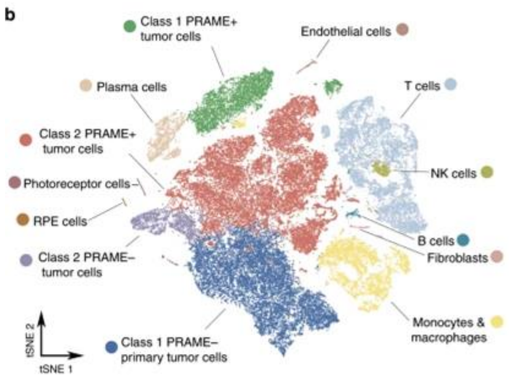
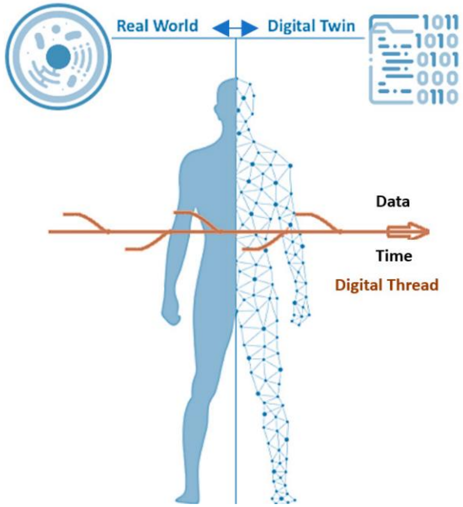
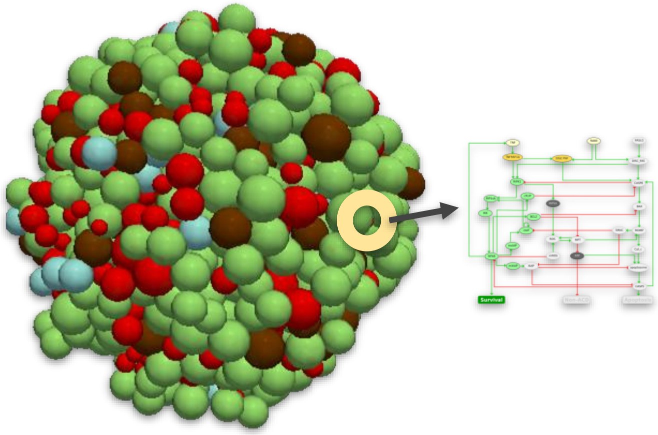
Computational challenges in the future of Molecular Biology and Personalised Medicine

3B4HPC 2022





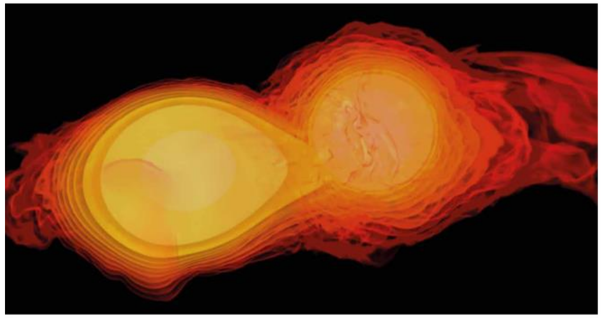


José Carbonell and Arnau Montagud
Computational Biology group

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

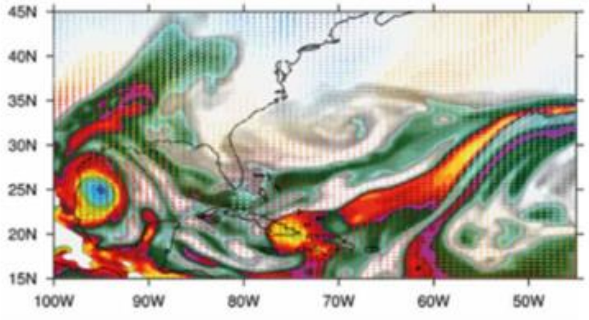
Simulations in Life Sciences

Astrophysics



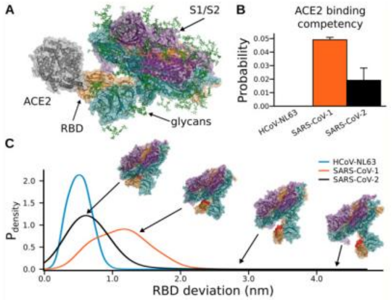
Luciano Rezzolla

Weather prediction



UPSCALE Project

COVID19

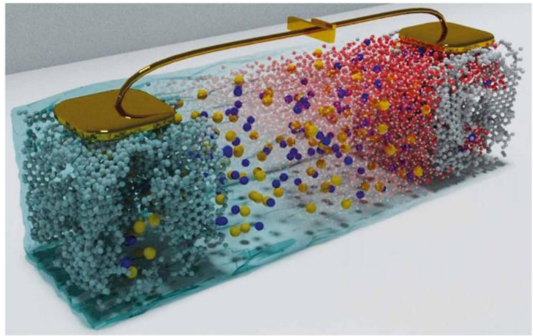


https://doi.org/10.1101/2020.06.27.175430

Energy

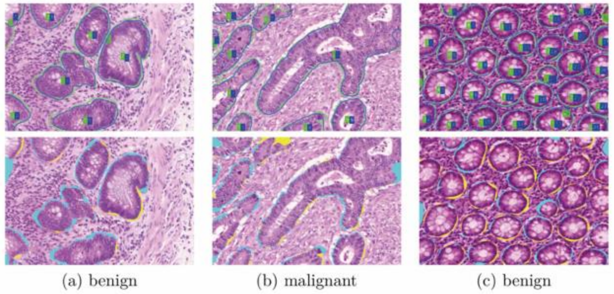
Frank Jenko, Marina Bécoulet

Materials design



Mathieu Salanne, EPFL.

