# Personalised medicine against cancer and viruses

Modelling cells and simulating how they work gives a boost to personalised treatment plans. The Per-MedCoE project combines clinical patient data with data related to the operation of genes, proteins and cells. The goal is to develop tools that can be used in precision medicine. The modelling of cells in detail is, however, a major undertaking, requiring a lot of computing power by supercomputers.
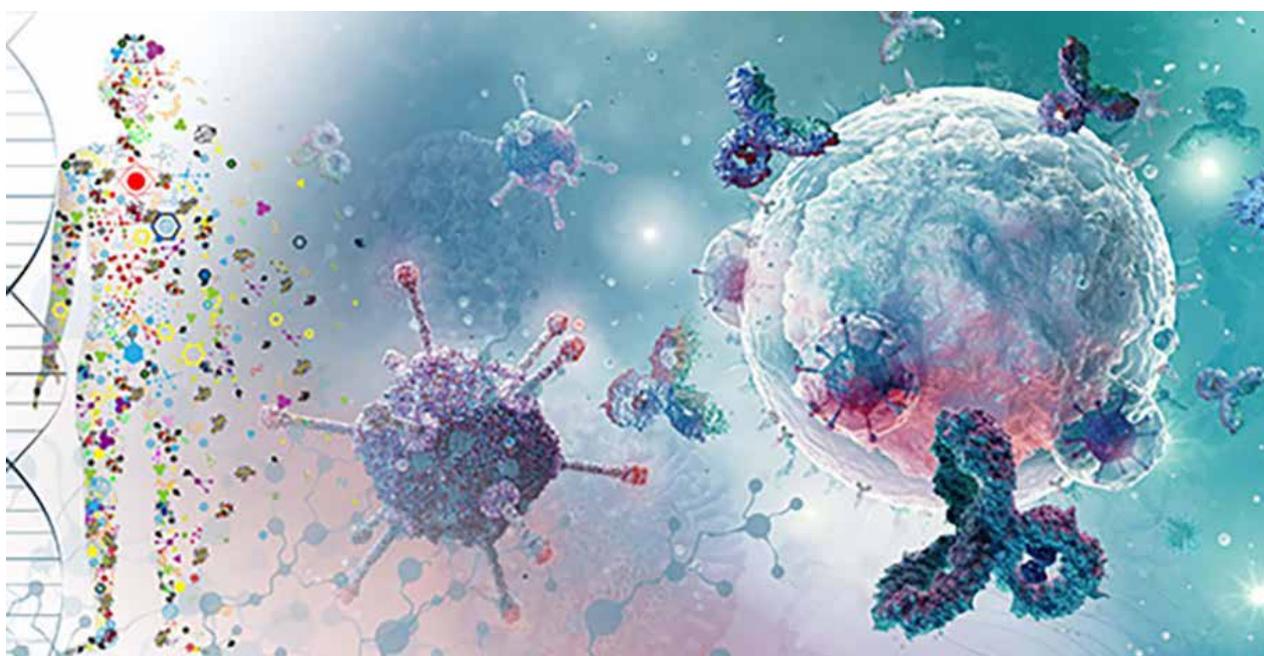
Personalised medicine will open up great opportunities in the future. The goal is to be able to combine a patient's clinical data with genetic data to create personalised treatment plans. The PerMedCoE project (HPC/Exascale Centre of Excellence in Personalised Medicine) is working to improve the compatibility of personalised medicine modelling software with next-generation exascale supercomputer systems. Their theoretical computing power is as much as 10^{18} operations per second. The project involves researchers from several European universities and hospitals, and focuses on four cellular-level modelling software systems based on open source code. Further to working on software development, the research project aims to make precision medicine tools easier to use and compatible with a number of European high-capacity computer centres.

"Our aim is that these four software systems could be used in many supercomputers," says Project Manager **Sampo Sillanpää** from CSC – Finnish IT Center for Science.

"At the moment this is technically very challenging to implement, because all high-performance computing environments are unique in terms of their system architecture."

The plan is to achieve seamless operation of software and data masses through jointly agreed technologies. In the PerMedCoE project, this is done by means of workflow software and what is known as container technology. Workflows are used to automate different steps required to analyse data. Further to the actual modelling step, a workflow may involve, for example, data pre-processing and steps relying on results produced by the modelling tool. Container technology can be used to specify a standardised environment in which scientific software is run in each high-performance computing environment taking part in the project. Once the software code and libraries and settings are placed in the container, they can be transferred from one computer to another.

