matrix Σ later on in the project.

Once μ and Σ have been elicited, generating an in silico profile is a relatively trivial task: one must sample a point in the 25-dimensional space, consistent with $N_d(x \mid \mu, \Sigma)$. But we can exploit the properties of the Gaussian distribution to produce a cohort consistent with some specific characteristics. Say, for instance, that our target population has a particular range of bacterial load, we would like then to produce virtual patients consistent with that specific profile. Formally, let x_1

represent the bacterial load and $x_{-1} = \{x_2, \ldots, x_{18}\}$, the rest of the features; we would like to sample from

$$\mathrm{N}_d(\boldsymbol{x}_{-1} \mid x_1, \boldsymbol{\mu}, \Sigma)$$

i.e. the conditional distribution of the rest of the features, given that bacterial load has a specific value. This is a standard procedure described formally in Appendix A. We can go even further and sort the list of features according to either their importance in determining the profile of a patient, or to the precision of their elicited mean, variance and covariance, and then proceed to sample from the conditional distributions, one at a time.

Achievements

According to the methodology described above to create a virtual in silico patient, we generated a first sample of 100 subjects-specific models with the aim to reproduce biological diversity. Table 2 shows a sample of 30 patients with the values of every single features.

NumAg	strain	Th1	Th2	IgG	TC	IL1	IL2	IL10	IL12	IL17	IL23	IFN1	IFNG	TNF	TGFB	LXA4	PGE2	ChemGQTY	VitaminD	MHC1	MHC2	Treg	Age	BMI
189	0,7064	43,0026	46,0171	4,5272	49,9449	5061,4484	555,2564	14,3961	166,0229	534,2345	516,9053	5534,1701	21,8631	13,0174	7,5336	2,9584	0,8896	9,147	47,433	2029	2077	56,5996	39	30,7239
249	0,4037	48,9355	51,9734	4,9051	51,5474	5074,8402	532,7486	14,9519	186,1316	515,3413	510,784	5443,0399	18,6891	17,8917	3,8182	2,6132	0,8228	11,8772	55,1513	2047	2061	56,1998	46	32,5742
187	0,5345	62,6795	52,7117	7,09	51,0602	4993,2883	514,3206	15,5863	158,926	510,4008	522,2833	5499,6464	20,1607	19,4982	6,8278	2,9206	0,748	11,7833	51,7372	2067	2036	51,0656	47	32,3395
253	0,1497	54,0043	61,9718	4,7221	43,8951	4986,7165	515,617	18,742	138,3579	524,738	512,9723	5445,4381	20,2626	10,8367	8,1019	1,9802	0,3019	7,0138	50,0939	2021	2014	47,2758	47	28,5157
175	0,2359	51,3442	43,3676	4,7365	51,1459	4925,4342	509,9842	18,8978	140,1573	500,5703	475,8469	5514,8748	20,2905	11,7617	5,659	2,4525	0,4114	12,1593	43,4858	2077	2047	50,3442	47	27,8206
229	0,8574	55,9904	42,5917	6,6368	48,2675	4954,0164	538,3406	17,4829	151,191	495,0603	484,9285	5484,8903	21,5391	13,9083	3,022	2,2785	0,3367	8,6198	49,4228	2072	2044	45,7794	47	27,2923
221	0,7263	50,7691	54,4967	4,9214	45,5858	4975,01	507,1036	16,1741	154,1892	512,229	485,329	5477,3615	18,6052	15,795	5,4948	2,4069	0,35	8,4031	48,8784	2034	1990	42,4486	49	29,2917
199	0,2529	46,6782	55,5033	3,667	53,9831	5019,5527	515,4431	17,3953	150,8441	502,38	460,1916	5458,0469	19,4956	9,6021	5,4425	2,3738	1,0689	10,5309	44,9965	2069	2029	46,2823	36	30,0874
158	0,341	43,7295	53,0081	4,051	59,4708	5007,7425	542,1148	20,7871	145,0323	498,4098	479,92	5460,3267	20,7865	12,383	6,7701	2,679	0,7108	9,7491	51,5819	2066	1992	51,981	42	29,6103
191	0,2069	40,1393	44,7984	3,6402	43,0434	4981,7051	519,9058	18,0114	138,6341	494,8496	503,2272	5395,4918	21,5357	7,9219	5,7284	1,9981	0,6403	8,3582	41,329	2024	2076	37,2982	41	26,8263
328	0,5697	47,7406	39,8609	4,3202	50,443	5018,1344	513,2773	17,841	144,1758	501,3239	505,7501	5519,1226	20,2162	12,055	5,8717	2,243	0,5254	7,8973	58,2549	2033	2065	40,006	46	25,8079
252	0,5409	34,5168	52,5371	5,44	42,9358	4973,132	496,5298	18,6227	143,7459	478,8328	500,9754	5493,0094	21,7386	12,0786	6,1195	2,0911	0,9006	8,2634	52,6502	1991	2041	49,1065	38	29,2932
261	0,496	53,841	55,7516	6,1979	56,1758	4946,8553	525,3766	17,8775	153,0256	502,4235	502,62	5587,8507	22,6664	13,0775	4,1341	3,2212	1,0254	9,729	59,3885	2077	2109	51,3286	55	29,8035
273	0,8468	51,7831	49,0332	7,3126	54,4582	5035,4812	550,429	16,6697	173,3459	505,559	504,4409	5510,072	21,5124	13,4563	6,7832	3,0472	1,403	12,7089	51,7225	2103	2107	58,9217	54	32,994
300	0,7984	42,2896	46,7656	6,0218	50,2861	4956,6442	500,785	16,6408	146,5688	508,8302	507,019	5487,8152	21,7575	12,2893	1,7823	3,0895	1,0844	11,6959	46,7568	2036	2037	54,5267	44	25,0303
245	0,3409	49,0301	43,4797	5,7352	50,107	4916,7829	522,1456	21,0991	134,5484	481,5512	497,4683	5430,4331	20,2877	15,0756	6,078	2,9428	1,0292	8,7435	57,8596	2031	2057	37,5741	42	31,7647
221	0,6765	52,9173	50,3888	6,7486	44,5091	4958,1908	484,4909	18,4464	153,8292	487,7736	476,6785	5491,8557	21,4735	13,6427	3,7297	2,3341	1,0161	8,9259	53,0422	2059	2032	50,8327	36	29,8958
122	0,8234	54,2139	44,9696	6,1404	51,3831	4911,5694	520,6547	17,554	163,8121	499,584	504,6211	5599,9925	22,115	14,7733	2,5084	2,2384	0,987	9,9874	44,9932	2039	2064	53,4504	38	27,2457
242	0,817	42,4162	50,9999	4,2083	50,9491	4983,226	527,727	17,3232	147,3087	494,8498	491,6612	5440,641	19,3185	12,3385	5,1726	2,6179	0,7806	9,303	53,4159	2041	2088	47,6792	47	30,6008
313	0,5595	42,5075	51,0508	6,1914	55,2109	5001,2331	514,2203	17,1889	155,1386	499,223	488,7785	5537,1736	20,3794	12,8552	3,1379	2,2685	0,3304	5,612	57,8411	2071	1992	47,4095	42	27,054
179	0,0263	42,2299	43,788	6,788	43,2682	5070,4856	520,669	21,9455	170,1706	525,4997	525,1979	5458,1266	19,6801	13,2225	2,6282	2,1743	1,1125	11,0477	45,1721	2020	2020	50,0213	36	32,62
345	0,3713	52,4289	44,0263	6,2036	47,7616	4975,009	541,1243	17,4899	155,8814	507,2089	502,6453	5467,4033	21,1679	13,2919	6,8168	2,7638	0,8652	14,8009	49,6341	2024	2112	55,1878	56	34,2884
245	0,6411	50,7486	52,029	5,8206	56,801	4992,5917	551,0074	15,562	173,0936	518,2232	514,6084	5569,5054	22,255	13,2625	8,3062	2,647	1,1576	12,9996	51,3303	2064	2059	54,2546	49	29,0993
251	0,5811	50,9855	50,9476	2,992	49,47	4988,3918	521,9707	20,6978	127,7217	489,5604	484,172	5438,5629	18,8945	12,8837	3,9893	2,4579	0,1879	5,2656	53,3541	2054	2069	41,7856	38	24,8173
234	0,2244	46,0625	41,5706	4,4013	56,7555	5003,6886	551,1547	20,7158	154,7522	501,4355	519,7695	5515,5011	20,754	14,2293	4,1069	2,1971	0,4892	10,2492	44,4583	2021	2017	44,8106	44	26,9518
186	0,4632	50,4897	46,9057	4,3641	46,9877	4869,5685	551,4614	20,0215	138,6745	496,335	500,3965	5476,0064	20,6685	12,4258	4,4918	3,1004	0,7385	6,991	47,4948	2080	2045	55,5919	45	30,598
154	0,3493	53,8546	49,2276	6,0139	44,4076	5014,575	515,4972	18,2619	152,165	502,5328	482,2956	5576,3162	20,5513	13,9025	5,666	2,6836	0,99	10,3478	56,3187	2011	2028	53,8609	41	27,1781
239	0,7162	57,5228	56,2092	6,293	47,7395	4961,3165	548,3835	17,8887	141,1583	520,8292	487,1472	5519,7006	20,3833	12,2471	3,2317	2,4364	1,1666	12,2289	45,6542	2057	2048	53,5762	45	28,0621
208	0,498	45,8048	54,4726	4,6544	52,1506	5064,606	539,2043	19,0403	160,0843	494,0983	520,2617	5638,1364	21,8383	14,0548	4,0847	2,8577	0,9645	6,3725	56,9065	2062	2026	44,9448	51	26,8387
209	0,5812	47,3296	56,9416	3,7732	52,5191	5118,0405	530,2351	17,2818	149,2191	501,746	492,4496	5513,7595	21,816	11,7202	7,0608	2,6921	0,4047	8,7601	47,4354	2043	2043	50,9277	44	32,8071

