HPC/Exascale Centre of Excellence in Personalised Medicine

D3.1. Use case work plans Version 1.0

Document Information

Contract Number	951773
Project Website	http://www.permedcoe.eu/
Contractual Deadline	M6, March 2021
Dissemination Level	PU
Nature	R
Author(s)	Laurence Calzone (IC), Arnau Montagud (BSC)
Contributor(s)	Jose Carbonell (BSC), Julio Sáez-Rodríguez (UKHD), Attila Gabor (UKHD), Laurent Heirendt (UNILU), Wei Gu (UNILU), Christophe Trefois (UNILU), and Vincent Noël (IC)
Reviewer(s)	Jesse Harrison (CSC), Adrian Thorogood (UNILU)
Keywords	Modelling applications, omics data, personalised models

Notice: *The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No "951773".*

Change Log

Version	Author	Date	Description of Change
V0.1	Laurence Calzone Arnau Montagud	17/02/21	Initial Draft, Table of content sent to Steering Board
V0.2	Laurence Calzone Arnau Montagud	04/03/21	Final Draft, document sent to reviewers
V0.3	Jesse Harrison Adrian Thorogood	18/03/21	Reviewers' comments received
V1.0	Laurence Calzone Arnau Montagud	22/03/21	Reviewers' comments integrated in the final document
V1.0	Alba Jené	30/03/21	Final Edits before submission to the EC.

Table of contents

Executive Summary

PerMedCoE use cases are scenarios that have been designed to demonstrate the use of different modelling core tools on HPC/Exascale supercomputers: MaBoSS, CelNOpt, COBRA and PhysiBoSS.

This deliverable describes detailed work plans for all the use cases considered in PerMedCoE, with a particular focus on the synergies between WP1, WP2 and WP3. The descriptions of the relevant use cases also include details on collaborations with domain experts, organised communities and coordination with other Centres of Excellence.

The data used in the use cases encompass both public and private datasets. Each of these use cases requires different methodologies, tools and data, demonstrating the broad range of biological scenarios that will benefit from using PerMedCoE core tools to perform computationally demanding analyses on different supercomputing environments.

Finally, the deliverable defines requirements in relation to addressing PerMedCoE use cases in pre-exascale environments. This information is detailed in each use case section and gathered in Annex 1.

1. Introduction

The purpose of PerMedCoE WP3 is to establish biologically relevant use cases to ensure a proper framework to test the capacities and pipelines established in WP1 and WP2. Five use cases were chosen to reflect a broad range of computationally demanding real-life scenarios, in terms of both the models employed and the types of data required. While four use cases were initially formulated, due to the emergence of the current worldwide pandemic, a fifth use case related to SARS-CoV-2 infection and COVID-19 progression was subsequently added.

Based on discussions between PerMedCoE partners, the use cases are arranged in the following order of priority:

- Use case 2: Drug synergies on cell lines
- Use case 5: COVID use case using single-cell data
- Use case 4: Tumour evolution based on single-cell omics and imaging
- Use case 1: Cancer diagnosis based on omics Information
- Use case 3: Personalised modelling of groups of rare-disease related patients

This order of priority was established based on three different criteria: “Feasibility of acquiring the data”, “Impact of HPC in use cases”, and “Impact of the tools and use cases in the targeted communities” (Table1).

Criteria	Description of criteria	Use case 1	Use case 2	Use case 3	Use case 4	Use case 5
Feasibility of acquiring the data	Data are public (2), already available via the consortium (1) or external (0)	0	2 (and 1)	0 (and 2)	1 (and 0)	2 (and 1)
Impact of HPC in use cases	Is HPC an enabling technology (2), would it speed analyses (1), or would the analyses not be affected (0)?	1	1	1	2	1
Impact of the tools and use cases in the targeted communities	Present and future value of the tools to researchers working in bioinformatics and the life sciences [1-3]	2	2	1	2	3

Table 1: Evaluation of different criteria to prioritise the PerMedCoE use cases. The numbers in each use case represent the priority criteria that are described in column 2 and that are specific to each of the criteria. For instance, for the “feasibility of acquiring data” criteria, 2 would correspond to public data, 1 to data from the consortium and 0 to external data

The work plans for each use case comprise the following topics: (1) definition of the use case and the computational issues likely to be met; (2) data used; (3) models used; (4) tools tested; (5) methodology or an outline of the work plan; and (6) requirements to run in a pre-exascale environment.

