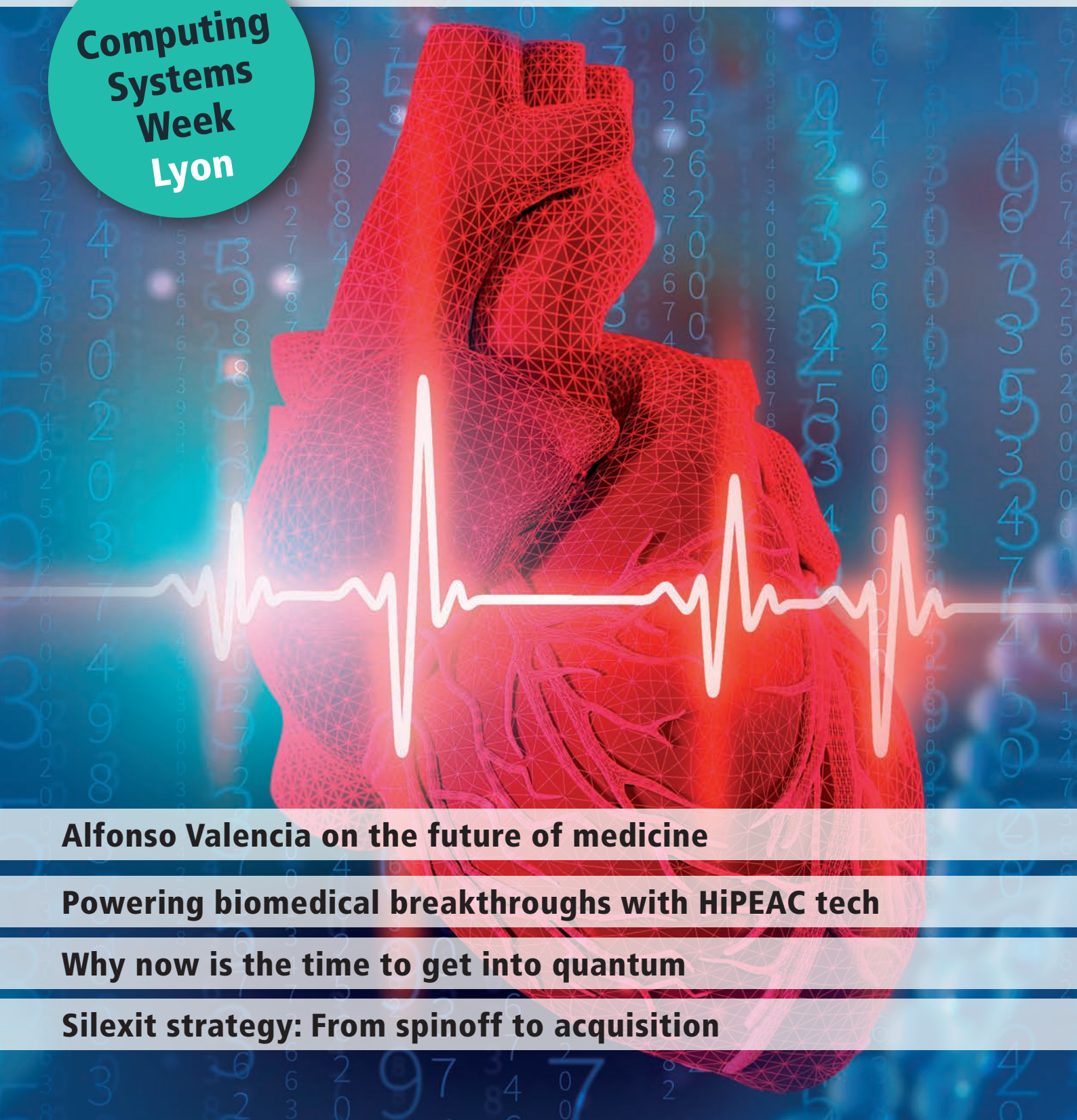
HiPEAC info 64

OCTOBER 2021

Computing Systems Week Lyon



Alfonso Valencia on the future of medicine

Powering biomedical breakthroughs with HiPEAC tech

Why now is the time to get into quantum

Silexit strategy: From spinoff to acquisition

Alfonso Valencia (Barcelona Supercomputing Center) is an internationally renowned researcher in the field of bioinformatics. We caught up with him to find out about personalized medicine, digital twins in biology, how artificial intelligence is revolutionizing biomedicine and more.