Table 2. A subset of 30 virtual in silico patients. Each virtual in silico patient is characterized by its own vector of features.

In the following figures we show the typical UISS-TB simulation framework when applied to the sample set of virtual in silico patients depicted in table 2. We show, for each biological entity, both the mean behavior (according to the entities the color line may vary) and the +/- SD (blue lines). We run a total of 90 simulations i.e., 30 for untreated virtual in silico patients, 30 for virtual in silico patients treated with RUTI vaccine and 30 for virtual in silico patients treated with ID93 vaccine.

In figure 2, it is depicted the dynamics of alveolar macrophages during two phases. The first one is during the active TB phase where both necrotic and MTB-infected populations increase. After the active phase, a latent phase is established and necrotic alveolar macrophages contribute to typical granuloma formation.

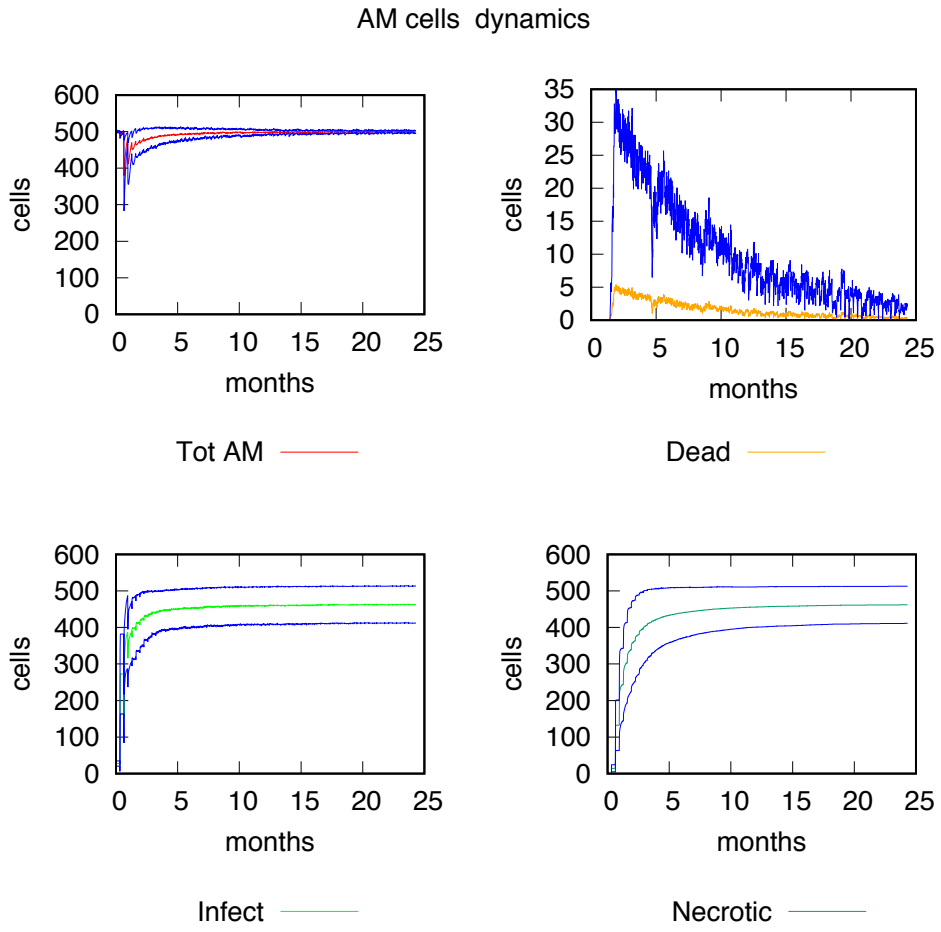

Figure 2. Alveolar macrophages population detailed dynamics.

Figure 3 shows the typical cellular response of CD4 T cells mounted against MTB infection. On average, CD4 Th type 17 population is the prominent response even if other CD4 subpopulations contribute to keep under control the infection.