To meet the objective of employing the use cases to address questions that are critical to advancing personalised medicine as a field and benefit from access to HPC clusters, we plan to prepare and disseminate pipelines of analyses tailored to each of the use cases. For instance, Python notebooks will be provided for individual cell models run in MaBoSS in Use Case 2. These notebooks will be optimised in collaboration with partners from other WPs, notably WP2.

One of the goals of WP2 is to place the core tools in building blocks to increase their usability and reproducibility. Combining different building blocks, PerMedCoE will offer pre-built workflows that will include standardised steps such as genomic analysis, integration of data in models, running simulations, etc. To underline that these workflows can be user-driven, we will use them in the use cases, improving the reproducibility of our analyses and its replication is several HPC clusters. In addition, we aim to have a meta-tool for workflow construction to try and cater for simple use cases not described in present deliverable. More information on this may be found in milestone MS07 and deliverable D2.2.

Finally, two other deliverables are related to the present use cases. A data management plan encompassing all the datasets featured in these use cases has been detailed in D3.2. Details concerning the appropriate ethics clearance steps toward data usage in the project are outlined in D5.6.

2. Description of the use cases

In this section, we describe each use case, the data, the tools, the models, and the requirements related to the consortium. Note that the order of the use cases reflects their priority in PerMedCoE.

2.1. Use Case 2: Drug synergies on cell lines

Definition of the use case

Use Case 2 is defined in the PerMedCoE grant proposal as follows:

« *This use case will find effective drug combinations for cancer by using drug-response experiments and publicly available databases, personalised cell-level models and BioExcel's GROMACS simulations.* »

The search for appropriate drug combinations using mathematical models involves providing a list of combined treatments that could lower the dose and the toxicity of therapies. A recent application of such approaches is the study of immunotherapies. Treatments combining the immune checkpoint inhibitors anti-CTLA4 and anti-PD1 have shown some improvement in patient response compared to single treatments, in particular in Merkel cell carcinoma, MSI-high cancers and Hodgkin's lymphoma [1,2].

Even when using a high-end workstation to search for the optimal combination of inhibitors in models, using more than 20 nodes can lead to a combinatorial explosion that brings computational issues. The study of these combinations is a good target to be parallelised as these simulations are independent of each other. The purpose of this use case is to optimise the search for personalised models of cancer cell lines by using compartmentalised, distributed workflows as well as pre-exascale-ready parallelised tools.

The output of this use case will provide a framework for integrating omics data into models, as well as molecular dynamics screenings and including PK/PD data (when available) with Boolean models. Further to these novel features, using distributed workflows will set the standard to be used when dealing with federated data repositories that are not allowed to leave the server or country.

Data

Both public and private datasets will be used for this use case. Public cancer cell line data from Sanger Institute, Broad Institute and GDSC will be mined and used for the selected cancer applications: gastric, colon and prostate.

Published data concerning cell line treatments will be used [16]. The data consist of eight cell lines for which growth is studied under combinations of targeted cancer therapies for a total of 171 combinations.

Private datasets will be shared by NTNU on gastric cancer cell lines. The dataset contains survival information, RNAseq, RPPA and mass spectrometry data pertaining to the AGS gastric cell line in response to treatments of single and double combinations of three drugs.

Models

This use case will make use of existing signalling models of gastric cancer [3], colon cancer (in progress) and prostate cancer (in progress) developed by members of the consortium (IC, UKHD, BSC) and collaborators (NTNU), as well as a generic cancer model [4]. Additionally, metabolic models such as RECON 3D [17] will be used to capture the cell line metabolism.

Models can be tailored to these cell lines, using PROFILE [18] or CORDA [21], and methods are already in place to simulate them using the core applications.

Tools

CellNOpt [5] will be used to personalise the models to cancer cell lines, on which drug treatments (node inhibitions) will be simulated.

MaBoSS [6,7] will first apply the PROFILE pipeline for model personalisation using transcriptomics and mutational information provided on cell lines, and automatic single and double perturbations of nodes of the model will be performed on the obtained models.

COBRA [18] will be used to simulate previously personalised metabolic models to provide insights into how drugs can perturb the metabolic flux landscape.