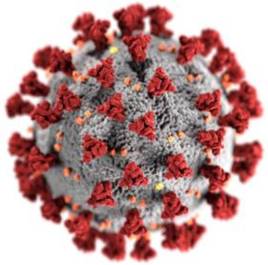
AI in biomedicine

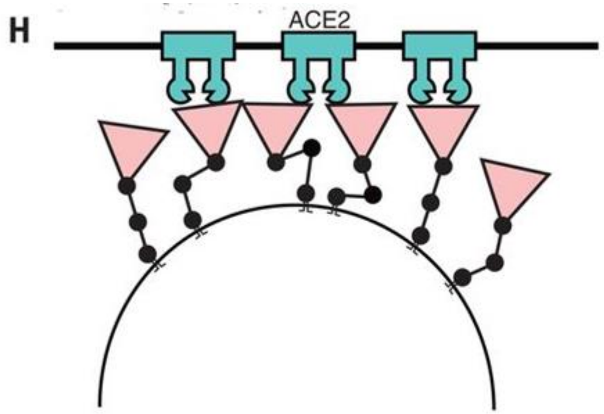


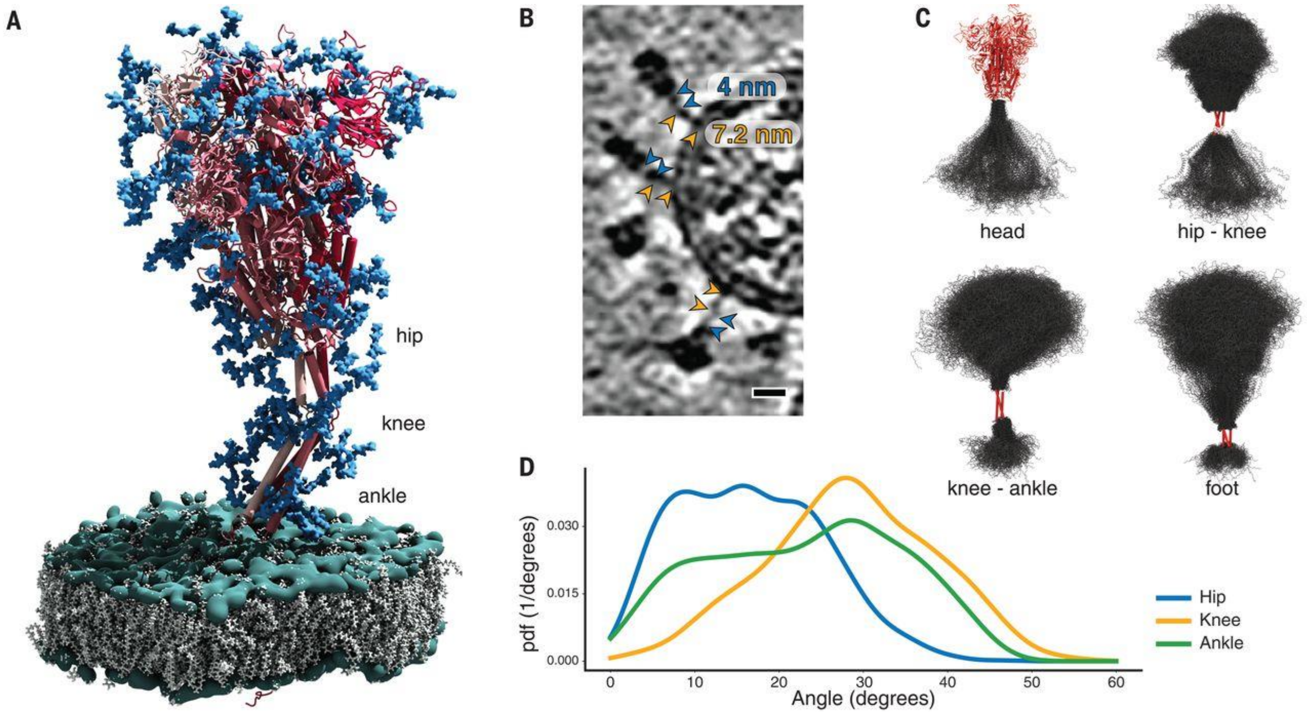
Luciano Rezzolla

Simulations in Life Sciences

Predicting the structure of SARS-CoV-2 S protein



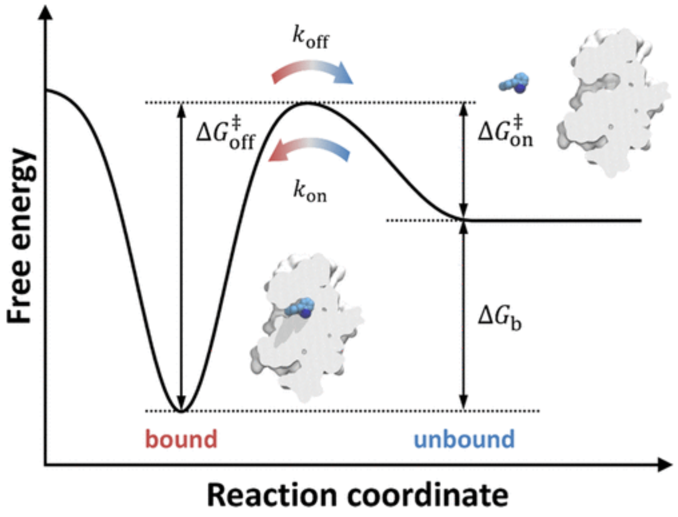


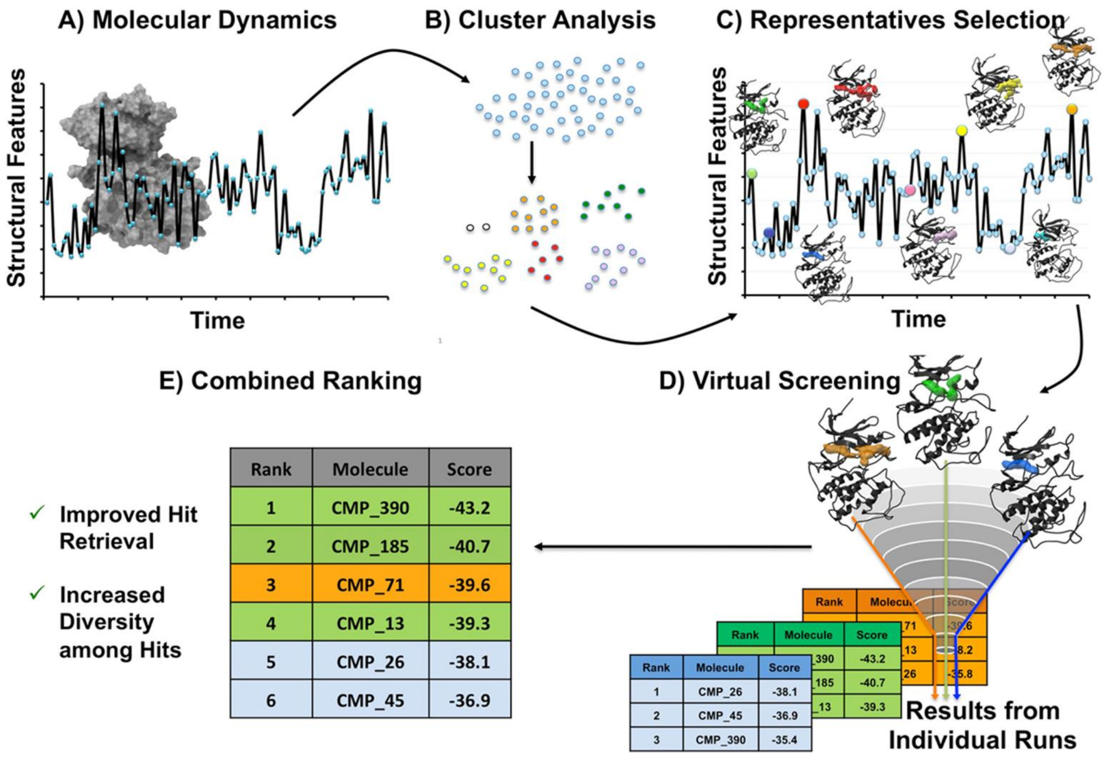


Turonˇová, B. et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science 370, 203–208 (2020)

Simulations in Life Sciences

Molecular dynamics guiding drug discovery





✓ Improved Hit Retrieval

✓ Increased Diversity among Hits

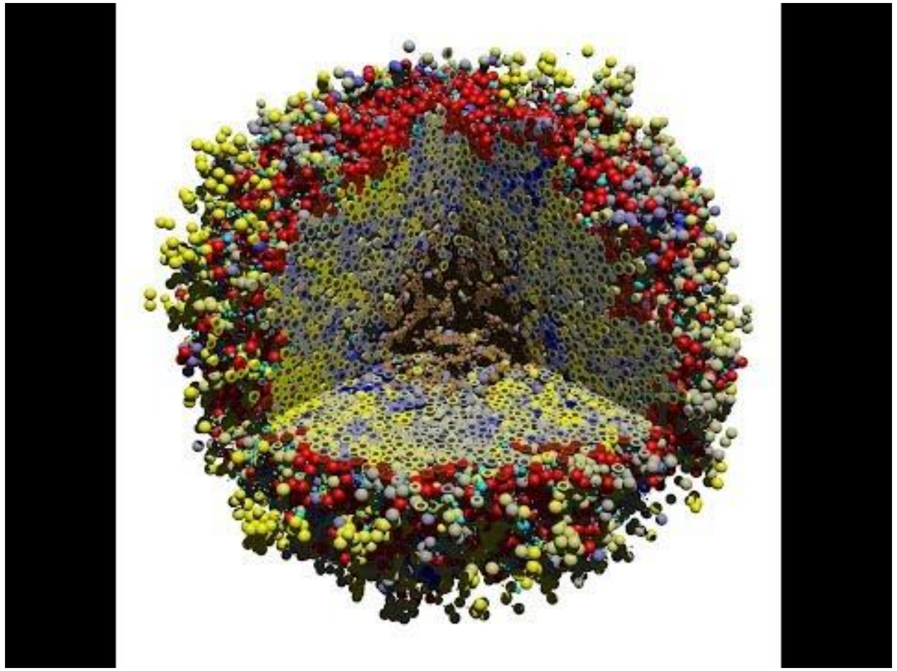
Rank	Molecule	Score
1	CMP_390	-43.2
2	CMP_185	-40.7
3	CMP_71	-39.6
4	CMP_13	-39.3
5	CMP_26	-38.1
6	CMP_45	-36.9

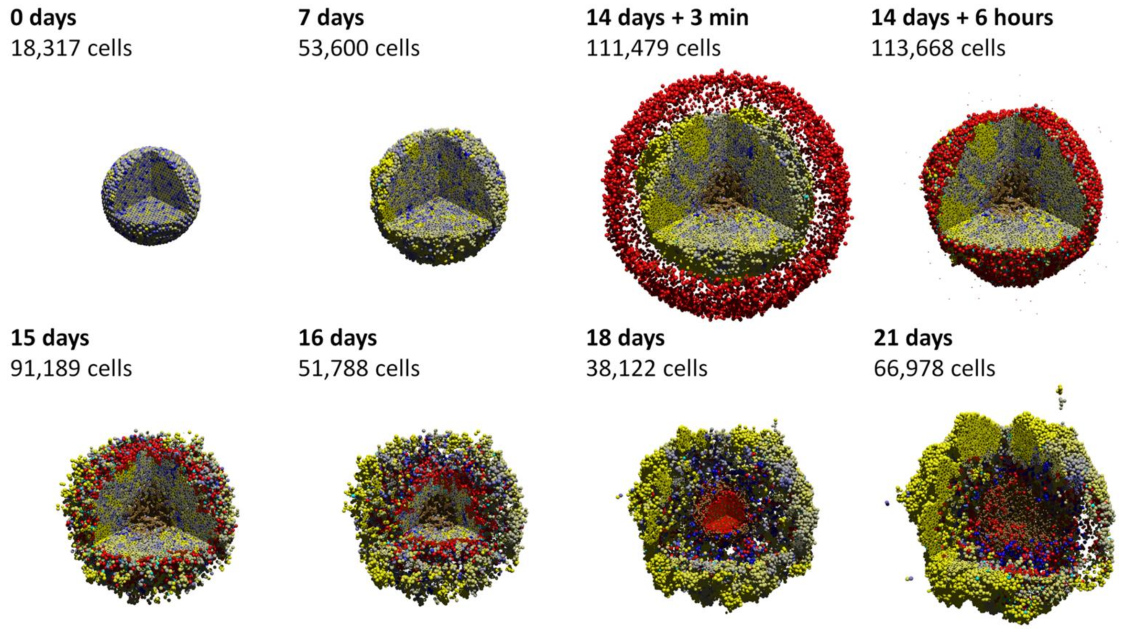
Rank	Molecule	Score
1	CMP_26	-38.1
2	CMP_45	-36.9
3	CMP_390	-35.4

Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli Journal of Medicinal Chemistry 2016 59 (9), 4035-4061

Simulations in Life Sciences

Cell-level simulations as a virtual microscope

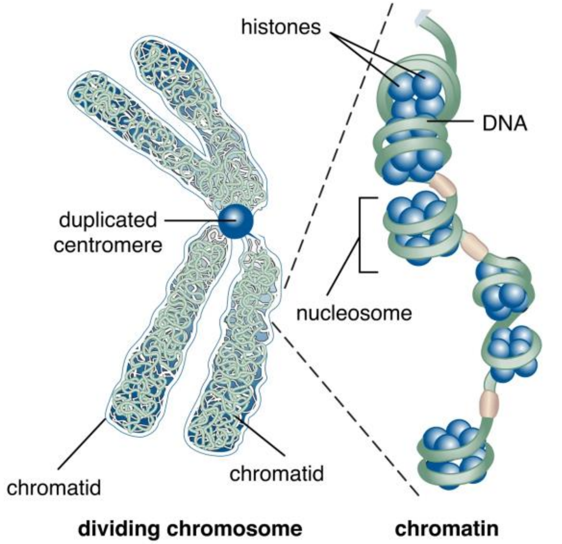




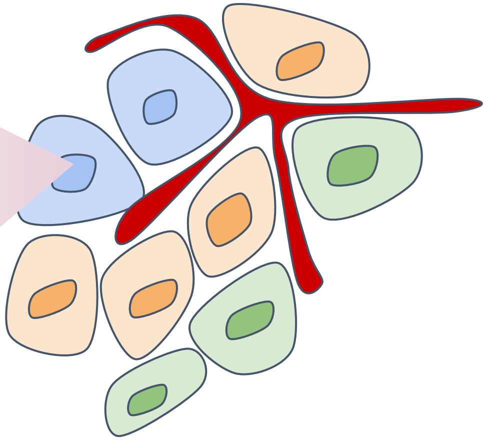
Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLoS Comput Biol. 2018

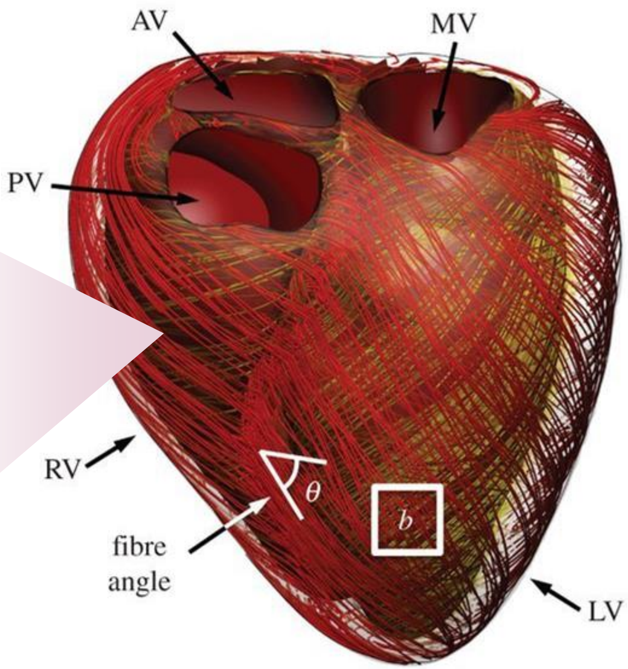
Simulations in Life Sciences

From cells to tissues and organs



© Merriam-Webster, Inc.