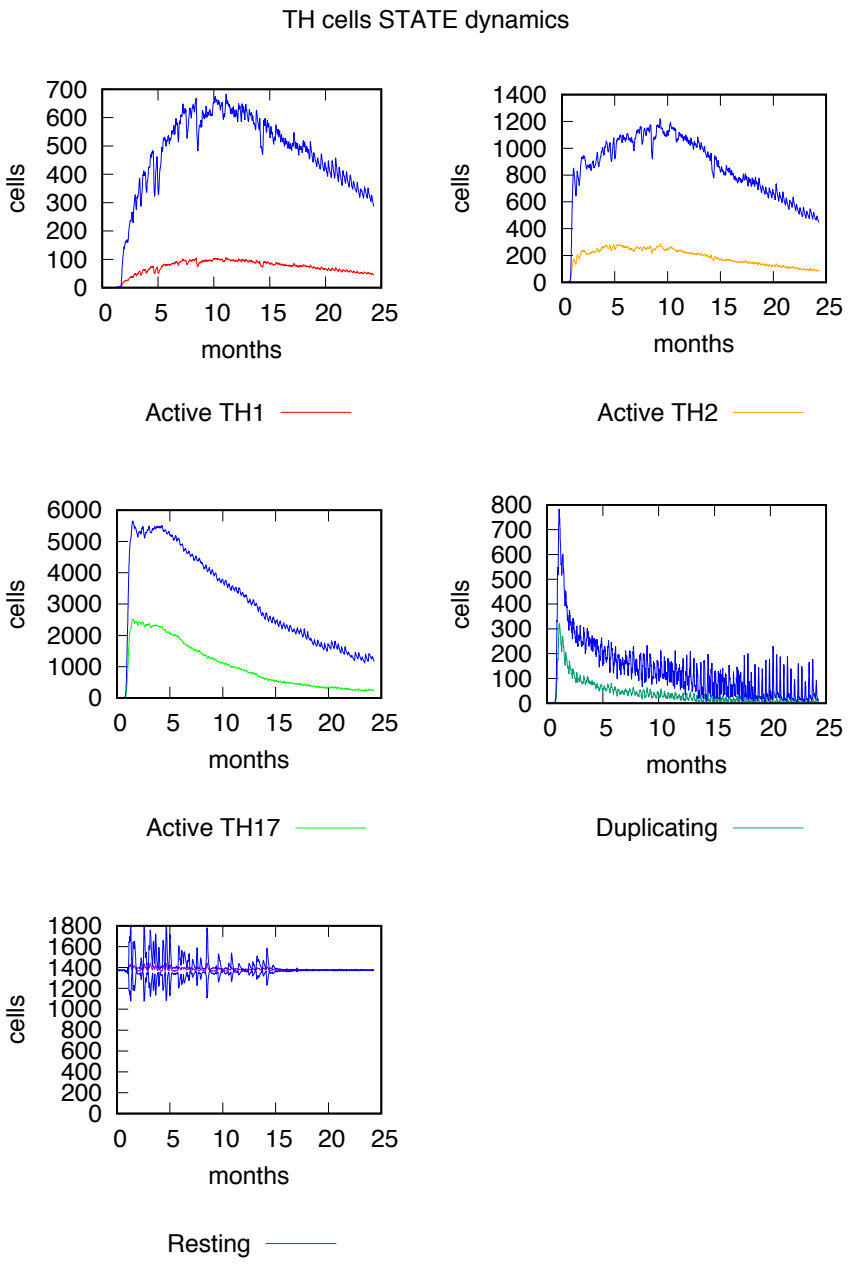

Figure 3. CD4 T cells detailed dynamics.

Figure 4 shows the dynamics of free MTB bacilli during the active phase, where they duplicate rapidly and infect target cells i.e., alveolar macrophages. At the same time, immune system cellular and humoral response are mounted accordingly. In particular figure 4 highlights the dynamics of specific IgG anti-MTB. Then, at a later phase, a typical low quantity of MTB bacilli shows the exhibition of latent infection.

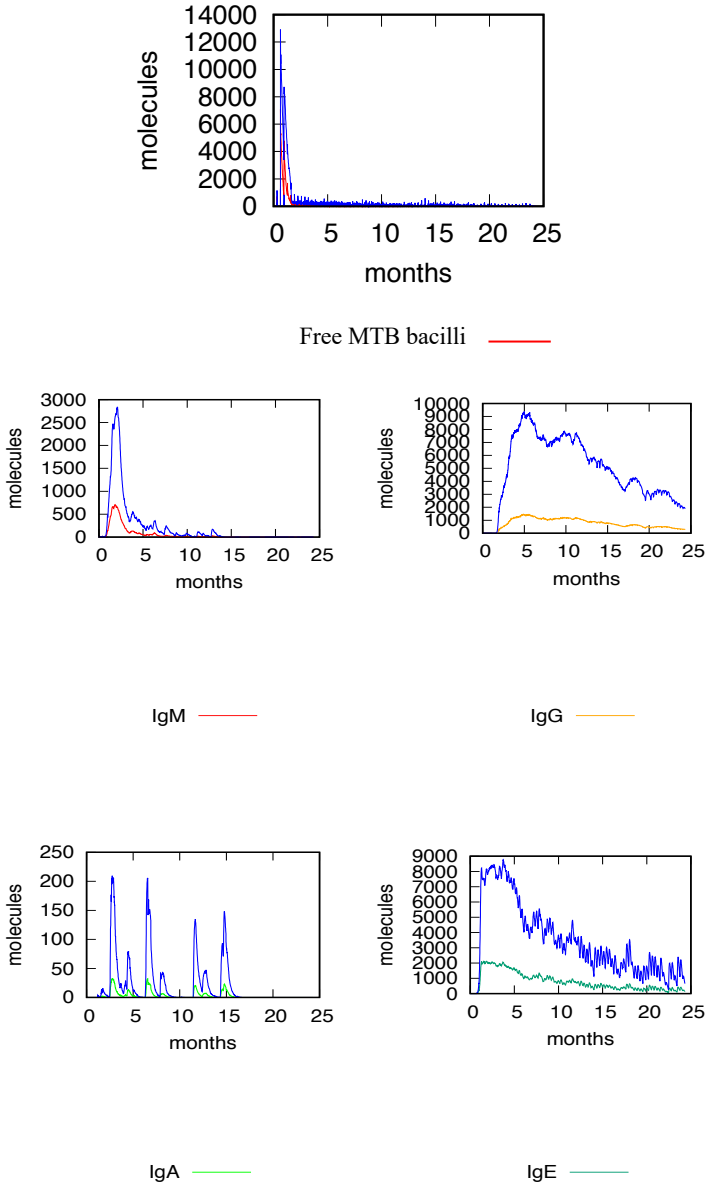

Figure 4. MTB bacilli and Ig classes dynamics

Figure 5 depicts two moments of typical granuloma formation during TB infection. One can observe the establishment of MTB latent chronic infection with some typical granuloma formation as a reservoir of MTB infection.

Figure 5. Typical granuloma formation during latent TB infection.

In the following figures (6-9), the average effects of the vaccination based on RUTI vaccine in a set of 30 virtual in silico patient with latent TB infection are shown. Figure 6 shows the alveolar macrophage dynamics after RUTI administration accordingly to the approved protocol. We can observe that the average necrotic AM population is considerably reduced indicating an effective immune response elicited by this vaccination strategy, decreasing the probability of disease reactivation.

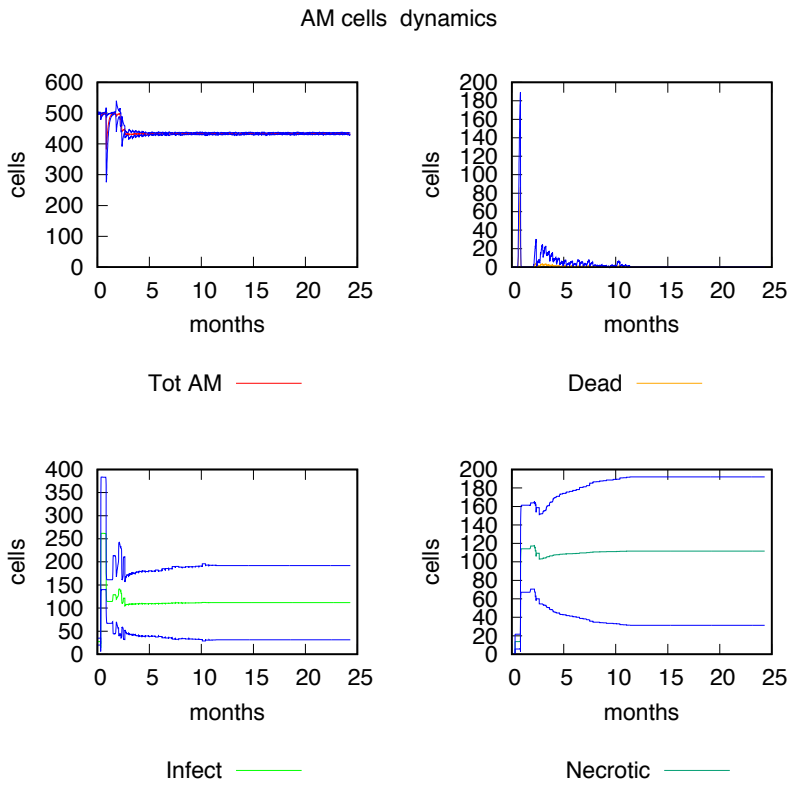

Figure 6. AM population detailed dynamics with RUTI vaccine administration.