Methodology to be used

Data will be obtained from publicly available databases including CCLE and GDSC (copy number, gene expression, drug response), AstraZeneca/DREAM (~900 combos across 85 cell lines), Merck drug screen (~580 combinations across 39 cell lines; [8]) and the ALMANAC study from NIH (~5000 combinations across 60 cell lines; [9]). A single dataset will be provided by collaborators at NTNU, Norway (171 combinations over 8 cell lines).

Using CellNOpt, MaBoSS and COBRA, cell-line-specific simulations will be run and the list of potential drug bindings will be screened to identify those that exhibit strong synergic interactions. COBRA will also be run using the personalised metabolic models and cell-line-specific models as inputs, with the resulting simulations providing complementary insights into how drug combinations can perturb the metabolic flux landscape.

Where possible, the identification of potential drugs and drug targets will be complemented with virtual screening and molecular dynamics simulations using the

pre-exascale version of GROMACS produced by BioExcel, thus expanding the catalogue of potential drugs and actionable therapeutic targets.

Requirements to run in a pre-exascale environment

- WP1

Use case 2 requires access to high-throughput simulations using MaBoSS, CellNOpt and/or COBRA to simulate thousands of cell lines at once. While the HPC-ready core applications are under preparation, model exploration methods already in place in the HPC clusters can be used to fit the parameters of the tailored models (transport rates, the effect on proteins, etc.). In addition, HPC-ready orchestrators can be used to parallelise otherwise linear code.

The use case will also require downstream analysis scripts supporting the use of these high-throughput simulations by enabling browsing, clustering and the identification of patterns in the data.

- WP2

Use Case 2 requires access to workflows for model personalisation. For further details on PerMedCoE worfklow design choices, see Milestone 07 (Design choices for building blocks).

Additional goals

- WP2

As additional goals, Use Case 2 would benefit from access to workflows enabling the execution of thousands of simulations at once, and distributed workflows that allow working with federated data repositories.

2.2. Use Case 5: COVID use case using single-cell data

Definition of the use case

This use case consists of modelling the cell types and intracellular pathways that have been identified as playing a role in COVID-19. The model will be based on ongoing efforts from the Disease map community [10,11] to construct individual maps of the SARS-CoV-2 virus and its interactions with the epithelial human host cells and the rest of immune cell types and processes that are at stake in the disease: virus replication, host defence mechanisms, innate immune response, etc. The goal of this use case is to provide a multiscale model of COVID-19 with the most relevant cell types and take advantage of multiscale modelling to suggest points of interventions to combat the virus.

By developing a simulation method that includes immune and epithelial human cells, virus infection, as well as genetic and environmental perturbations, Use Case 5 can help address key research questions concerning SARS-CoV-2 and its relationship with

human hosts. For example, such a method could be used to interrogate how different hosts with different degrees of COVID-19 symptoms respond to and fight the virus, tackling a current blind spot in the field.

This use case was not in the original project submission and was introduced at the kick-off meeting in October 2020. One of the features of HPC Center of Excellence is to maintain awareness of the field of study and remain flexible enough to focus on emerging topics requiring the development of new tools and use cases. The consortium deemed that introducing Use Case 5 was justified given the urgency of the COVID-19 pandemic, as well as the feasibility of COVID-19 modelling using the PerMedCoE core tool set.

Data

Most of the data used as part of Use Case 5 are freely accessible, while access to specific datasets will have to be agreed upon with data providers. An example dataset that will be used is GSE148729 () that compiles single-cell and bulk RNA-seq from three human cell lines (H1299, Caco-2 and Calu-3 cells) infected by SARS-CoV-1/2.

Models

The models simulated with PhysiBoSS will be adapted from maps provided by the Disease Map community by means of the CaSQ tool [12] and adapted by manual curation. If the models are unavailable or do not include the expected proteins, logical models will be sought for in existing model databases (e.g., Cell Collective , GINsim [13], BioModels). Examples of topics addressed by the models include the differentiation of macrophages, lymphocytes, and the division and death of epithelial cells.

Tools

MaBoSS will be used to rapidly evaluate the effect of all mutants on cell behaviour.

PhysiCell will be extensively used in this use case, as well as versions of PhysiCell incorporating MaBoSS (PhysiBoSS) and MPI (PhysiCell-MPI).