"For the first time, we have real biological data with which we can create digital twins"

What exactly does personalized medicine involve?

A topic which has been spoken about for some years now, personalized medicine basically encompasses techniques of genomic, precision and data-oriented medicine. Thanks to advances in medical instruments and devices, we can now measure things that we were unable to measure in the past. This means that medicine is increasingly being practised with more information and more precision, and thereby can become more and more tailored to people's individual characteristics.

As an example, many drugs currently used to treat disease have never been tested on a specific demographic group, such as older people. This could have serious consequences, given that our metabolism changes as we age.

Personalized medicine also provides a more holistic way to investigate different therapies. For instance, the most common way of approaching cancer research until now was to study each type of cancer individually. Researchers chose a particular type of cancer, identified a drug, and undertook a large number of clinical trials.

With personalized medicine, however, researchers can probe how the cancer actually develops and focus on the mutations which take place. This has led to the discovery that the mutations in one type of cancer – and therefore the drugs used to treat it effectively – may be the same as those in another, as was the case in a study where breast cancer was treated with drugs originally intended for pancreatic cancer.

By creating a highly personalized patient profile, it is possible to find the most suitable treatment for that patient, with applications beyond cancer. There are hereditary diseases, for example, where some patients experience very severe symptoms while others have minor symptoms. Personalized medicine allows us to better understand why this happens.

How is the concept of digital twins being applied in biomedicine?

The concept of digital twins, which has been around for a long time in engineering, is now being increasingly applied within the field of personalized medicine. In engineering, digital twins are created using real data; today, for the first time, we have real biological data with which we can create something similar. It's important to note, however, that biological systems are much more complex than mechanical systems, embodying many more elements; the system of a human being is far more heterogeneous and less predictable than that of a satellite, for example.

We are still missing many parts of the puzzle, but two technologies in particular have been fundamental in enabling this development. First, the ability to sequence

"Machine learning – from natural language processing to image recognition – and cryptography are two areas of vital importance in this field"

individual cells, allowing us to view what is going on in each cell. Second, the synthesis of artificial organs, allowing us to try out different drugs and carry out experiments on organs. Of course, this implies a serious computational challenge, given the necessity of simulating billions of cells as they grow and differentiate.

Three European-funded Centres of Excellence (CoEs) are doing relevant work in this area: BioExcel, which focuses on simulations at the biomolecular level; PerMedCoE, which undertakes simulations at the cellular level; and CompBioMed, which is working on simulations of organs.

What are the main requirements of biomedical applications in terms of computing technology?

It goes without saying that serious computational horsepower is required to carry out these kinds of analysis. In general, the most important consideration for biomedical applications is the ability to process vast amounts of data; related to this, security needs to be taken extremely seriously, in order to prevent confidential patient data becoming compromised. Hence machine learning, particularly on deep neural networks – from natural language processing to image recognition – and cryptography are two areas of vital importance in this field.

As for hardware, systems used for biomedical applications are generally federated, complex, closed and secure. They need large amounts of memory and rely on accelerators such as graphics processing units (GPUs) to carry out the necessary processes. Another key consideration is the ability to carry out the computation at the most appropriate point on the compute continuum; there may be instances where it is important to process the data locally, on small devices at the edge, while simulations will likely require a large computing infrastructure.

How is artificial intelligence revolutionizing biomedical research?