Then, in figure 7, a strong Th1 response is induced with a down-regulation of Th2 response, with the induction of immunological memory, while in figure 8 high levels of IFN-γ are present, in good agreement with the results presented in specific literature.

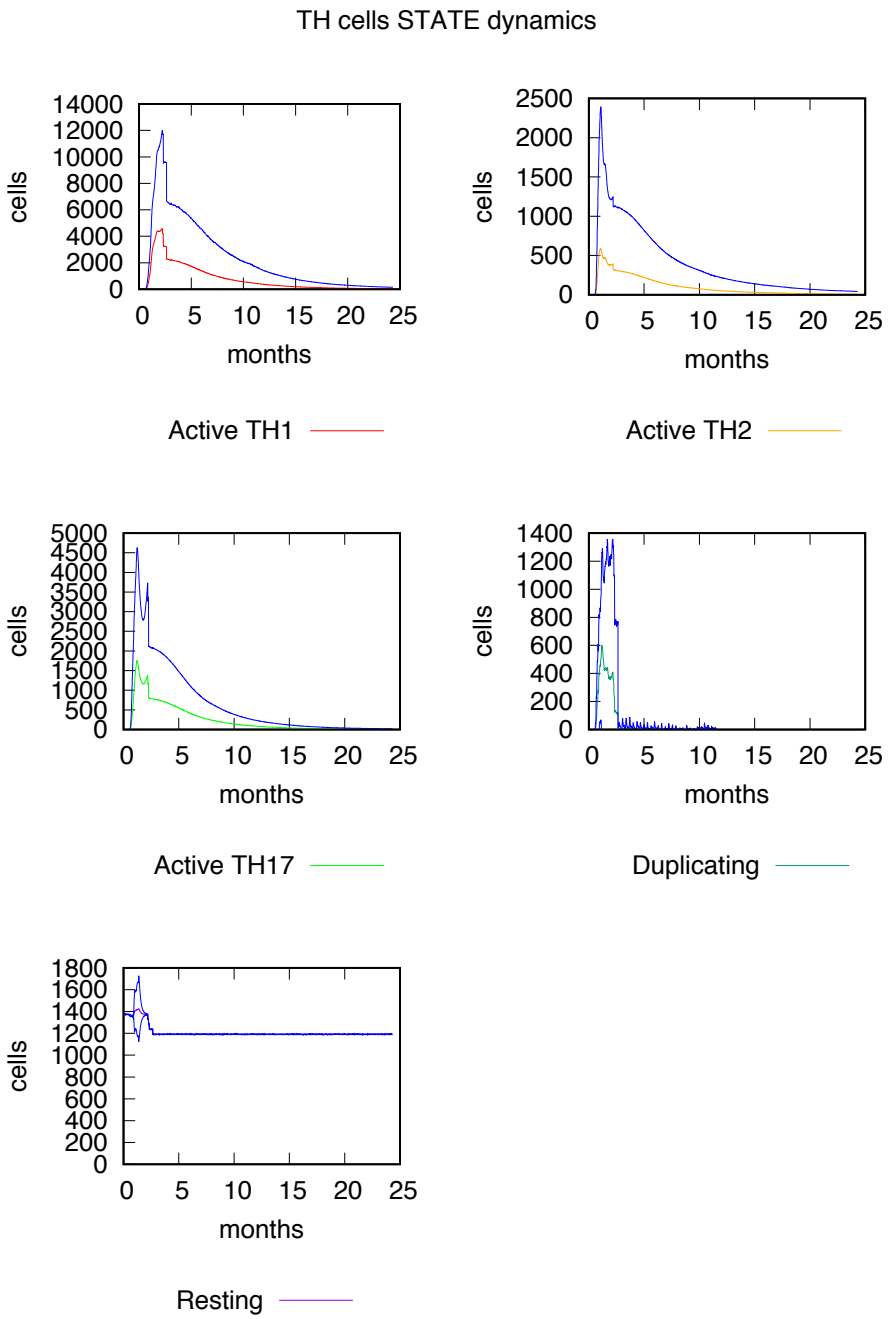

Figure 7. CD4 T cell population detailed dynamics.

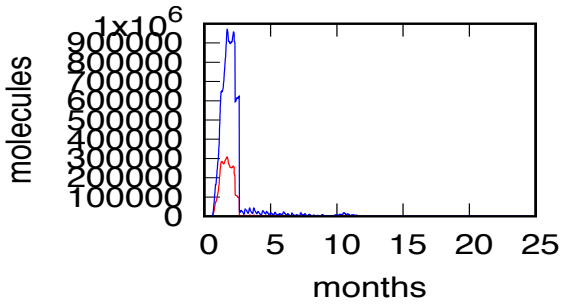

Figure 8. IFN-gamma levels

Finally, in figure 9 Lipoxin A4 (LXA4) and prostaglandin E2 (PGE2) levels are depicted. LXA4 promotes necrosis, while PGE2 is a proapoptotic factor. When necrosis process is favored, the AM becomes necrotic and contributes to the MTB spread. UISS-TB is able to reproduce this behavior as pointed out by figure 6. As one can see, after RUTI vaccine administration, LXA4 level decreases favoring a pro-apoptotic behavior, and consequently reducing necrotic alveolar macrophages.