Methodology to be used

As part of Use Case 5, models will be built using maps available via the Disease Maps community (https://disease-maps.org/) that will be translated into logical models using the CasQ tool [12]. The omics data from single-cell experiments will be explored and fit to personalise models to different patient groups (by age, gender, nationality, etc.) [19]. MaBoSS will be used to screen potential patient-specific biomarkers and therapeutic targets.

Following analyses employing MaBoSS,PhysiBoSS will be used tobuild upon existing models of SARS-CoV-2 with PhysiCell (, PC4COVID, [14]), with this tool providing access a multiscale modelling framework that incorporates environmental and population

descriptions as well as intracellular evaluation. Institut Curie has already developed a working example of such a model that will be extended.

In parallel with these efforts, development work in relation to PhysiCell-MPI will be extended to PhysiBoSS. This will increase the scope of the simulations by several orders of magnitude and enable the simulation of complex behaviours.

Requirements to run in a pre-exascale environment

- WP1

As part of Use Case 5, developmentf work related to PhysiBoSS and must be extended to PhysiCell-MPI (resulting in a tool termed PhysiBoSS-MPI).

- WP2

This use case requires workflows incorporating model exploration methodologies, as well as features facilitating strategies for the exploration of the model parameter space. This would enable multiple-parameter fitting using HPC clusters.

Further, the workflows employed as part of Use Case 5 should enable monitoring of simulations to kill those simulations that are not useful or to modify parameters to return to a specific parameter space (termed “trajectory analysis during execution”).

Additional goals

- WP1

A further goal for Use Case 5 is to incorporate CellNOpt as part of PhysiCell, following the addition of PhysiBoSS 2.0.

- WP2

This use case could benefit from access to complete single-cell analysis workflows.

2.3. Use Case 4: Tumour evolution based on single-cell omics and imaging

Definition of the use case

Use Case 4 is defined in the PerMedCoE grant proposal as follows:

« *Simulations of millions of individual cells, each one of them with their own omics data and individual models of metabolism, regulation and signalling. Cells will interact with each other (agent-based models) and with the environment, modelled as fluids, from where cells get nutrients and input information, in gradients related to their relative position to the other cells.* »

It is well established that the tumour microenvironment impacts intracellular cell signalling and, consequently, the fate of each cell. If most mathematical models focus on individual cell types, we aim at building a comprehensive model with tumour cells

and immune cells that can interact and respond to the status of their microenvironment. In particular, we are interested in how cells respond to anti-tumoral treatments and how it affects tumorigenesis in patients that show resistance to these treatments.

The goal of this use case is to simulate real-sized tumours and provide a tool to test novel drugs or novel treatments, which is a much-needed framework in the field.

Data

Single-cell data will be gathered from collaborators at the MDC, Germany, co-leading the LifeTime initiative and European experts in single-cell omics data generation and analysis.

Published live cells imaging data will be used to reproduce cell invasion characteristics (similar to [15]).

As this use case needs cell-type-specific data to have cell-type-specific simulations, an example dataset will be the two-time point tumorigenesis tracing with scRNA-seq on a mouse model from MDC (GSE124425). This dataset allows having a single-cell resolution level to simulate tumour growth.

Models

We will use existing models, both published and in construction, of cell metabolism and cell invasion. The intracellular model of cell invasion in which we are working on considers the signalling pathways that are activated by DNA damage or the presence of some molecular entities in the microenvironment, such as growth factors.

We will use generic models such as RECON 3 for metabolic model simulations that may be personalised to capture characteristics of the specific cell line in use.

We are also exploring the possibility to use image analysis methods in ML toolbox.

Tools

The tools used for this use case will include COBRA for the metabolic models and MaBoSS, PhysiBoSS and/or CellNOpt for the models of the tumour evolution.

Tools will be deployed on HPC environments as soon as possible. PhysiCell-MPI is currently under beta testing, PhysiCell + MaBoSS is ready to be used and needs to be incorporated into PhysiCell-MPI. The scaling of the other core applications is still under development.

Methodology

The molecular characteristics of the single-cell data used for this use case will be analysed with the omics analysis workflows discussed in WP2.

The data will be organised in multi-layer networks, which will be used to reconstruct cell-specific gene-interaction networks that will lead to cell-specific models.

Depending on the available data, intracellular simulations will be done with COBRA software (metabolic data) orwith MaBoSS or CellNOpt (molecular descriptions).

Images from tumours will be gathered to replicate their behaviours in intercellular simulations using PhysiCell or PhysiBoSS.