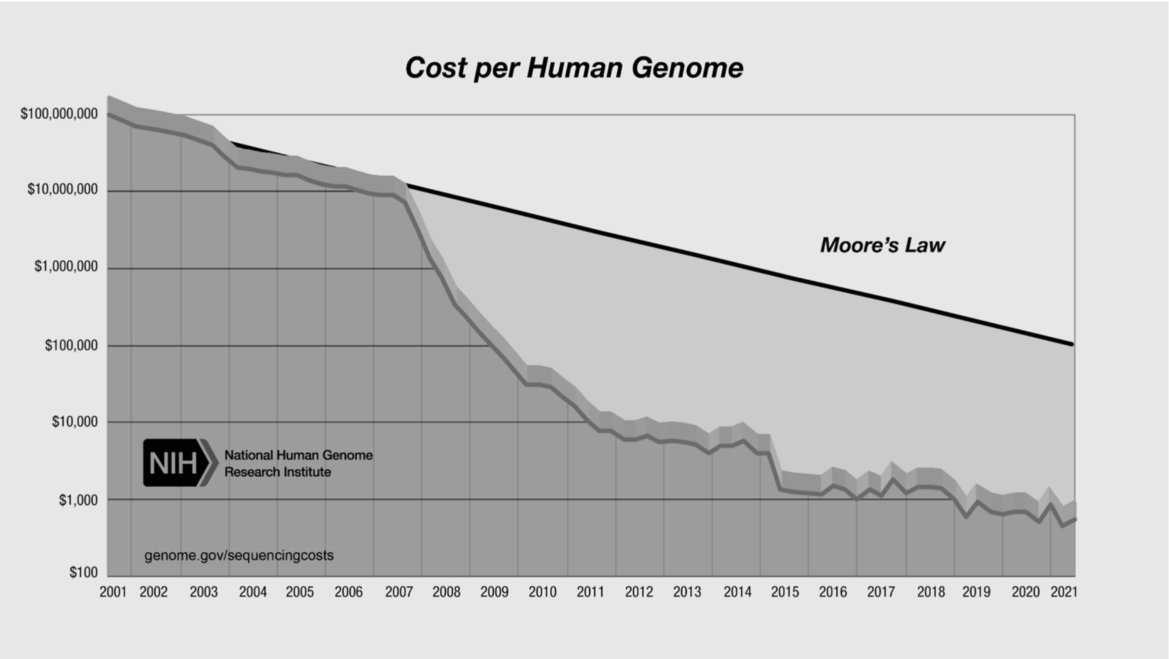


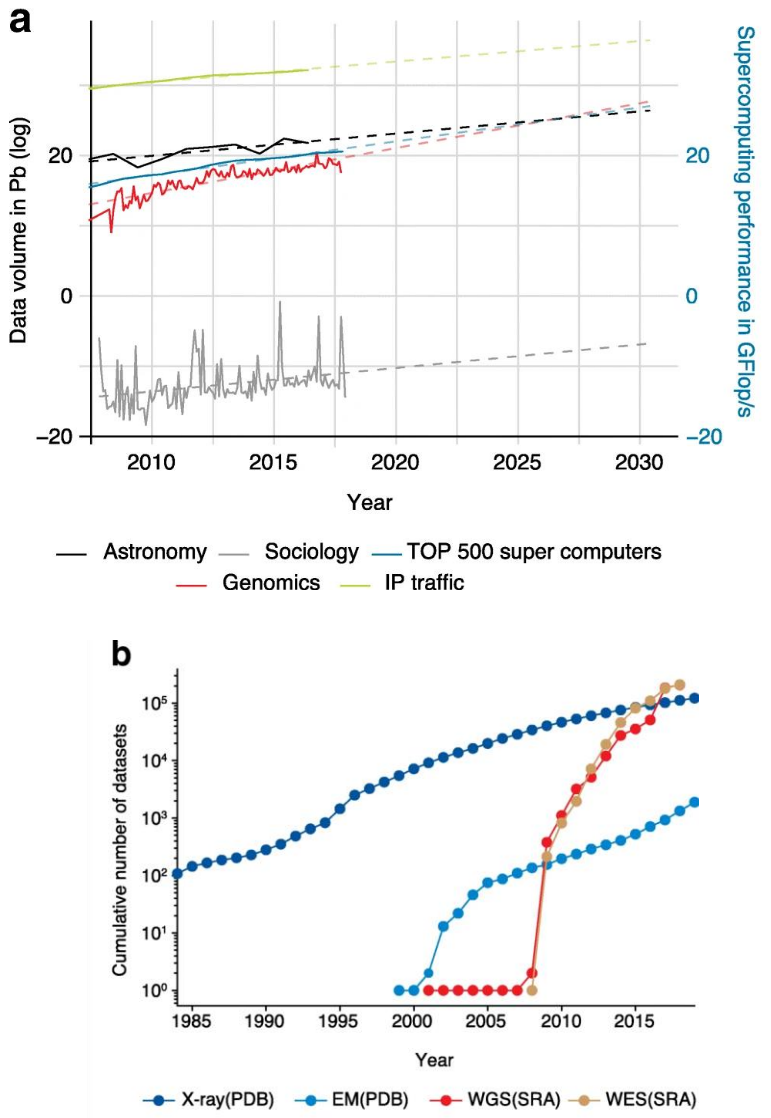


From Chabiniok R et al. Interface Focus. 2016

Towards a big data landscape

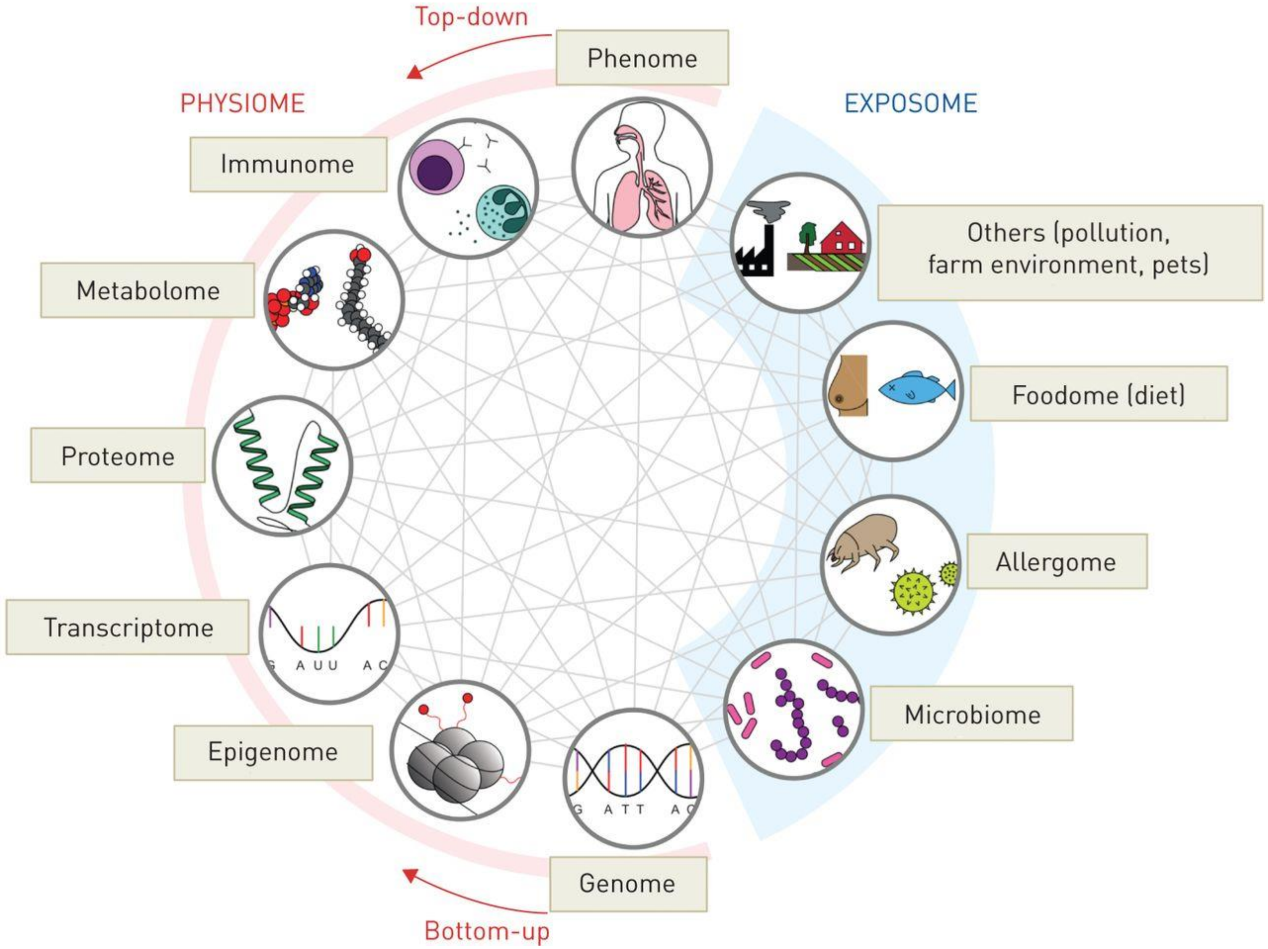
More and more data



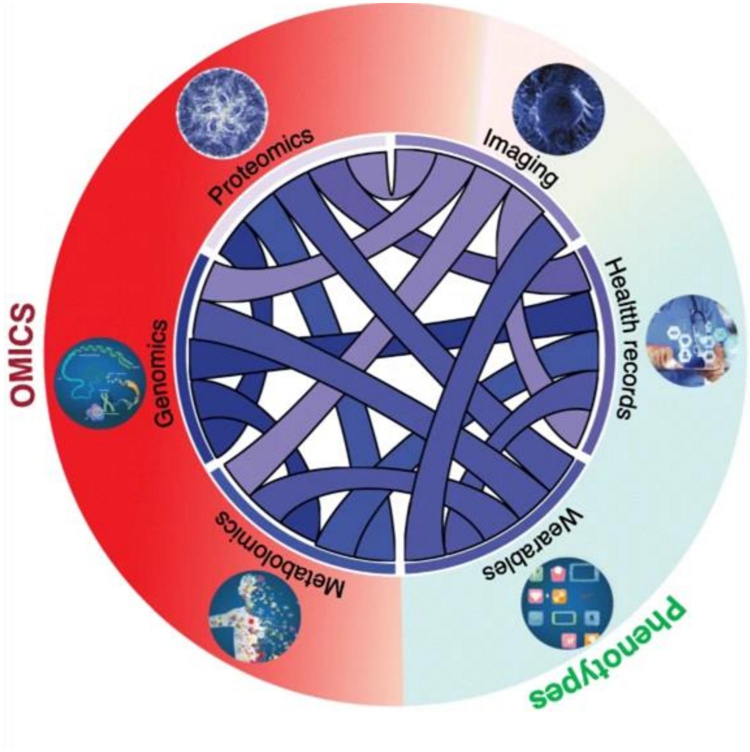


Navarro FCP et al. Genome Biol. 2019

Integration of multiple omics



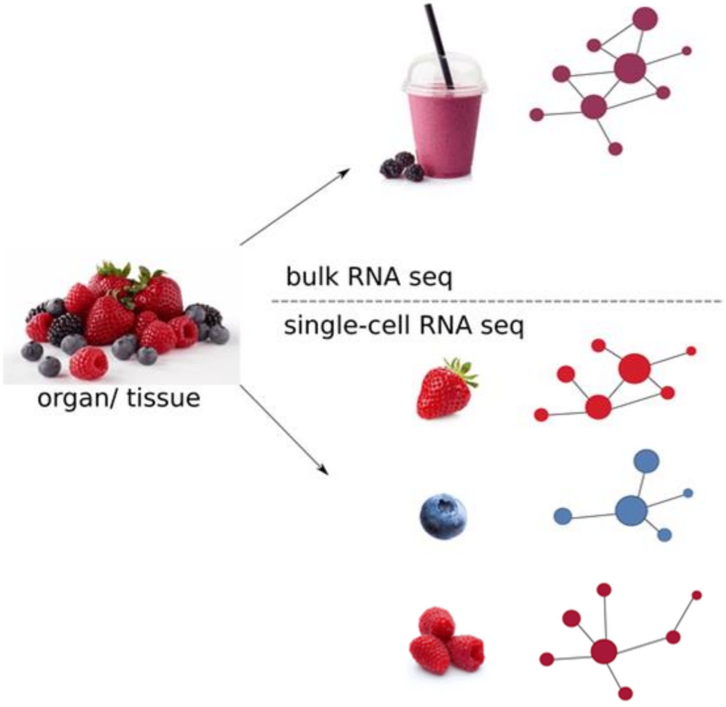
Howard H.F. Tang, et al.European Respiratory Journal 2020