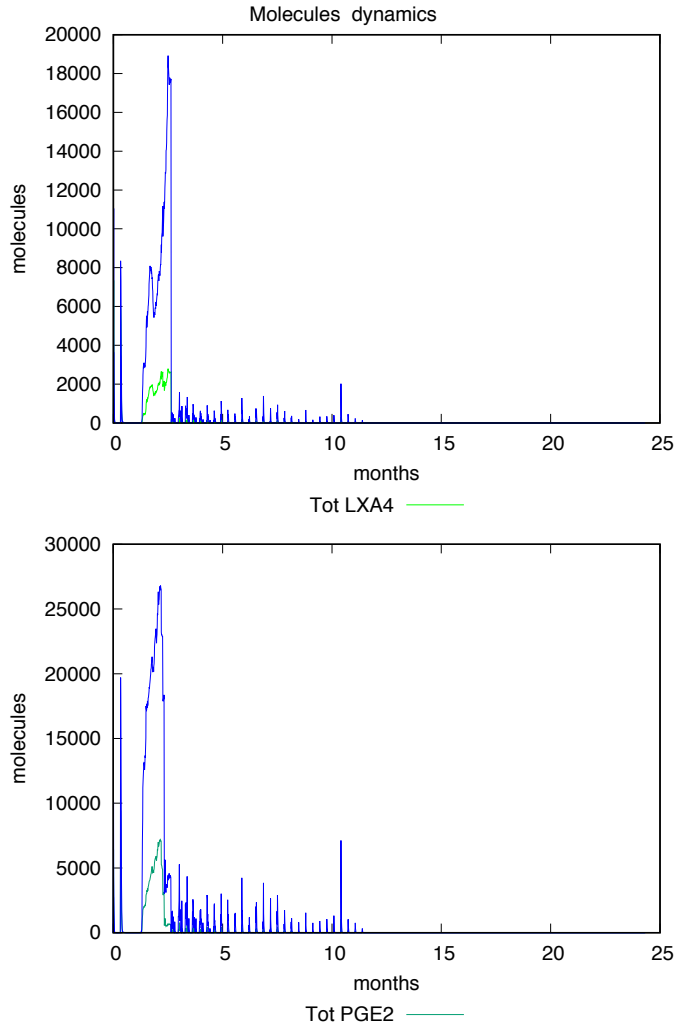

Figure 9. LXA4 and PGE2 levels.

In the next figures (10-13), the average effects of the vaccination based on ID93 vaccine in a set of 30 virtual in silico patient with latent TB infection are shown. Figure 10 shows the alveolar macrophage dynamics after ID93 administration. We can observe that the average necrotic AM population is considerably reduced indicating an effective immune response elicited by this vaccination strategy, decreasing the probability of disease reactivation.

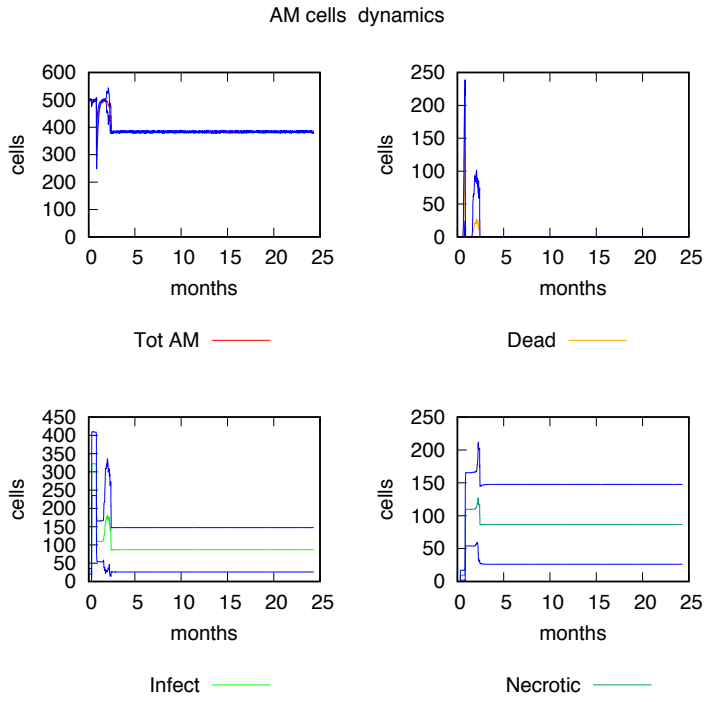

Figure 10. AM population detailed dynamics with ID93 vaccine administration.

Then, in figure 11, a strong Th1 response is induced with a down-regulation of Th2 response, with the induction of immunological memory, while in figure 12 high levels of IFN-γ are present, in good agreement with the results presented in specific literature.

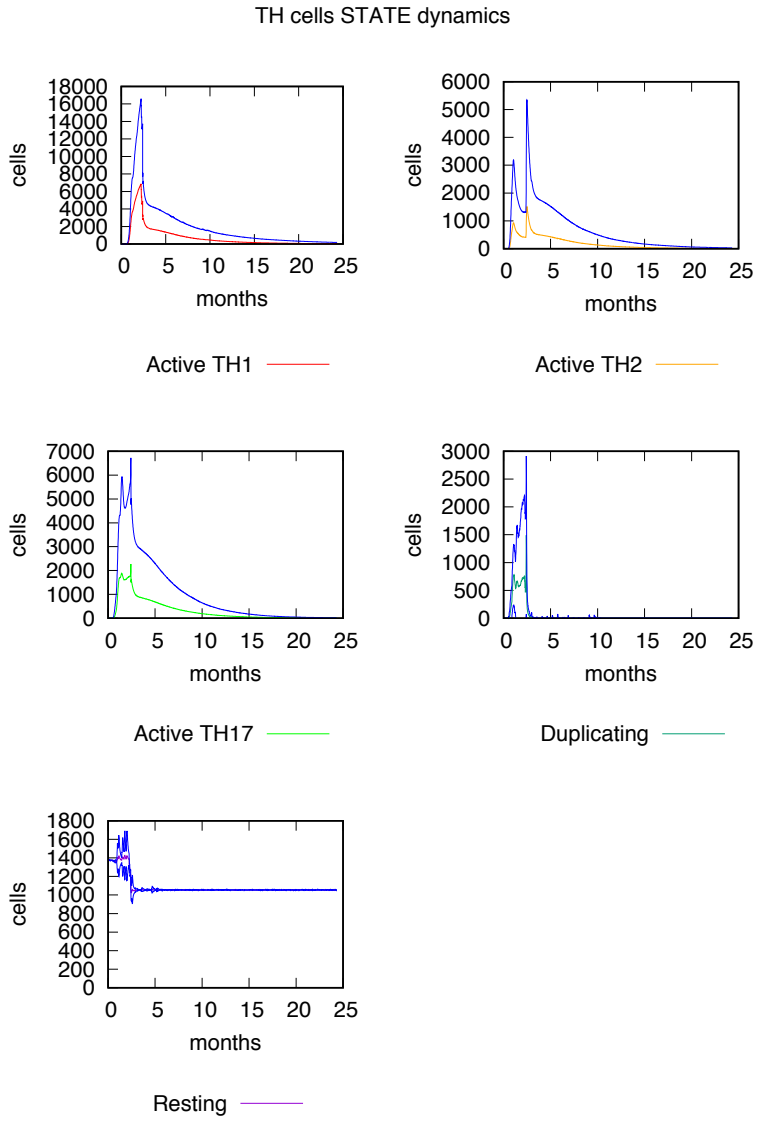

Figure 11. CD4 T cell population detailed dynamics.

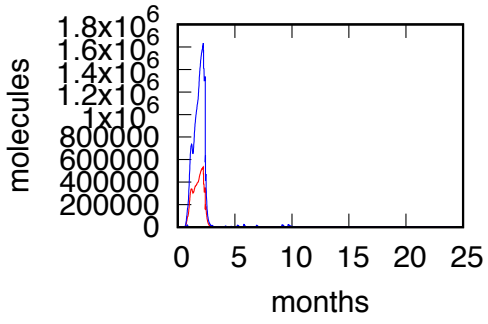

Figure 12. IFN-gamma levels

Finally, in figure 13 Lipoxin A4 (LXA4) and prostaglandin E2 (PGE2) levels are depicted. LXA4 promotes necrosis, while PGE2 is a proapoptotic factor. When necrosis process is favored, the AM becomes necrotic and contributes to the MTB spread. UISS-TB is able to reproduce this behavior. As one can see, after ID93 vaccine administration, LXA4 level decreases favoring a pro-apoptotic behavior, and consequently reducing necrotic alveolar macrophages.