We will predict tumour evolution, accounting for (1) variability in cell types, such as immune cells, (2) heterogeneity in the environment topological, chemical or temporal, and (3) cancer cells-other cells-environment interactions and communication.

Finally, we will discuss the results with the LifeTime single-cell community collaborators and propose new hypotheses for testing.

In this use case, HPC is an enabling technology to have a tumour simulation of billions of cells from a starting seeding cell with genetic and environmental perturbations. These huge simulations are needed to simulate real-sized tumours and are a step towards having digital twins on which to test novel drugs or novel treatments.

Requirements to run in a pre-exascale environment

- WP1

To implement realistic tumour simulations, Use Case 4 requires PhysiBoSS-MPI.

- WP2

We need several workflows for this use case: for the single-cell analyses, for the integration of these data in multi-layers, and to have model exploration techniques.

Additional goals

- WP1

To incorporate CellNOpt and COBRA inside PhysiCell, following the addition of PhysiBoSS 2.0.

- WP2

To have workflows to analyse anatomopathological images and compare them to simulations outputs. Verifying that the simulations are successful will require their comparison against anatomopathological images of real tumours.

2.4. Use Case 1: Cancer diagnosis based on omics information

Definition of the use case

Use Case 1 is defined in the PerMedCoE grant proposal as follows:

« *This use case will propose cancer treatments for individual patients using associated clinical information employing omics data and personalised cell-level models.* »

In this use case, we plan to investigate the possibility of complementing the current generation of bioinformatics methods with cellular models, which provide mechanistic descriptions and testable hypotheses instead of current statistical approximations. For this, we aim to decipher patient-specific molecular characteristics of individual cancer cases out of omics data that help us identify biomarkers and therapeutic targets.

The discussion in the field on what is the degree of accuracy of the patient-specific molecular characteristics found by the models with regards to the patients' data is quite open and needs to be addressed. In addition to that, it will be useful for the community to identify if the model's degree of granularity only captures differences at the level pf subgroups, or whether it can also reach the level of individual patients.

In this particular case, we will focus on the study of Chronic Lymphocytic Leukemia (CLL), which is a specific type of blood cancer characterised by the alteration and subsequent accumulation of B lymphocytes [18]. Our simulations will contribute to understanding how the stochastic accumulation of patient-specific somatic mutations changes the behaviour of B lymphocytes progressively from normal to tumoral phenotypes, consequently modifying the tumour microenvironment.

Data

The use case will employ Chronic Lymphocytic Leukemia data from IDIBAPS hospital, Spain. This dataset consists of different data from 551 patients: mutations, copy number alterations, germline variants, DNA methylation and expression.

Models

We will use models of signalling pathways (e.g., using a generic model as the basis [4]) as well as metabolism (e.g., starting from RECON 3D) and will obtain patient-specific ones using personalisation techniques (PROFILE and CORDA).

Furthermore, we will explore the possibility to include additional logical models [19] to represent the transdifferentiation process experienced by lymphoid cells. Also, we will evaluate the inclusion of logical models of macrophages cells [20] to illustrate the interplay between the tumour B lymphocytes and the innate immune system.

Tools

One or several of the intracellular core tools considered in the project will be used: MaBoSS, CellNOpt or COBRA.

Methodology

Different omics and clinical data will be gathered from collaborators at IDIBAPS hospital, Spain and organised in a multi-layer network. Key molecular patient-specific characteristics will be derived from omics data.

Then, MaBoSS, COBRA and/or CellNOpt will be used, and intracellular simulations will be run using the calibrated patient-specific models of genetic and metabolic interactions to decipher patient-specific molecular characteristics of individual cancer cases (Chronic Lymphocytic Leukemia in the case of IDIBAPS). In particular, at the end of the executions, the main simulation features (such as the number of tumour cells) will be gathered and correlated with patient-specific clinical variables, including survival probability or specific subtype.

Requirements to run in a pre-exascale environment

- WP1

Use Case 1 requires HPC-ready simulations using MaBoSS, CellNOpt and/or COBRA. While HPC-ready versions of the core applications are being developed, model exploration methods already in place in the HPC clusters can be used to fit the parameters of the tailored models (transport rates, the effect on proteins, etc.). In addition, HPC-ready orchestrators can be used to parallelise otherwise linear code.

- WP2

The use case requires access to workflows with model exploration methodologies. This would allow us to fit many parameters using HPC clusters.