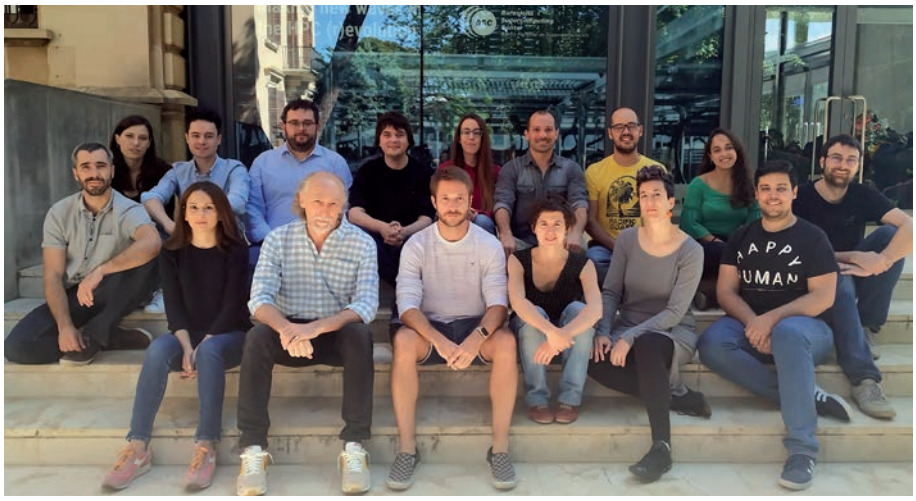
The roots of what we call 'artificial intelligence' can be traced back to research into natural language processing in the 1960s and 1970s. Over half a century later, machine-learning technologies are delivering spectacular advances, partly thanks to a change in approach. In the area of natural language processing, for example, one of the main topics investigated within my department, rather than trying to get the system to 'learn' an entire language, today's systems simply learn its structure.

A high-profile example of this shift in approach is AlphaFold, the programme developed by DeepMind which in 2020 solved a scientific problem first posed 50 years ago. Specifically, this software programme managed to predict, with very high precision, how proteins fold and form the three-dimensional structures which determine their function.

Predicting what form these strings of amino acids will take is extremely difficult, given that exploring all the different structures possible would need an impossible amount of computational time. It is hard to overstate the importance that this achievement will have; it will undoubtedly revolutionize biology research. This year, DeepMind published the algorithm – implemented on neural networks – with which they managed to solve this problem. An invention which required much imagination, the solution is partly intuitive, rather than being based on brute force alone.

While artificial intelligence is therefore driving incredible progress in biomedicine, bias is a major issue, and one which already has a problematic history in biomedical research. In the Life Sciences department at BSC, we have a group named BioInfo4Women. As well as promoting female bioinformaticians, this group examines scientific problems related to bias in biomedicine. One striking example is the fact that proportionally more women than men die of a heart attack, as women tend to exhibit different symptoms which are not as well known or understood, and they are at risk of not receiving the required attention or treatment.

Similarly, there are many well-known examples of bias in artificial intelligence systems. In the end, the quality of the system and the trustworthiness of its results depend on the quality of the data and algorithms used to build it.



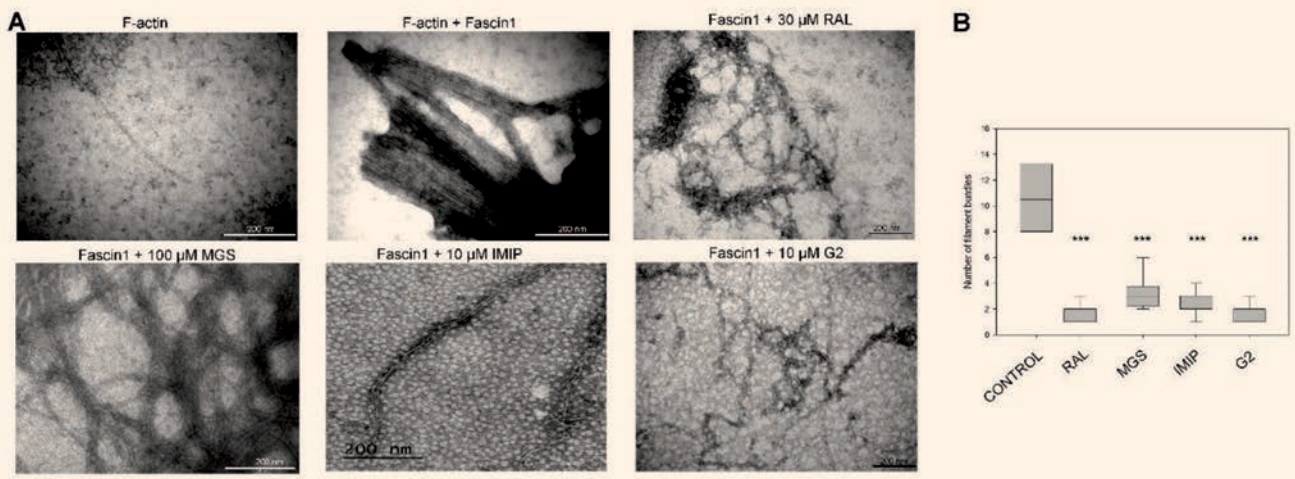
Alfonso Valencia with members of the life sciences department at Barcelona Supercomputing Center

Supercomputing infrastructure was provided by a number of HPC centres across Europe to undertake simulations and screening of the pharmacological compounds. For the HPC simulations, the workload manager Slurm was used, while Autodock Vina along with other in-house scripts and software along with singularity containers were used for the computational drug discovery aspects.