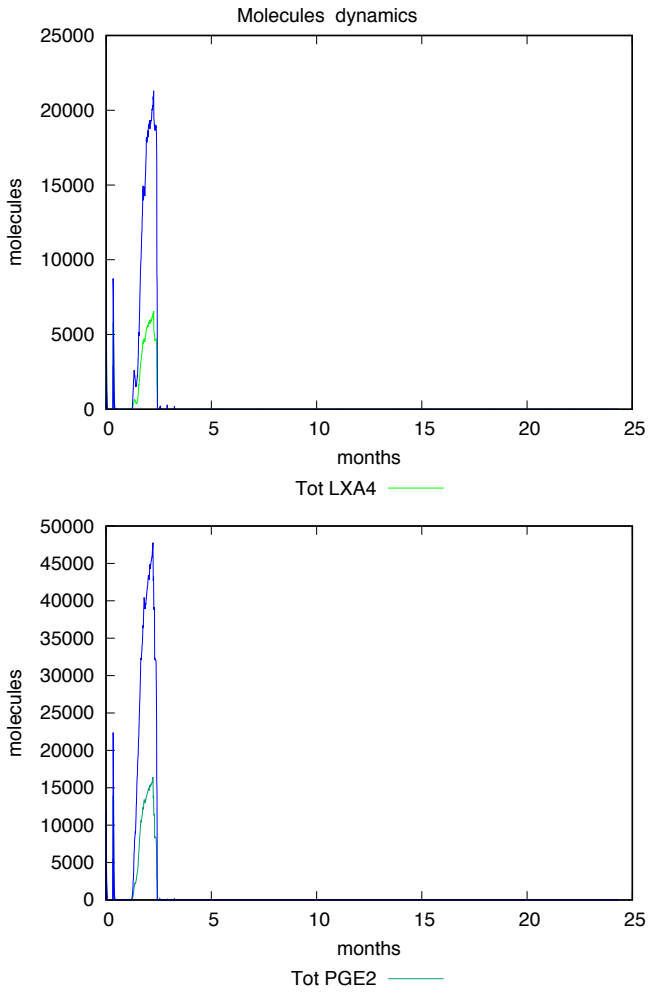

Figure 13. LXA4 and PGE2 levels.

It is worth to mention that, looking at the +/- SD data, UISS-TB is able to identify very few virtual in silico patients that are actually not responding to vaccine stimuli. At the same time, there are also few virtual in silico patients that are very good responders. As an example, we reported in figures (14-17) the AM population dynamics, CD4 T cell population dynamics and IFN-gamma levels of a specific case of both not responder and responder, respectively for RUTI and ID93 vaccine.

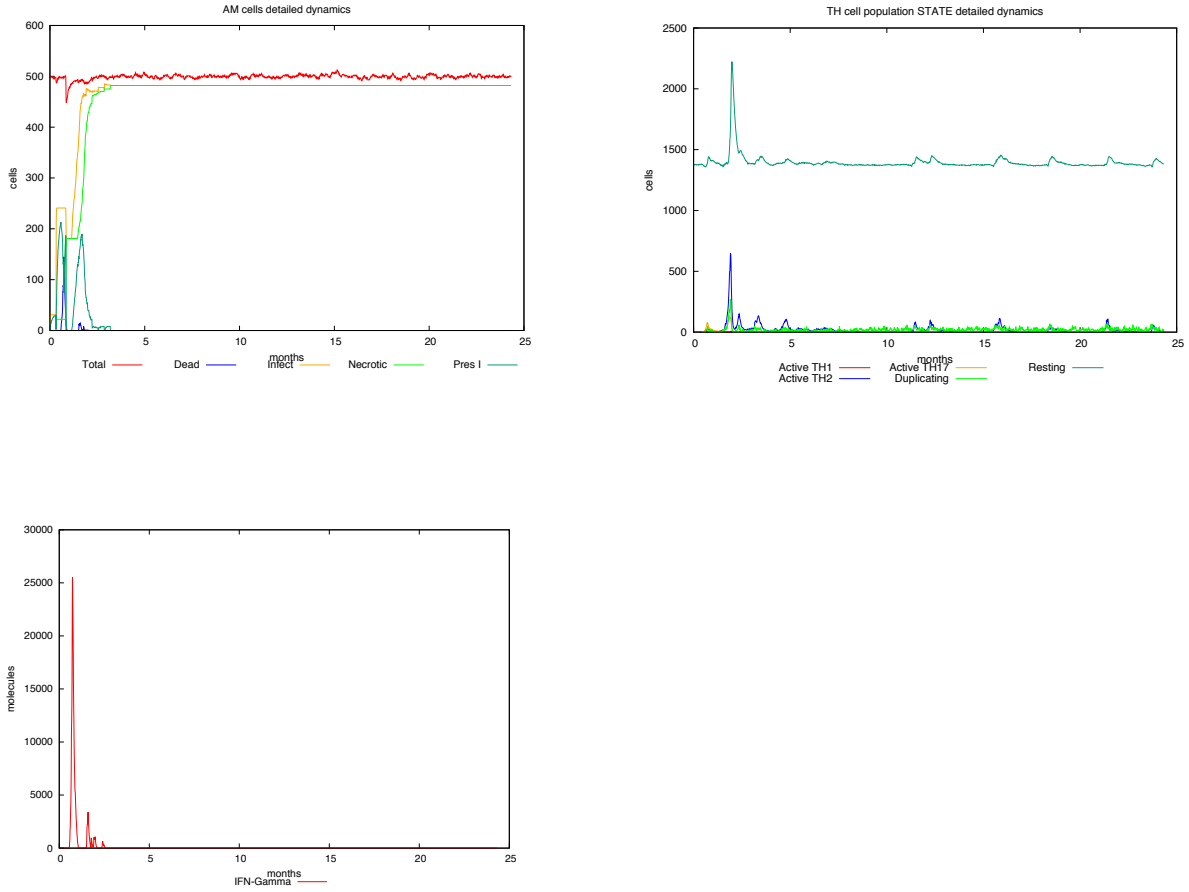

Figure 14. AM population dynamics, CD4 T cell population dynamics and IFN-gamma levels in a not responder case after RUTI administration.