Workflows are also required to retrieve signatures from omics datasets, also called “Clustering and analysis of simulation results” in WP2. This would allow comparing these signatures against the outputs of the models.

Additional goals

- WP2

An additional goal pertaining to Use Case 1 is to have workflows for network reconstruction from omics datasets.

2.5. Use Case 3: Personalised modelling of groups of rare-disease related patients

Definition of the use case

Use Case 3 is defined in the PerMedCoE grant proposal as follows:

« *This use case will focus on simulations of individual disease characteristics, such as the degree of severity, in support of the rare diseases’ diagnoses where only a small number of cases are available* »

The use case will focus on the identification of genetic and epigenetic bases that explain different degrees of affectation related to consanguinity, thus identifying common characteristics and differences that determine their degree of severity of the diseases. Studying ways to better tailor data to models in problems with a scarcity of patients and most of them with family relationships is a useful problem for the community.

This use case was deemed to have the least priority in our evaluation (see Introduction). It does not require any additional functionalities for the core tools in comparison to the other use cases. The rest of the use cases already constitute a comprehensive collection of use cases on which to showcase PerMedCoE's core tools. Thus, we will focus on them and leave this use case as a back-up.

Data

Where use case will employ different omics and clinical data from the CHEO Research Institute, Canada.

Models

We will use different models for each disease and will personalise them to have patient- and disease-specific models. Depending on the available data, we will work on signalling pathway models or also metabolic models.

Tools

Depending on the available data, we will work on signalling pathway tools (CellNOpt, MaBoSS) or also metabolic modelling tools (COBRA).

Methodology

We will gather omics and clinical data from collaborators at the CHEO Research Institute, Canada and organise them in multi-layers.

The data will be processed with existing Genome analysis High Content methods (for discussion on workflows, see WP2) to derive disease-specific molecular markers at the individual level to describe the genetic and epigenetic basis of the observed different degrees of diseases affectation between patients that are connected by different degrees of consanguinity, as typical in rare disease research.

Relevant gene-interaction networks contextualised with patient-specific omics information will provide intra-cellular models for individual patients.

Modelling the effect of intra-cellular genetic variations with CellNOpt or MaBoSS will make it possible to capture common characteristics, as well as the individual differences that determine their degree of severity of the disease.

Requirements to run in a pre-exascale environment

- WP2

To conduct genome analyses as part of this use case, we need the High Content methods workflow ready.

3. Partners' involvement

Some partners have been designated as responsible for each use case. Each of them will ensure the proper management and realisation of the results.

Regular meetings with members involved in WP1 and WP2 will be planned and will ensure proper communication between all parties.

	Data source	Curie	BSC	UNILU	MDC	UKHD	IRB	CSC
PM in T3.2		14	12	9	8	7	4	4
UC1	External		R					X
UC2	Public + Consortium	R2	X			R1		
UC3	External + Public							
UC4	Consortium + External	X	R1	R2	X	X	X	
UC5	Public + Consortium	R1	R2			X		

Table 2: Partners assignments to each use case. X means involvement, R sole responsible and R1 and R2 co-responsible. PM: person-months.

The persons to contact for each of the use cases are the following:

- UC1: Jose Carbonell (BSC)
- UC2: Julio Sáez-Rodríguez (UKHD), Laurence Calzone (IC)
- UC3: to be determined
- UC4: Arnau Montagud (BSC), Wei Gu (UNILU) / Christophe Trefois (UNILU)
- UC5: Vincent Noël (IC), Arnau Montagud (BSC)

4. References

[1] Ribas, A.; Wolchok, J.D. Cancer immunotherapy using checkpoint blockade. *Science (80-.).* 2018, *359*, 1350–1355.

[2] Darvin, P.; Toor, S.M.; Sasidharan Nair, V.; Elkord, E. Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp. Mol. Med.* 2018, *50*.

[3] Flobak Å, Baudot A, Remy E, Thommesen L, Thieffry D, et al. (2015) Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. PLOS Computational Biology 11(8): e1004426.

[4] Fumiã HF, Martins ML. Boolean network model for cancer pathways: predicting carcinogenesis and targeted therapy outcomes. PLoS One. 2013 Jul 26;8(7):e69008. doi: 10.1371/journal.pone.0069008.