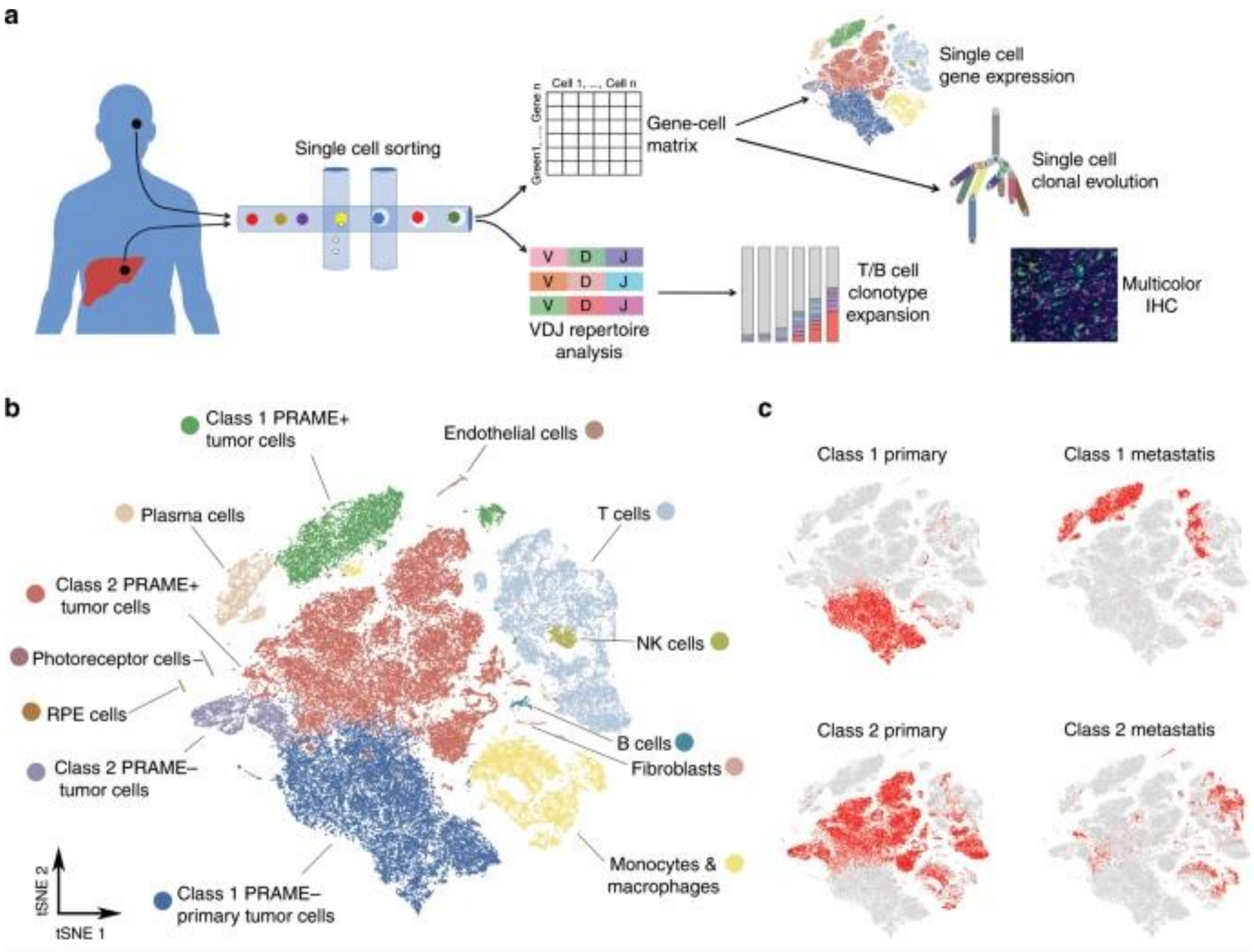
Navarro, F.C.P., Mohsen, H., Yan, C. et al. Genomics and data science: an application within an umbrella. Genome Biol 20, 109 (2019)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

From Bulk to Single-cell sequencing



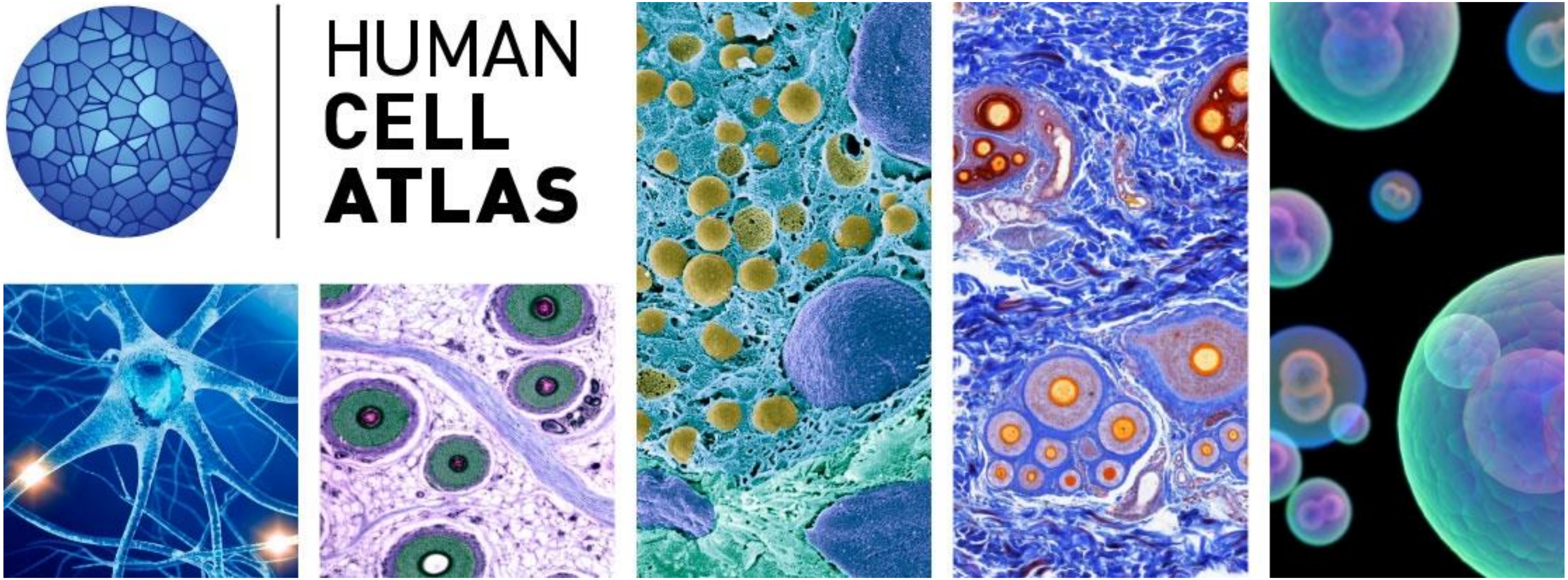
Adapted from Lisa Maria Steinheuer et al. Bioarxiv 2021



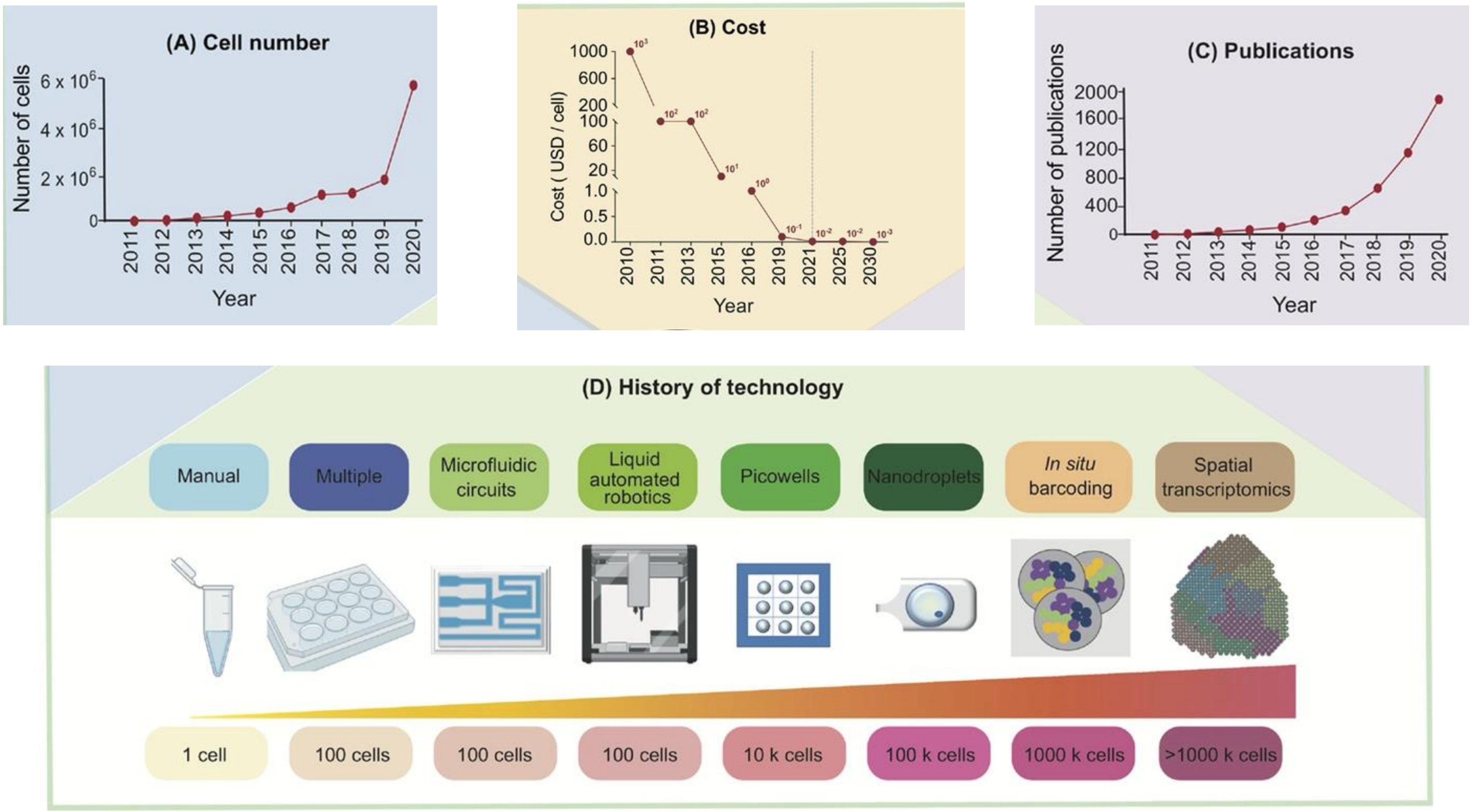
Durante MA et al. Single-cell analysis reveals new evolutionary complexity in uveal melanoma. Nat Commun. 2020