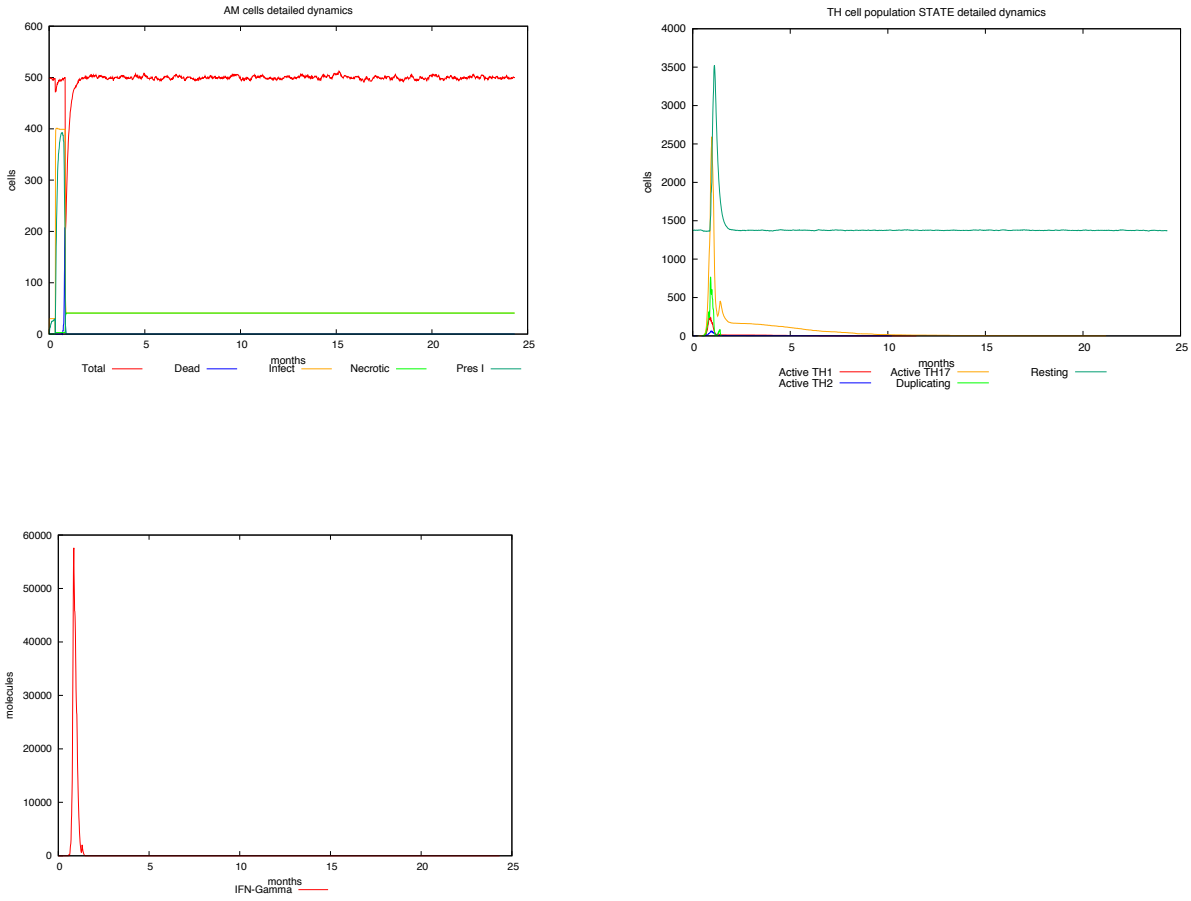

Figure 15. AM population dynamics, CD4 T cell population dynamics and IFN-gamma levels in a responder case after RUTI administration.

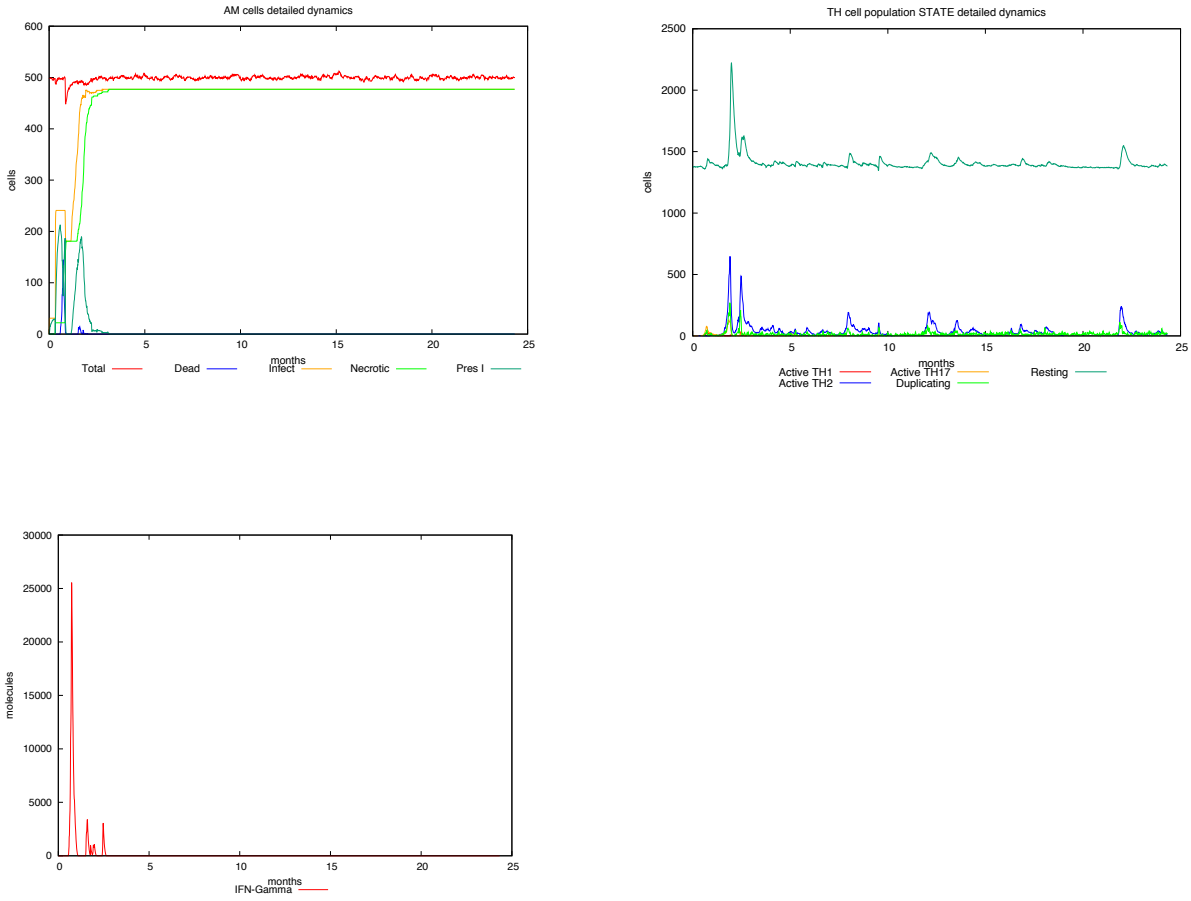

Figure 16. AM population dynamics, CD4 T cell population dynamics and IFN-gamma levels in a not responder case after ID93 administration.