[5] Morris MK, Melas I, Saez-Rodriguez J. Construction of cell type-specific logic models of signaling networks using CellNOpt. Methods Mol Biol. 2013;930:179-214. doi: 10.1007/978-1-62703-059-5_8.

[6] Stoll, G. E. Viara, E. Barillot and L. **Calzone** (2012) Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. BMC Systems Biology.2012, 6:116.

[7] Stoll G, Caron B, Viara E, Dugourd A, Zinovyev A, Naldi A, Kroemer G, Barillot E, **Calzone L.** MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics. 2017 Jul 15;33(14):2226-2228.

[8] O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, Li J, Kral A, Lejnine S, Loboda A, Arthur W, Cristescu R, Haines BB, Winter C, Zhang T, Bloecher A, Shumway SD. An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. Mol Cancer Ther. 2016 Jun;15(6):1155-62. doi: 10.1158/1535-7163.MCT-15-0843. Epub 2016 Mar 16. PMID: 26983881.

[9] Holbeck SL, Camalier R, Crowell JA, Govindharajulu JP, Hollingshead M, Anderson LW, Polley E, Rubinstein L, Srivastava A, Wilsker D, Collins JM, Doroshow JH. The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. Cancer Res. 2017 Jul 1;77(13):3564-3576. doi: 10.1158/0008-5472.CAN-17-0489. Epub 2017 Apr 26. PMID: 28446463; PMCID: PMC5499996.

[10] Ostaszewski et al. ,COVID-19 Disease Map, a computational knowledge repository of SARS-CoV-2 virus-host interaction mechanisms

[11] Kinza Rian, Marina Esteban-Medina, Marta R. Hidalgo, Cankut Çubuk, Matias M. Falco, Carlos Loucera, Devrim Gunyel, Marek Ostaszewski, María Peña-Chilet, Joaquín Dopazo. Mechanistic modeling of the SARS-CoV-2 disease map. bioRxiv 2020.04.12.025577

[12] Sara Sadat Aghamiri, Vidisha Singh, Aurélien Naldi, Tomáš Helikar, Sylvain Soliman, Anna Niarakis, Automated inference of Boolean models from molecular interaction maps using CaSQ, *Bioinformatics*, Volume 36, Issue 16, 15 August 2020, Pages 4473–4482,

[13] Gonzalez AG, Naldi A, Sánchez L, Thieffry D, Chaouiya C. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. Biosystems. 2006 May;84(2):91-100. doi: 10.1016/j.biosystems.2005.10.003.

[14] Michael Getz, Yafei Wang, Gary An, Andrew Becker, Chase Cockrell, Nicholson Collier, Morgan Craig, Courtney L. Davis, James Faeder, Ashlee N. Ford Versypt, Juliano F. Gianlupi, James A. Glazier, Sara Hamis, Randy Heiland, Thomas Hillen, Dennis Hou, Mohammad Aminul Islam, Adrianne Jenner, Furkan Kurtoglu, Bing Liu, Fiona Macfarlane, Pablo Maygrundter, Penelope A Morel, Aarthi Narayanan, Jonathan Ozik, Elsje Pienaar, Padmini Rangamani, Jason Edward Shoemaker, Amber M. Smith, Paul Macklin. Rapid community-driven development of a SARS-CoV-2 tissue simulator bioRxiv 2020.04.02.019075

[15] Ferrari, R., Martin, G., Tagit, O. *et al.* MT1-MMP directs force-producing proteolytic contacts that drive tumor cell invasion. *Nat Commun* 10, 4886 (2019).

[16] Flobak, Å., Niederdorfer, B., Nakstad, V.T. et al. A high-throughput drug combination screen of targeted small molecule inhibitors in cancer cell lines. Sci Data 6, 237 (2019).

[17] Brunk, E., Sahoo, S., Zielinski, D. et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol 36, 272–281 (2018).

[18] Palsson Bernhard, Systems Biology: Constraint-based Reconstruction and Analysis, ISBN: 978-1-107-03885-1

[19] Béal J, Montagud A, Traynard P, Barillot E, Calzone L. Personalization of Logical Models With Multi-Omics Data Allows Clinical Stratification of Patients. Front Physiol. 2019 Jan 24;9:1965. doi: 10.3389/fphys.2018.01965. eCollection 2018.