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

From Bulk to Single-cell sequencing



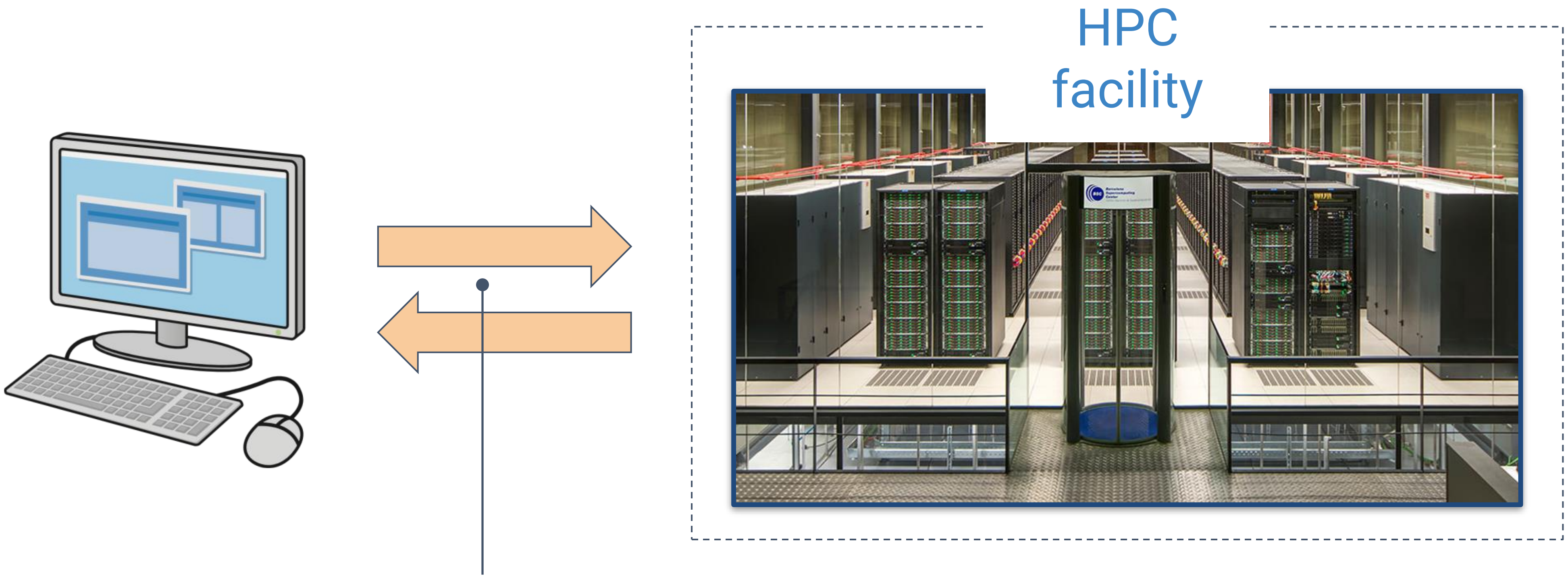
From Bulk to Single-cell sequencing



Jovic D et al. Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med. 2022 Mar;12(3):e694

From laptops to HPC centres

Designing appropriate user interfaces



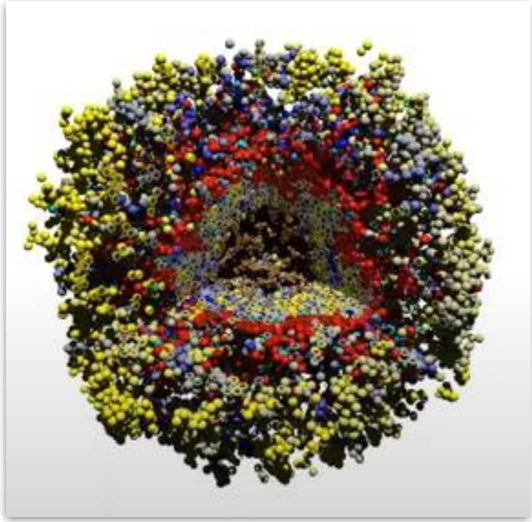
HPC facility

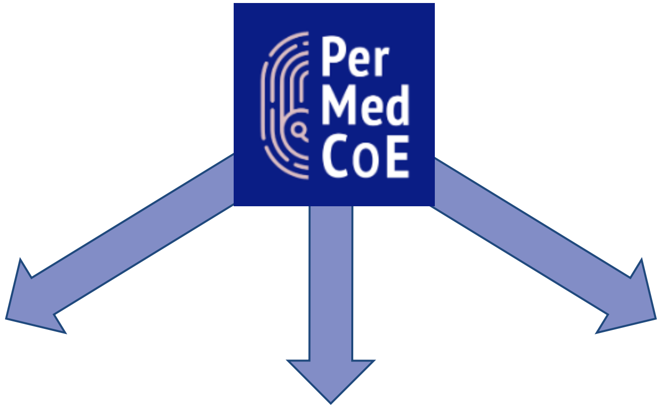
User interface (profile-based)

From laptops to HPC centres

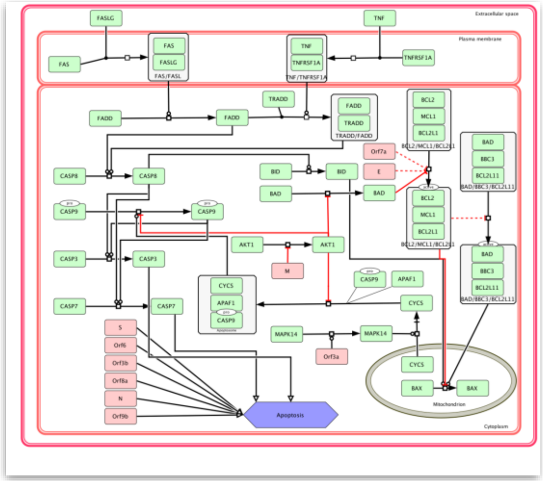
Towards an ecosystem of tools adapted to HPC environments

Cell level simulations



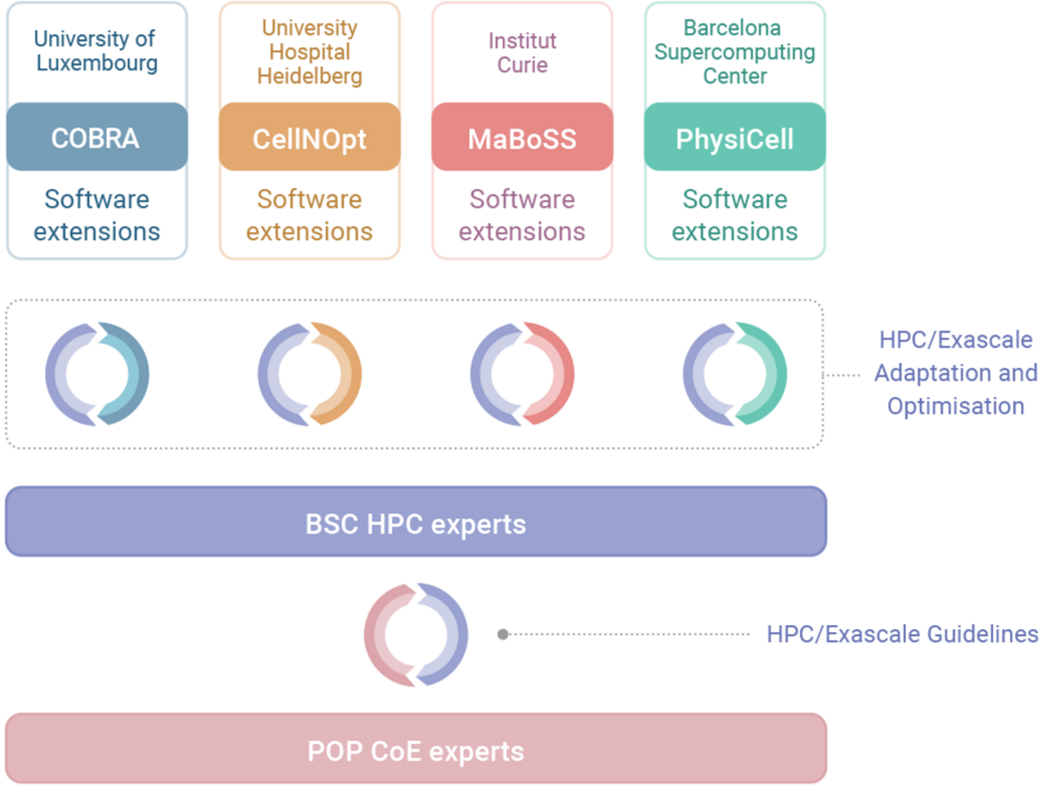


Molecular Pathways



HPC

From laptops to HPC centres



University of Luxembourg — COBRA — Software extensions

University Hospital Heidelberg — CellNOpt — Software extensions

Institut Curie — MaBoSS — Software extensions

Barcelona Supercomputing Center — PhysiCell — Software extensions

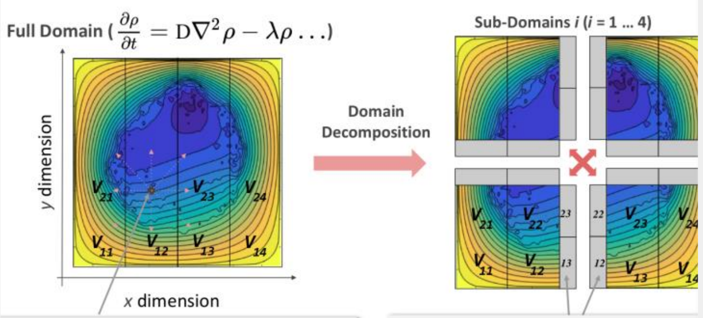
HPC/Exascale Adaptation and Optimisation

BSC HPC experts

HPC/Exascale Guidelines

POP CoE experts

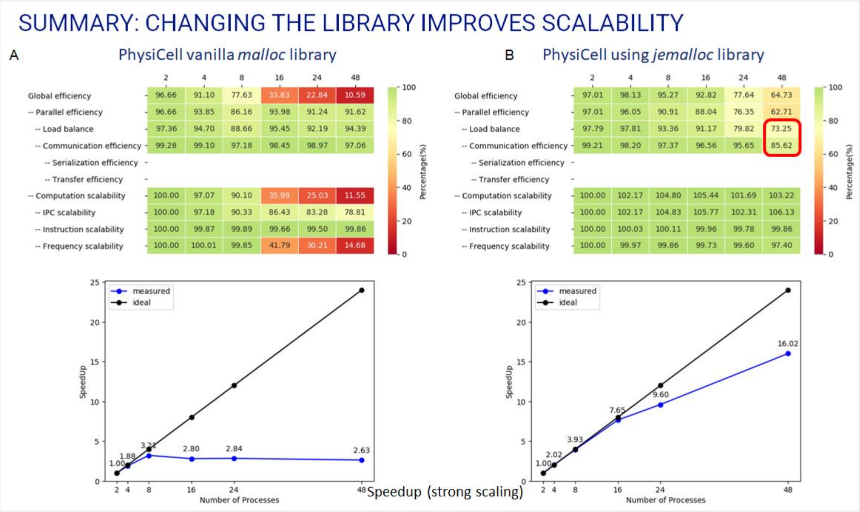
Roadmap to scalability



Full Domain ($\frac{\partial \rho}{\partial t} = \mathrm{D}\nabla^2 \rho - \lambda \rho \ldots$)
y dimension
x dimension
V_{11} V_{12} V_{13} V_{14}
V_{21} V_{23} V_{24}
Domain Decomposition
Sub-Domains *i* (*i* = 1 ... 4)
V_{21} V_{22} 23 22 V_{23} V_{24}
V_{11} V_{12} 13 12 V_{13} V_{14}