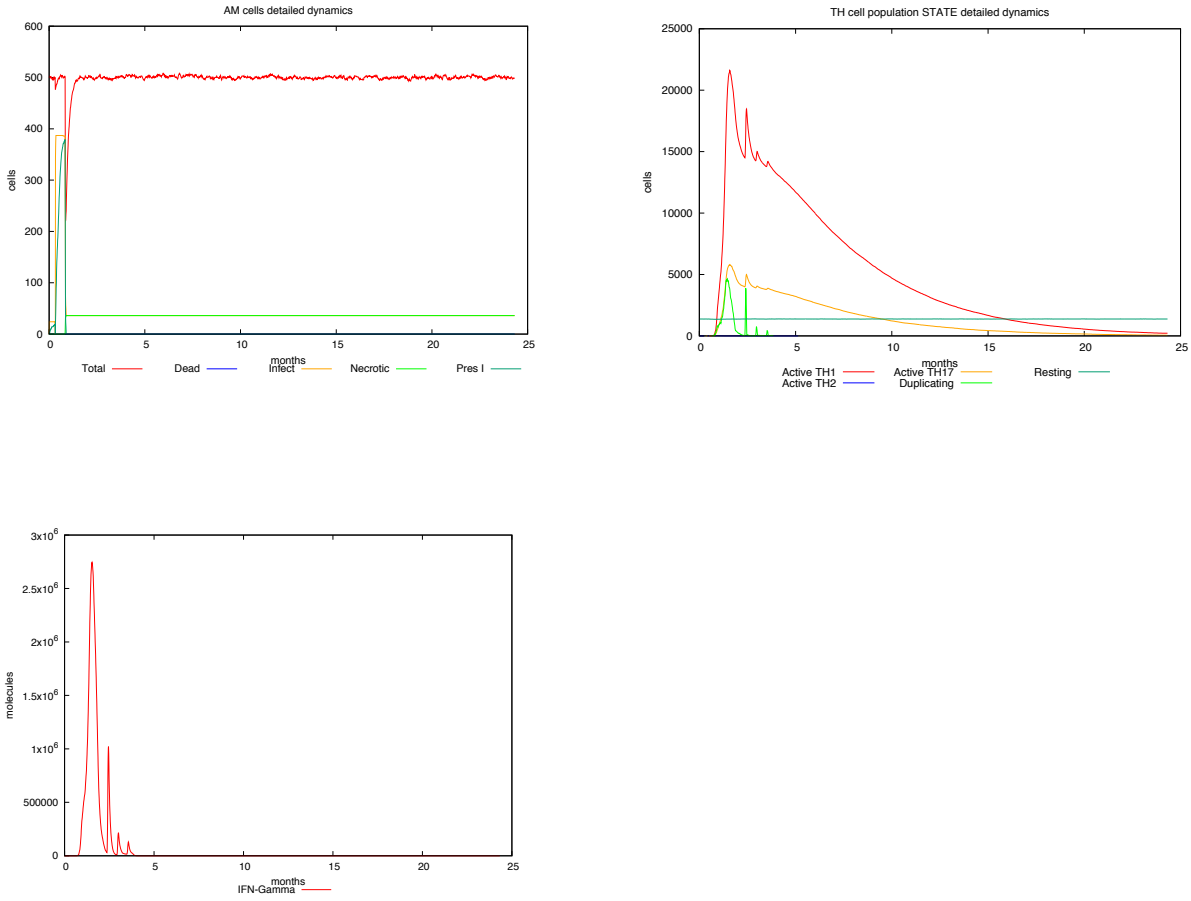

Figure 17. AM population dynamics, CD4 T cell population dynamics and IFN-gamma levels in a responder case after ID93 administration.

Summarizing, UISS-TB reveals that not-responders patients are identified by immune system repertoire, virulence of MTB strain that lead to an insufficient CD4 T cell Type 1 response with the correspondent low levels of IFN-gamma.

Conclusions

The simulations results are in very good agreement with available data coming from retrospective vaccine data coming from IDRI and Archivel Farma partners. In particular, the UISS-TB simulation platform shows reliability in capturing the main landmarks of both induced immune response mirroring the IFN-γ, CD4 T cells, LXA4 and PGE2 dynamics. Moreover, from the spatial point of view, the granuloma dynamics along with MTB spread were appropriately reproduced during the two most widespread scenarios characterizing pulmonary tuberculosis i.e., latent and latent with reactivation.

Using the first estimate of the covariance matrix Σ, a first cohort of 100 virtual patients was generated, in order to confirm the feasibility of the whole procedure. The entire workflow has been encoded in a program written in R programming language, making trivial the generation of any larger virtual cohort. Preliminary results appear to be consistent with the disease progression reported in the literature for non-treated patients.

These results complete task 2.7, as detailed in the Description of the Action, and close the activities associated to WP2. The work on the virtual cohorts will now continue in WP5, with the completion of the systematic review, and the generation of reference virtual populations to be used as part of the technical validation of UISS-TB. Once the data from the clinical trials are available, we will regenerate the virtual cohorts, and use the Bayesian statistical model developed in WP3 to explore specific use cases, such as that of in silico-augmented clinical trials, where virtual and physical patients are combined.

Appendix A

Let x = {x_1, ..., x_d} be the vector of features as described above. Assume you are interested in only one element of x, say x_i, i = 1, . . . , d. The marginal distribution of x_i is

$N(x_i \mid \mu_i, \sigma_i^2)$,

$$f(x_i \mid \mu_i, \sigma_i^2) = \frac{1}{2\pi\sigma_i^2} \exp\left[-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right]$$

where μ_i is the i-th entry of μ and σ_i^2 is the (i, i) entry of Σ.

Now, assume you have observed/fixed the values of (without loss of generality) the first d − q entries of x, so x = {x_f, x_r} with $x_r \in R^q$ and $x_f \in R^{d-q}$. The conditional distribution of x_r, given that $\boldsymbol{x}_f = \boldsymbol{a}$, $f(\boldsymbol{x}_r \mid \boldsymbol{x}_f = \boldsymbol{a}) = N_q(\boldsymbol{x}_r \mid \boldsymbol{\nu}, \Omega)$ with

$$\boldsymbol{\nu} = \boldsymbol{\mu}_r + \Sigma_{rf}\Sigma_{ff}^{-1}(\boldsymbol{a} - \boldsymbol{\mu}_f) \quad \text{and} \quad \Omega = \Sigma_{rr} - \Sigma_{rf}\Sigma_{ff}^{-1}\Sigma_{fr}$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{ff} & \Sigma_{fr} \\ \Sigma_{rf} & \Sigma_{rr} \end{pmatrix} \quad \text{with sizes} \quad \begin{pmatrix} (d-q)\times(d-q) & (d-q)\times q \\ q\times(d-q) & q\times q \end{pmatrix}.$$

Ω is the Schur complement of Σ_{rr} in Σ; i.e. to calculate Ω, one inverts Σ, drops the rows and columns

q × (d − q) q × q corresponding to the variables being conditioned upon, and then inverts back to get Ω.

For illustration, assume d = 2, i.e. x = {x_1, x_2}, μ = {μ_1, μ_2} and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

The conditional distribution, f(x1 | x2) is Gaussian with mean

$$\mathbb{E}[x_1 \mid x_2] = \mu_1 + \rho_{12}\frac{\sigma_1}{\sigma_2}(\mu_2 - x_2) \quad \text{and variance} \quad \mathbb{V}[x_1 \mid x_2] = \sigma_1^2(1 - \rho_{12}^2),$$

where $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}}$ is the correlation coefficient.

	Str	NAg	Th1	Th2	IgG	TC	IL1	IL2	IL10	IL12	IL17	IL23	IFN1	IFN_γ	TNF_α	TGF_β	LXA4	PGE2	CK	VD	MHC1	MHC2	Treg	Age	BMI
Str	1																								
NAg	+	1																							
Th1	+	+	1																						
Th2	+	+	+	1																					
IgG	+	+	+	+	1																				
TC	+	+	+	+	+	1																			
IL1	+	+	+	+	+	+	1																		
IL2	+	+	+	+	+	+	+	1																	
IL10	−	−	−	−	−	−	−	−	1																
IL12	+	+	+	+	+	+	+	-	−	1															
IL17	+	+	+	+	+	+	+	+	−	+	1														
IL23	+	+	+	+	+	+	+	+	−	+	+	1													
IFN1	+	+	+	+	+	+	+	+	−	+	+	+	1												
IFN_γ	+	+	+	+	+	+	+	+	−	+	+	+	+	1											
TNF_α	+	+	+	+	+	+	+	+	−	+	+	+	+	+	1										
TGF_β	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	1									
LXA4	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	1								
PGE2	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	1							
CK	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	1						
VD	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	1					
MHC1	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	1				
MHC2	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	1			
Treg	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	1		
Age	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1	
BMI	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1

Table 3. Correlation matrix of UISS-TB model inputs. We think all features have positive correlations, but IL-10, which is negatively correlated with the rest. We need to quantify these values for a successful generation of virtual patients.