A patent for the application of this antiviral compound in the colorectal cancer context has been filed and the group is now contacting companies with a view to licensing it. The group is also investigating whether this discovery has implications for other forms of treatment. See paper in 'Further reading', below.

FURTHER INFORMATION:

B. Alburquerque-González et al. 'The FDA-Approved Antiviral Raltegravir Inhibits Fascin1-Dependent Invasion of Colorectal Tumor Cells In Vitro and In Vivo'
Cancers 2021, 13(4), 861
doi.org/10.3390/cancers13040861
Podcast episode detailing computation process for this research:
bit.ly/CDD_blind_docking_podcast



As shown in these images, the stimulation of the actin-bundle protein Facsin 1 is prevented by raltegravir (taken from Figure 4 from doi.org/10.3390/cancers13040861)

COVID-19 VIRAL MODELLING AND INVESTIGATING CANCER DRUG COMBINATIONS WITH PERMEDCOE

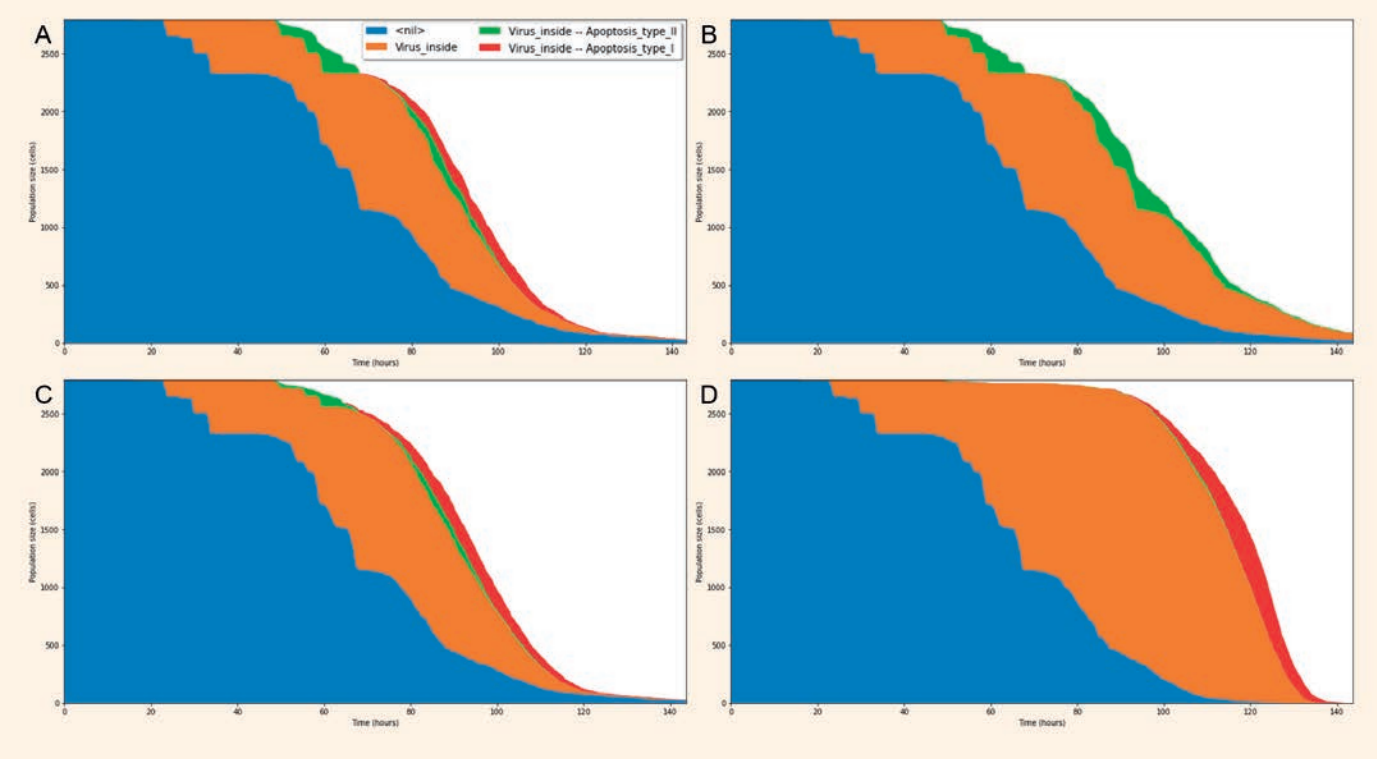
PerMedCoE, the HPC/Exascale Centre of Excellence in Personalised Medicine, aims to harness the power of exascale computing to solve problems in the field of personalized medicine. While exascale computing promises faster, more capable computers, much of these machines' potential will go unused unless current high-performance computing (HPC) software is adapted.

