HPC workflow compartmentalises a multiscale simulation aimed at the personalisation of COVID-19 treatments and facilitates its scalabilty

Vincent Noël[1,*], Javier Conejero[2,*], Jose Carbonell[2], Miguel Vázquez[2], Gaurav Saxena[2], Miguel Ponce de Leon[2], Thalia Diniaco[2], David Vicente[2], Emmanuel Barillot[1], Laurence Calzone[1], Rosa M Badia[2], Alfonso Valencia[2,3,$] and Arnau Montagud[2,$]

1 Institut Curie, INSERM U900, Mines ParisTech, University PSL, Paris, France
2 Barcelona Supercomputing Center, Barcelona, Spain
3 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
(*) co-first authors
($) co-last authors

As a result of the unprecedented COVID-19 pandemic, researchers now have a vast amount of new publications that cover different clinical and molecular aspects of the disease. In spite of this, many underlying molecular events that lead SARS-CoV-2 infection to the observed clinical traits of COVID-19 remain unclear, especially the multiple mechanisms across different space and time scales that are involved through a set of complex interactions.

Multiscale modelling frameworks prove useful in integrating such mechanisms, as in the study of viral infection, human host cell demise and immune cells response. However, these frameworks usually lack the potential to consider many use cases simultaneously in workflows. Thus, we have used PyCOMPSs (Tejedor *et al.*, 2017) to orchestrate the simulation of thousands of patients in scalable workflows that use Singularity containers (Kurtzer *et al.*, 2017) as building blocks.

We used our own PhysiBoSS, which integrates an agent-based tool to simulate cells (PhysiCell (Ghaffarizadeh *et al.*, 2018)), a time-continuous Markovian simulator for intracellular models (MaBoSS (Stoll *et al.*, 2017)) and an efficient MPI implementation (Saxena *et al.*, 2021). PhysiBoSS was used on top of the PC4COVID initiative (Getz *et al.*, 2020) that studies the dissemination of the SARS-CoV-2 virus on lung epithelia. We incorporate cell- and pathway-specific Boolean models to detail the interactions of virus and human cells from the COVID-19 Disease Map initiative (Ostaszewski *et al.*, 2020), which integrates and formalises mechanistic knowledge using current systems biology standards.

We devised a workflow that compartmentalises the tasks in building blocks, such as analysing single-cell data using the seurat R package, using single-cell data to tailor the Boolean models to their cell types, identifying interesting interventions, running the simulations and evaluating their simulations' results. This workflow was run for each patient for up to a total of 2000 patients and in replicates to take care of the intrinsic stochasticity of the modelling tools.

We tested the scalability of this workflow in MareNostrum4 with notable results. The strong scaling is dependent on the scaling of the different building blocks considered. The weak scaling works very well for the parallel building blocks, and the losses are constrained to the few linear parts present in the workflow.

Finally, this work is part of the European HPC/Exascale Centre of Excellence in Personalized Medicine (PerMedCoE, http://permedcoe.eu/), whose purpose is to adapt multiscale modelling to supercomputing environments and to provide an easy-to-use interface to systems biology end users.

References can be found in the attached notes.

Main domain: Research on COVID-19
Sub-domain: Multiscale simulations