[20] Marku M, Verstraete N, Raynal F, Madrid-Mencía M, Domagala M, Fournié JJ, Ysebaert L, Poupot M, Pancaldi V. Insights on TAM Formation from a Boolean Model of Macrophage Polarization Based on In Vitro Studies. Cancers (Basel). 2020 Dec 7;12(12):3664. doi: 10.3390/cancers12123664.

[21] Schultz A, Qutub AA: Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. PLoS Comput Biol 2016, 12:e1004808.

5. Annex I - MS08: Use cases' requirements to run in a pre-exascale environment

MS08: Use cases' requirements to run in a pre-exascale environment defined and fed to WP1 and WP2. The content of this part is replicated from the MS08 document prepared beforehand.

Use Case 2: Drug synergies on cell lines

- Requirements
- WP1

To have high-throughput simulations using MaBoSS, CellNOpt and/or COBRA to simulate thousands of cell lines at once. While waiting for the HPC-ready core applications, model exploration methods already in place in the HPC clusters can be used to fit the parameters of the tailored models (transport rates, the effect on proteins, etc.). In addition, HPC-ready orchestrators can also be used to parallelise otherwise linear code.

To have downstream analysis scripts of these high-throughput simulations that allow browsing, clustering and uncovering interesting properties.

- WP2

To have workflows to personalise models.

To have workflows to analyse thousands of drug synergy simulations at once.

  - Additional goals
- WP2

To have workflows to analyse thousands of simulations at once. To have distributed workflows that allow working with federated data repositories.

Use Case 5: COVID use case using single-cell data

- Requirements
- WP1

To merge the developments of PhysiBoSS and the ones from PhysiCell-MPI (termed PhysiBoSS-MPI).

- WP2:

To have workflows with model exploration methodologies, also called "strategies for the exploration of the parameter space" in the WP2. This would allow us to fit many parameters using HPC clusters.

To have workflows to monitor online the simulations to kill simulations that are not useful or to modify parameters to go back to a space of parameters deemed interesting. We referred to this as "Trajectory analysis during execution" in WP2.

- Additional goals
  - WP1

To incorporate CellNOpt inside PhysiCell, following the add-on structure of PhysiBoSS 2.0.

  - WP2

This use case could benefit from having single-cell analysis workflows ready.

Use Case 4: Tumour evolution based on the single-cell omics and imaging

- Requirements

  - WP1

To have realistic tumours, we need PhysiBoSS-MPI implemented.

  - WP2

We need several workflows for this use case: for the single-cell analyses, for the integration of these data in multi-layers and to have model exploration techniques.

- Additional goals

  - WP1

To incorporate CellNOpt and COBRA inside PhysiCell, following the add-on structure of PhysiBoSS 2.0.

  - WP2

To have workflows to analyse anatomopathological images and compare them to simulations' outputs. The ultimate verification that the simulations are successful will be the comparison against anatomopathological images of real tumours.

Use Case 1: Cancer Diagnosis Based on omics Information

- Requirements

  - WP1

To have HPC-ready simulations using MaBoSS, CellNOpt and/or COBRA would allow scaling the simulations of these models. While waiting for the HPC-ready core applications, model exploration methods already in place in the HPC clusters can be used to fit the parameters of the tailored models (transport rates, the effect on proteins, etc.). In addition, HPC-ready orchestrators can also be used to parallelise otherwise linear code.

  - WP2

To have workflows with model exploration methodologies. This would allow us to fit many parameters using HPC clusters.

To have workflows of bioinformatics analysis to retrieve signatures from omics datasets, also called “Clustering and analysis of simulation results” in WP2. This would allow comparing these signatures against the outputs of the models.

- Additional goals

- WP2

To have workflows of network reconstruction from omics datasets.

Use Case 3: Personalised Modelling of groups of rare-disease related patients

- Requirements

- WP2

To have genome analysis, we need the High Content methods workflow ready.

Acronyms and Abbreviations

Each term should be bulleted with a definition.

Below is an initial list that should be adapted to the given deliverable.

- CA – Consortium Agreement
- D – deliverable
- DoA – Description of Action (Annex 1 of the Grant Agreement)
- EB – Executive Board
- EC – European Commission
- GA – General Assembly / Grant Agreement
- HPC – High-Performance Computing
- IPR – Intellectual Property Right
- KPI – Key Performance Indicator
- M – Month
- MS – Milestones
- PM – Person month / Project manager
- UC – Use Case
- WP – Work Package
- WPL – Work Package Leader