Individual Voxel: stores the values of each molecule concentration. Connected to other voxels through Moore neighborhood (PDE solver)

Ghost (Halo) Cells: needed to update boundary voxels in a tran way. Needed to exchange information between neighbour voxe



SUMMARY: CHANGING THE LIBRARY IMPROVES SCALABILITY

A PhysiCell vanilla *malloc* library

	2	4	8	16	24	48
Global efficiency	96.66	91.10	77.63	33.83	22.84	10.59
-- Parallel efficiency	96.66	93.85	86.16	93.98	91.24	91.62
-- Load balance	97.36	94.70	88.66	95.45	92.19	94.39
-- Communication efficiency	99.28	99.10	97.18	98.45	98.97	97.06
-- Serialization efficiency						
-- Transfer efficiency						
-- Computation scalability	100.00	97.07	90.10	35.99	25.03	11.55
-- IPC scalability	100.00	97.18	90.33	86.43	83.28	78.81
-- Instruction scalability	100.00	99.87	99.89	99.66	99.50	99.86
-- Frequency scalability	100.00	100.01	99.85	41.79	30.21	14.68

B PhysiCell using *jemalloc* library

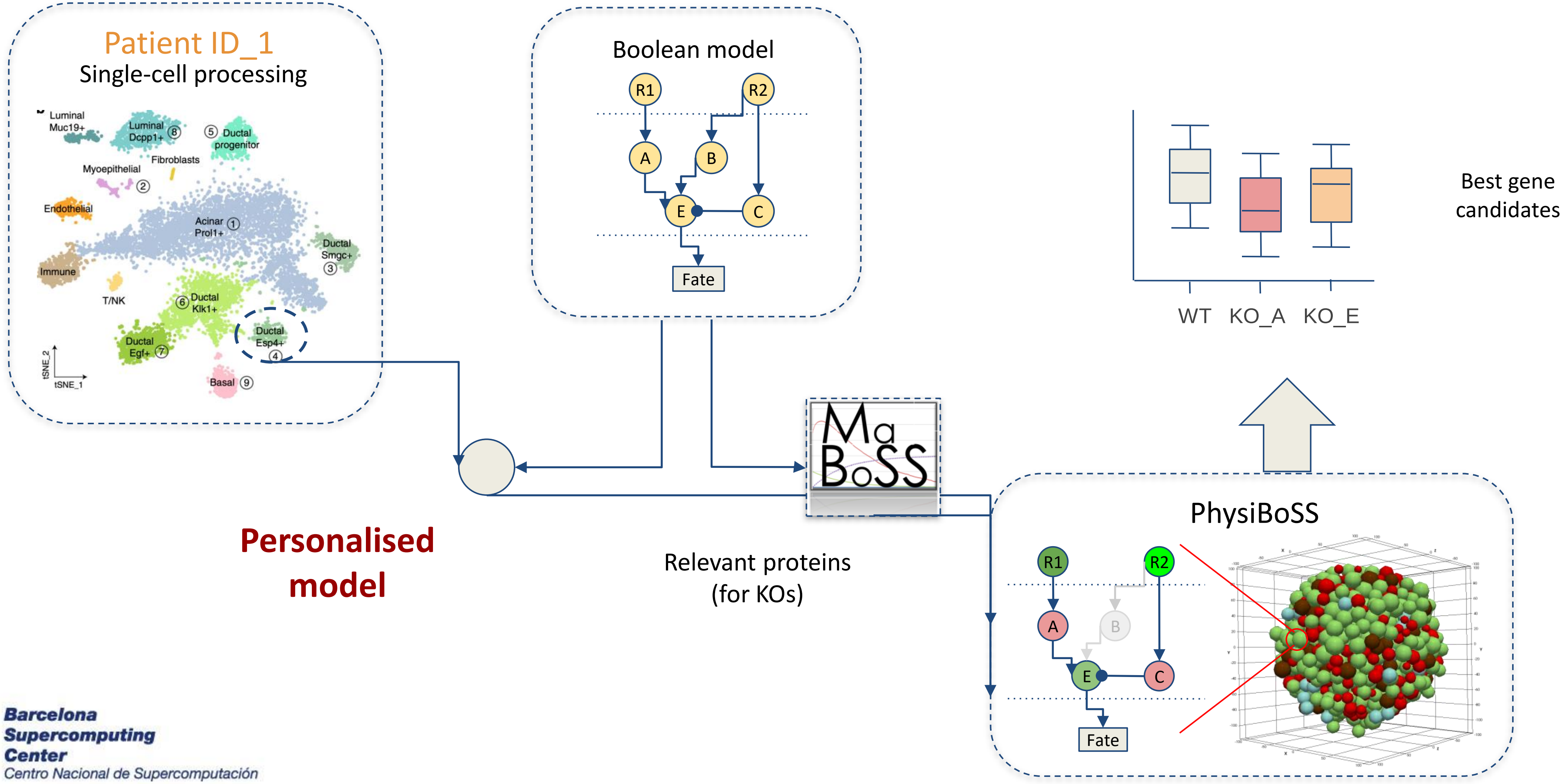
	2	4	8	16	24	48
Global efficiency	97.01	98.13	95.27	92.82	77.64	64.73
-- Parallel efficiency	97.01	96.05	90.91	88.04	76.35	62.71
-- Load balance	97.79	97.81	93.36	91.17	79.82	73.25
-- Communication efficiency	99.21	98.20	97.37	96.56	95.65	85.62
-- Serialization efficiency						
-- Transfer efficiency						
-- Computation scalability	100.00	102.17	104.80	105.44	101.69	103.22
-- IPC scalability	100.00	102.17	104.83	105.77	102.31	106.13
-- Instruction scalability	100.00	100.03	100.11	99.96	99.78	99.86
-- Frequency scalability	100.00	99.97	99.86	99.73	99.60	97.40

Speedup (strong scaling)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

From laptops to HPC centres

Solving real-life use cases



Patient ID_1

Single-cell processing

Boolean model

Personalised model

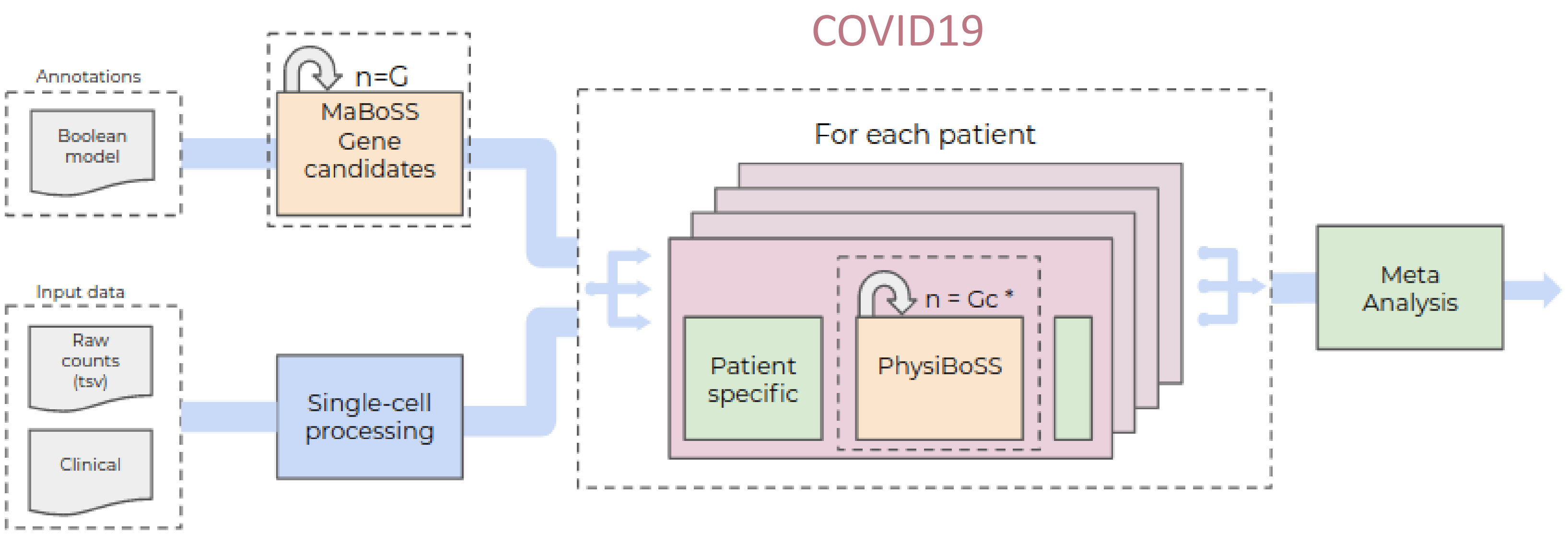
Relevant proteins (for KOs)

PhysiBoSS

Best gene candidates

Building block and workflows

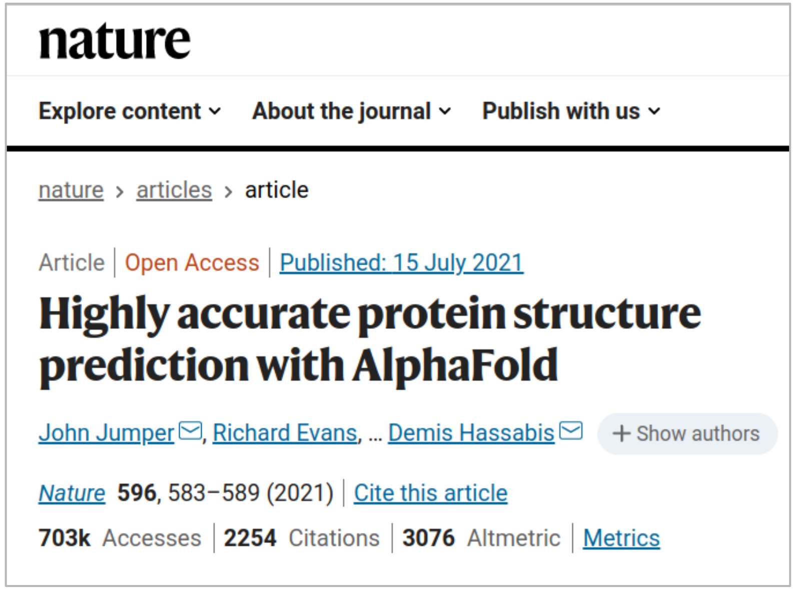
Leveraging HPC resources while improving reproducibility



apptainer.org COMPSs nextflow

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Artificial intelligence in Life Sciences