To avoid this, PerMedCoE is testing four software tools for use on future exascale computers. These four tools will be applied in different use cases, reflecting a broad range of computationally demanding real-life biomedical scenarios where cell-level models are used. Cell-level models study the mechanisms disrupted in cells where disease is present or when a drug is detected. These models can study perturbations on metabolism, signalling pathways and cell populations, among other areas. We caught up with Arnau Montagud and José Carbonell (Barcelona Supercomputing Center) to find out more.

Drug combinations for cancer treatment

'As is well known, cancer drugs often result in unpleasant side effects, notably from the high concentrations of the drugs used,' says Arnau. 'If better combinations of drugs were found, this could allow drug doses to be reduced and resistances to be avoided.'

To contribute to this field, PerMedCoE is studying the effects of different drug combinations and concentrations in the simulation of cell-line specific models. 'Rather than taking generic signalling models found in online resources such as the Atlas of Cancer Signaling Network (ACSN), cell-line specific models capture the specificities of a given cell line and how drugs interact with it,' explains Arnau.



Simulation of mutants and heterogeneous cell populations. The PerMedCoE framework can simulate wild type cells (A) and study how knockouts, such as FADDs, alter the model's behaviour (B). It also allows the study of different proportions of cells with an inhibited virus' M protein, such as 50% of M knockout cells (C) and 95% of M knockout cells (D)

Eventually, the aim is to integrate the results obtained by the BioExcel Centre of Excellence. With a focus on biomolecular simulations, BioExcel aims to identify and repurpose drugs for different diseases.

Multi-scale modelling of the COVID-19 virus and patient tissue

'Thanks to tremendous, global efforts, we now have vaccines to help prevent infection by COVID-19,' says José. 'However, once a patient is infected, we are still struggling to find the right drugs for treatment.'

Another PerMedCoE use case aims to help solve this by studying the COVID-19 disease using multi-scale models and single-cell data. 'Multi-scale models consider different time and space scales. Our multi-scale model takes a lung epithelial cell layer – that is, the layer of cells lining the inner space of lungs – that responds to viral infection through signalling pathways. The model allows us to organize the available knowledge on molecular mechanisms, identify biomarkers and propose therapeutic targets,' explains José.

In addition, the project aims to use single-cell data from patients to personalize these models. 'Single-cell granularity allows researchers to study the diversity of different cells, their various maturation stages and how these affect their signalling and behaviour,' José notes. 'By integrating these data, we hope to help explain why different patients experience varying levels of severity of the disease.'

Scaling up software for cell-level simulations

One of the key challenges within the project is to take existing tools, developed in a range of programming languages, and port them efficiently to an HPC environment. 'A good example is our work on PhysiCell, which demonstrates the PerMedCoE approach,' says Arnau. 'PhysiCell is a physics-based tool designed to simulate the evolution and dynamics of a heterogeneous cell population – think of a tumour, for example. It uses an agent-based paradigm to model cell-specific behaviours, allowing us to simulate a wide variety of real-life scenarios.'

Written in C++, PhysiCell will need to be refactored to use the message passing interface (MPI) protocol so that the computations and memory can be distributed over different machines. 'Our strategy for PhysiCell focuses on dividing the physical space where cells are located in subdomains, each of which is managed by an individual MPI node. This allows us to scale up individual simulations significantly and achieve simulations of billions of cells,' adds José.

Find out more on the PerMedCoE website: permedcoe.eu

PerMedCoE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 951773.