"The software and data are, in a manner of speaking, packaged in a single box, making it easy to move it from one environment to another. CSC have several container technology experts, so the tools can be transferred from one platform to another," says Sillanpää.
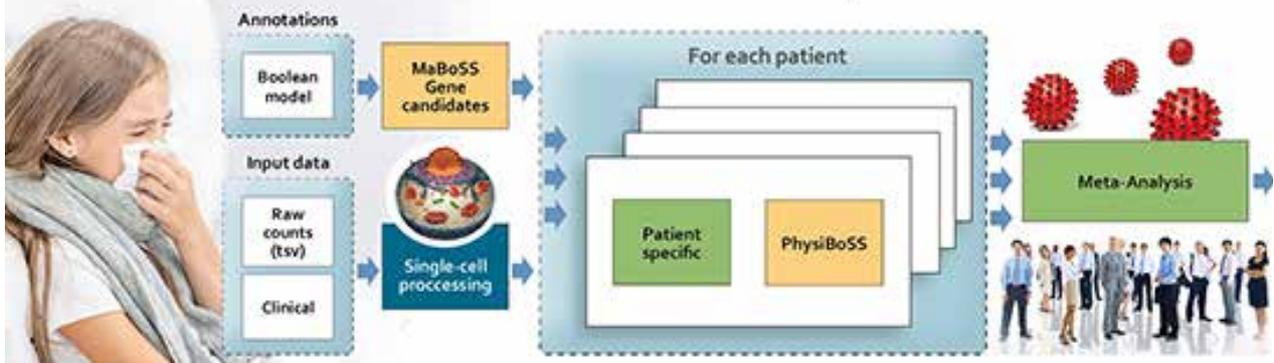
The COVID-19 disease and different cells' population behaviours are studied using multi-scale models and single-cell data. MaBoSS is software that enables to simulate populations of cells and to model stochastically (Boolean modelling) the intracellular mechanisms that are deregulated in diseases. PhysiBoSS combines MaBoSS with PhysiCell, open source software for simulating large systems of cells. 3D tissues can be studied on standard desktop computers. With PhysiBoSS scientists can analyse the effect of genetic alterations of individual cells at the population level.

"The containers enable experts to create user-friendly workflows. Workflows in the PerMedCoE project consists of several building blocks, each performing a specific precision-medicine calculation. One building block may pre-process data, a second one carrying out the actual analysis, and a final one delivering the outcome to the end user. This means that the users may not even have to know how many building blocks the automation contains, but focus on analysing the results."

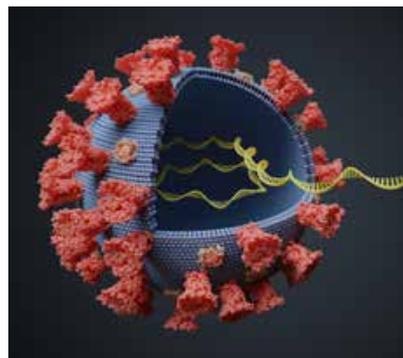## Modelling COVID-19 at the cellular level

The usefulness of technologies built in the project is assessed by means of various use cases. Workflows are used to analyse which disturbances may be caused at the cellular level by diseases and how the drugs that have been administered actually work. The models can be used to examine cellular metabolism and signal transmission.

"In PerMedCoE use cases, we make use of publicly available genomic data. Now we can study samples taken from coronavirus patients and look for markers in the genomic data to learn which patient groups are particularly susceptible to the dangerous forms of the disease."

The project models the skin's epithelial tissue that reacts to a coronavirus infection by calling various immune cells to work against the virus. This may help to identify patient groups that are susceptible to the serious form of the disease.

"The idea is that we are able to run several models simultaneously for individual patients. This results in efficient analysis of sufficiently large amounts of data, so that the modelling results can be used for personalised medicine," says Senior Data Scientist **Jesse Harrison** of CSC.

When modelling the COVID-19 use case, cellular-level RNA sequence data is used. Single-cell RNA sequencing (scRNA-seq) may reveal regular interaction between genes, cell lineages, cellular difference and the cellular frame of reference in its environment.

Another key use case of the project concerns cancer diagnostics. The goal is to create modelling tools for the prediction of cancer tumours and the development of patient-specific treatments. The material has been collected by the Wellcome Institute and the Massachusetts General Hospital Cancer Center. The database contains more than a thousand tumour-tissue cell lines.