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

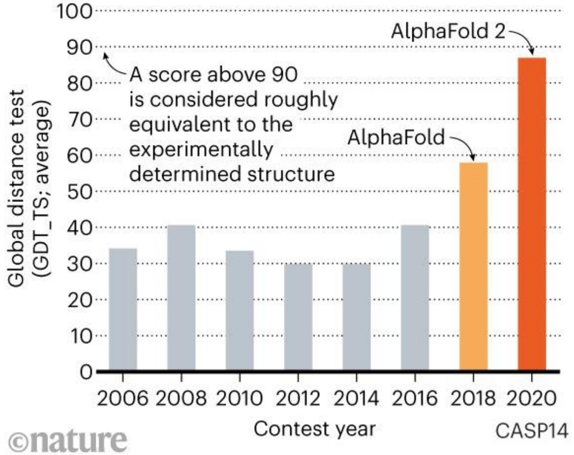
Article | Open Access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

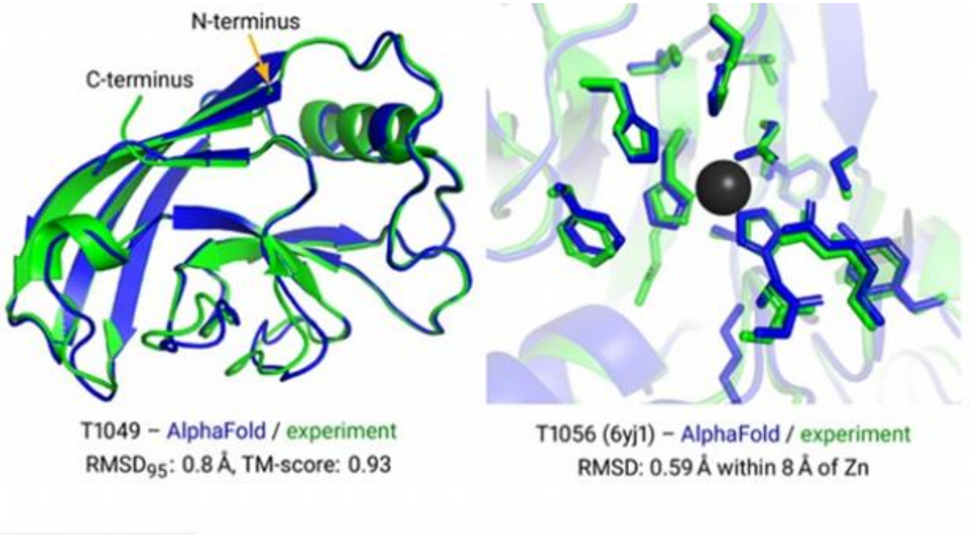
John Jumper, Richard Evans, ... Demis Hassabis + Show authors

Nature **596**, 583–589 (2021) | Cite this article

703k Accesses | **2254** Citations | **3076** Altmetric | Metrics

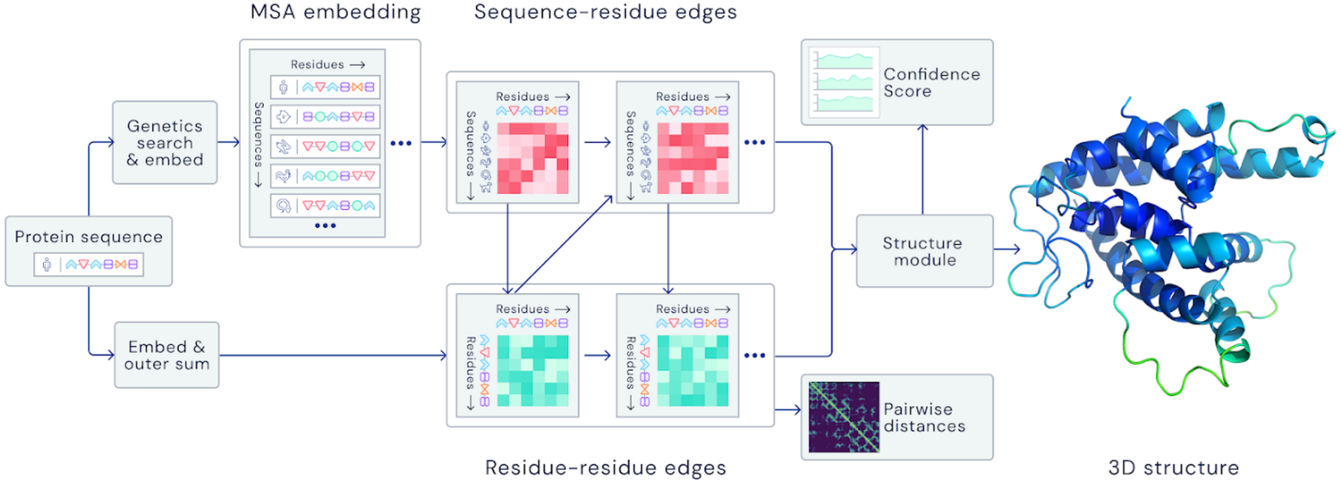


©nature



T1049 – AlphaFold / experiment
RMSD$_{95}$: 0.8 Å, TM-score: 0.93

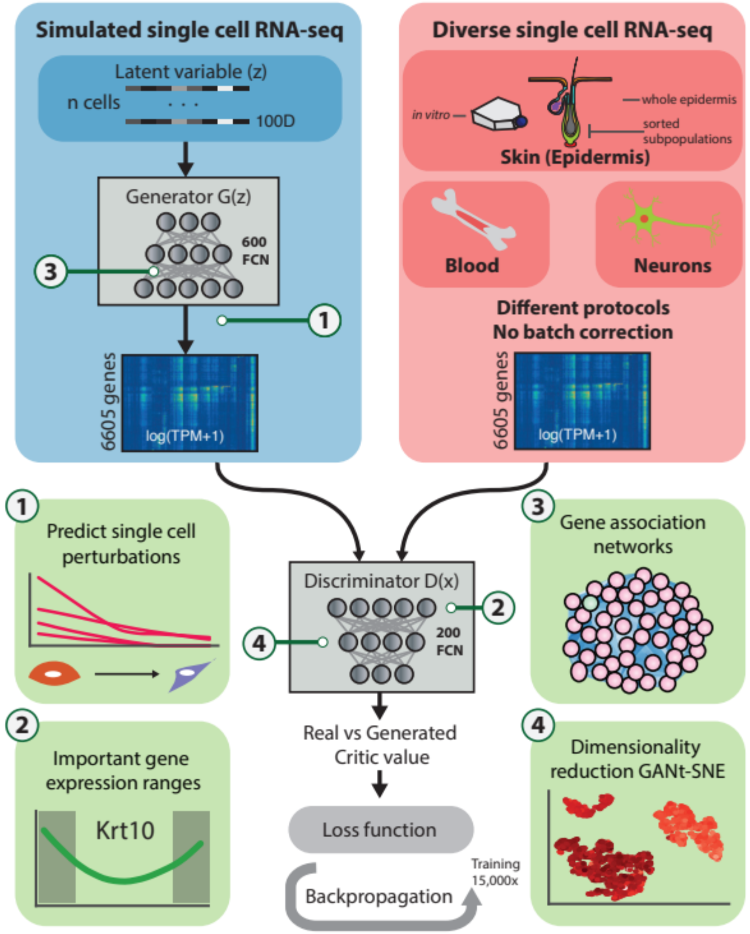
T1056 (6yj1) – AlphaFold / experiment
RMSD: 0.59 Å within 8 Å of Zn

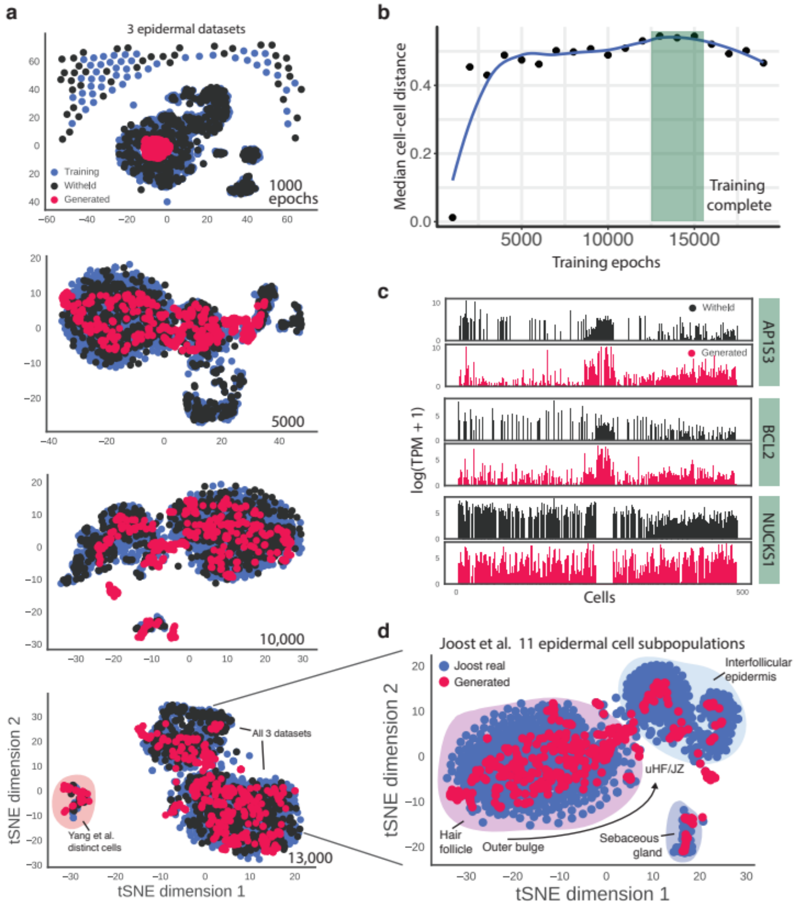


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Artificial intelligence in Life Sciences