"The project aims, among other things, to identify new drug combinations that could be useful in cancer treatment," says Jesse Harrison.

This will hopefully lead to more personalised cancer treatments and faster diagnostics.

"In order for this to become reality, closer collaboration will be needed between high-performance computing centres and medical organisations. This is because we are talking about large amounts of data, and analysing such large-scale personalised data sets is simply not possible with a desktop computer."

The results and tools of PerMedCoE are open to all researchers.

"When the project ends in summer 2023, we will have updated versions of modelling tools developed from open source code, and these will be made available to the research community. The project will also create new expertise to support the use of precision medicine tools in CSC computing environments."

Ari Turunen
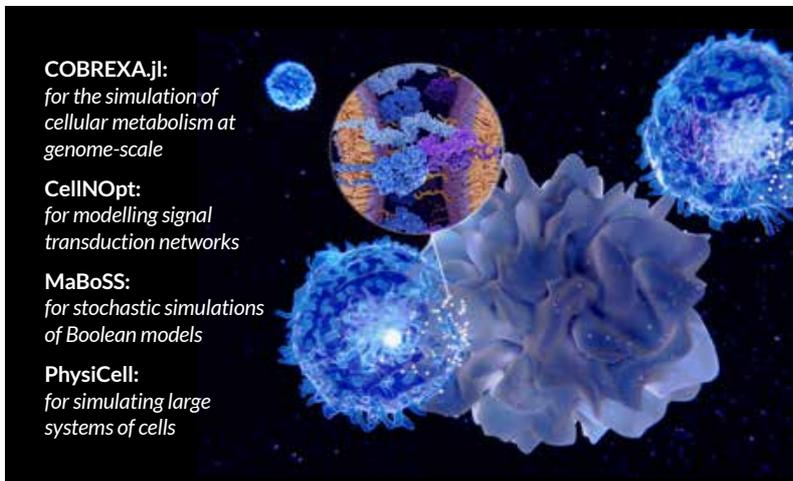
COBREXA.jl:
for the simulation of cellular metabolism at genome-scale

CellNOpt:
for modelling signal transduction networks

MaBoSS:
for stochastic simulations of Boolean models

PhysiCell:
for simulating large systems of cells

How cells work at multiple levels, ranging from the individual cells to large populations of cells? CSC – IT Center for Science and the Barcelona Supercomputing Center (BSC), alongside ten other academic and commercial players, kicked off in October 2020 a European Commission Centre of Excellence a project called HPC/Exascale Center of Excellence for Personalised Medicine (PerMedCoE). The project develops cell-level modelling software suitable for high-performance computing. Thanks to high-performance computing, biological data such as genomics and proteomics can be made part of precision medicine, because data can be analysed much faster. Diagnoses, for example, should in future be possible to make within hours or days. PerMedCoE is part of the ELIXIR Finland's development programme.

Beating cancer

The EU is funding a number of projects that will enable personalised patient treatment in the future. Cancer is an example of a disease that is extremely individualised, whether it is breast, lungs, liver or prostate cancer.

For example, the *Conquering Cancer: Mission Possible* programme under Horizon Europe will – according to **Esa Pitkänen**, researcher at FIMM (the Institute for Molecular Medicine Finland) – pave the way to future cancer research and treatments. The ambitious programme strives to understand the mechanisms that lead to cancer, to discover methods for early detection of cancers, and to achieve breakthroughs in personalised cancer medicine.

"What is common to all these goals is a versatile and comprehensive use of health data by means of new calculation methods. Artificial intelligence algorithms based on machine learning have indeed already achieved some encouraging results in terms of, for example, digital pathology. The next leaps will be made by combining various data sources in order to make individualised cancer screening and treatment recommendations," says Pitkänen confidently.

Within the programme, cancer patients become active participants in cancer treatment development, for example by being able to send their health data securely to researchers. This also gives the patients new research data about their own illness.

"It is important that as treatments develop, people are given an equal opportunity to benefit from new treatments regardless of their background. I am glad to see that this has been taken into account in the programme's recommendations. There is also special emphasis on cancers among children and young people."

MORE INFORMATION:

HPC/Exascale Centre of Excellence in Personalised Medicine
https://permedcoe.eu

CSC – IT Center for Science
is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.
http://www.csc.fi
https://research.csc.fi/cloud-computing

ELIXIR
builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.
http://www.elixir-finland.org
http://www.elixir-europe.org