Using Deep Learning to simulate scRNA-seq

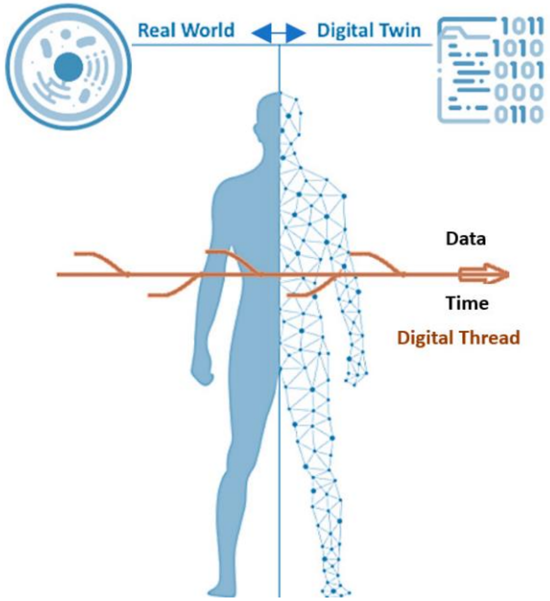


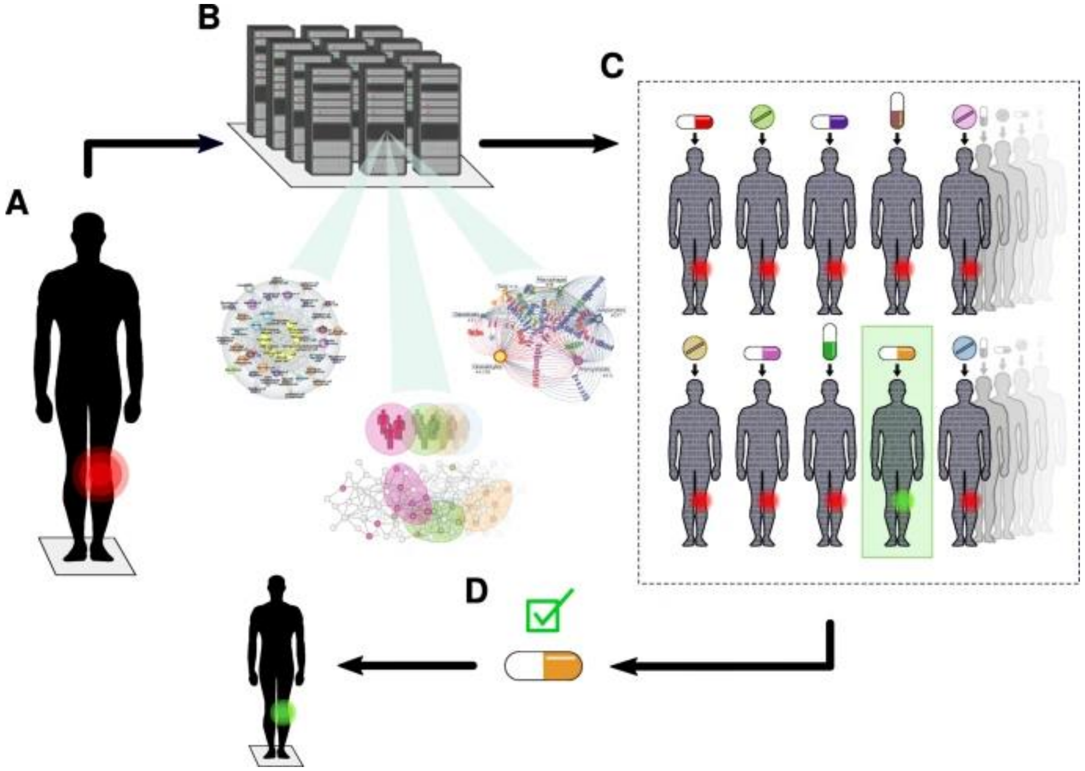


Arsham Ghahramani et al. Generative adversarial networks simulate gene expression and predict perturbations in single cells. bioRxiv 262501

Personalised Medicine and Digital Twins

Towards patient-specific treatments

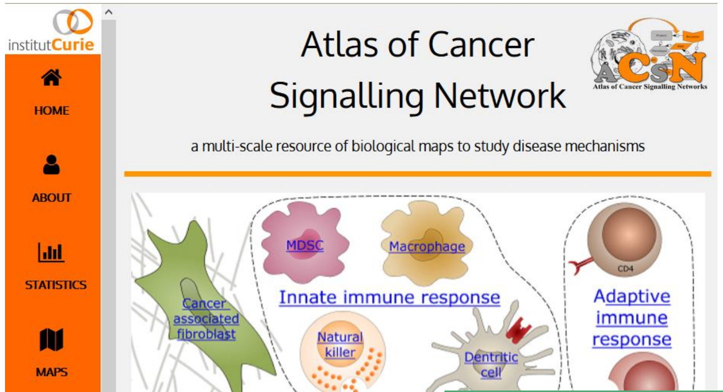




Björnsson, B., Borrebaeck, C., Elander, N. et al. Digital twins to personalize medicine. Genome Med 12, 4 (2020).

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

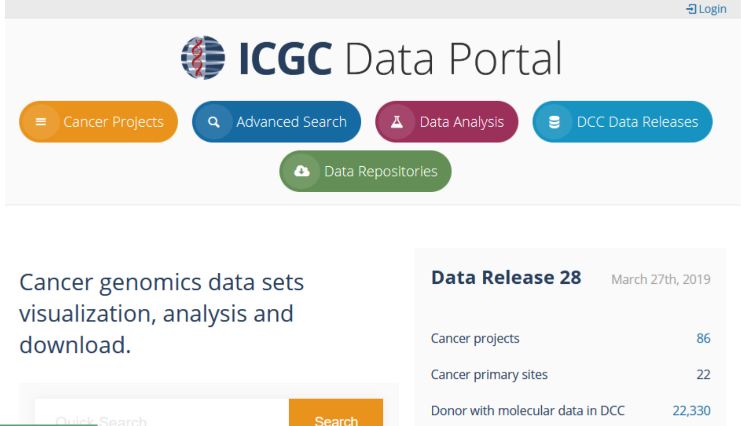
Databases where to find information



Atlas of Cancer Signalling Network

a multi-scale resource of biological maps to study disease mechanisms

ACSN was created and is maintained by the team "Com



ICGC Data Portal

Cancer genomics data sets visualization, analysis and download.

Quick Search | Search

KRAS G12D, DO35100, MU7870, FI998, apoptosis, ne Census, imatinib, GO:0016049

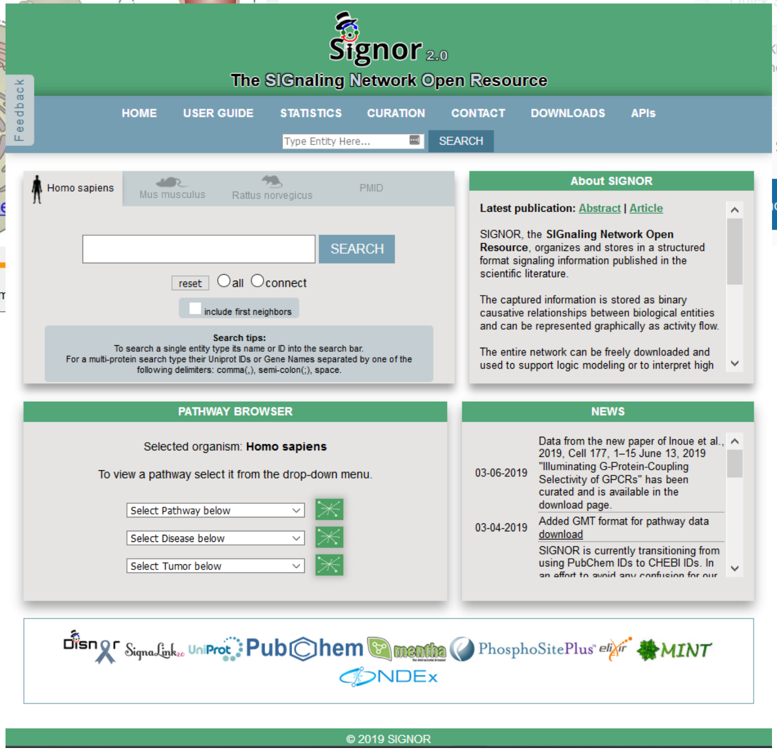
Search

ors | By genes | By mutations

Data Release 28 March 27th, 2019

Cancer projects	86
Cancer primary sites	22
Donor with molecular data in DCC	22,330
Total Donors	24,289
Simple somatic mutations	

Downlo



Signor 2.0

The SIGnaling Network Open Resource

○all ○connect

□ include first neighbors

Search tips:
To search a single entity type its name or ID into the search bar.
For a multi-protein search type their Uniprot IDs or Gene Names separated by one of the following delimiters: comma(,), semi-colon(;), space.

About SIGNOR

Latest publication: Abstract | Article

SIGNOR, the **SIGnaling Network Open Resource**, organizes and stores in a structured format signaling information published in the scientific literature.

The captured information is stored as binary causative relationships between biological entities and can be represented graphically as activity flow.

The entire network can be freely downloaded and used to support logic modeling or to interpret high

PATHWAY BROWSER

Selected organism: **Homo sapiens**

To view a pathway select it from the drop-down menu.

Select Pathway below

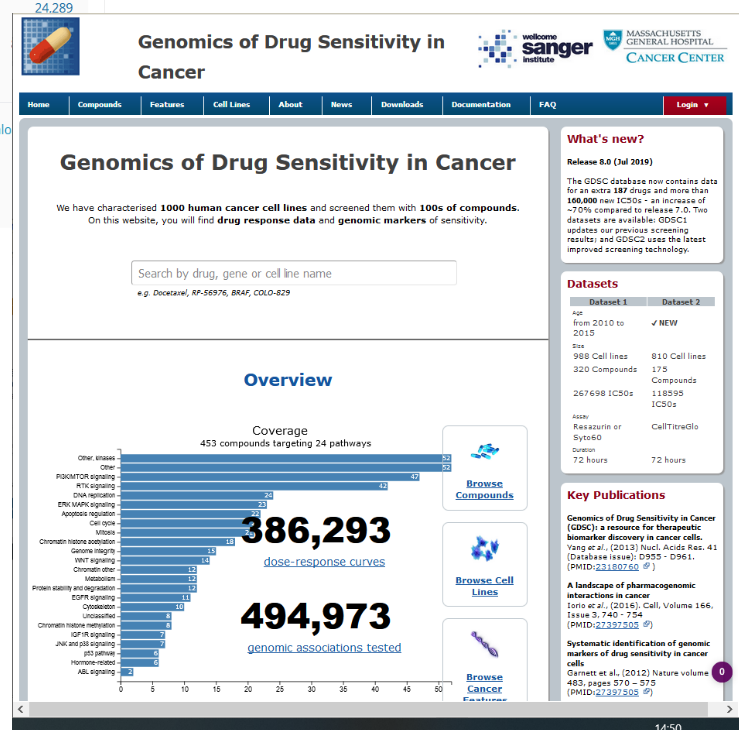
Select Disease below

Select Tumor below

NEWS

03-06-2019	Data from the new paper of Inoue et al., 2019, Cell 177, 1–15 June 13, 2019 "Illuminating G-Protein-Coupling Selectivity of GPCRs" has been curated and is available in the download page.
03-04-2019	Added GMT format for pathway data download
	SIGNOR is currently transitioning from using PubChem IDs to CHEBI IDs. In an effort to avoid any confusion for our

© 2019 SIGNOR



Genomics of Drug Sensitivity in Cancer

Genomics of Drug Sensitivity in Cancer

We have characterised **1000 human cancer cell lines** and screened them with **100s of compounds**. On this website, you will find **drug response data** and **genomic markers** of sensitivity.

Search by drug, gene or cell line name

e.g. Docetaxel, RP-56976, BRAF, COLO-829

Overview

Coverage
453 compounds targeting 24 pathways

386,293 dose-response curves

494,973 genomic associations tested

What's new?

Release 8.0 (Jul 2019)

The GDSC database now contains data for an extra **187** drugs and more than **160,000** new IC50s - an increase of ~70% compared to release 7.0. Two datasets are available: GDSC1 updates our previous screening results; and GDSC2 uses the latest improved screening technology.

Datasets

	Dataset 1	Dataset 2
Age	from 2010 to 2015	✓ NEW
Size	988 Cell lines	810 Cell lines
	320 Compounds	175 Compounds
	267698 IC50s	118595 IC50s
Assay	Resazurin or Syto60	CellTitreGlo
Duration	72 hours	72 hours

Key Publications

Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Yang *et al.*, (2013) Nucl. Acids Res. 41 (Database issue): D955 - D961. (PMID:23180760)

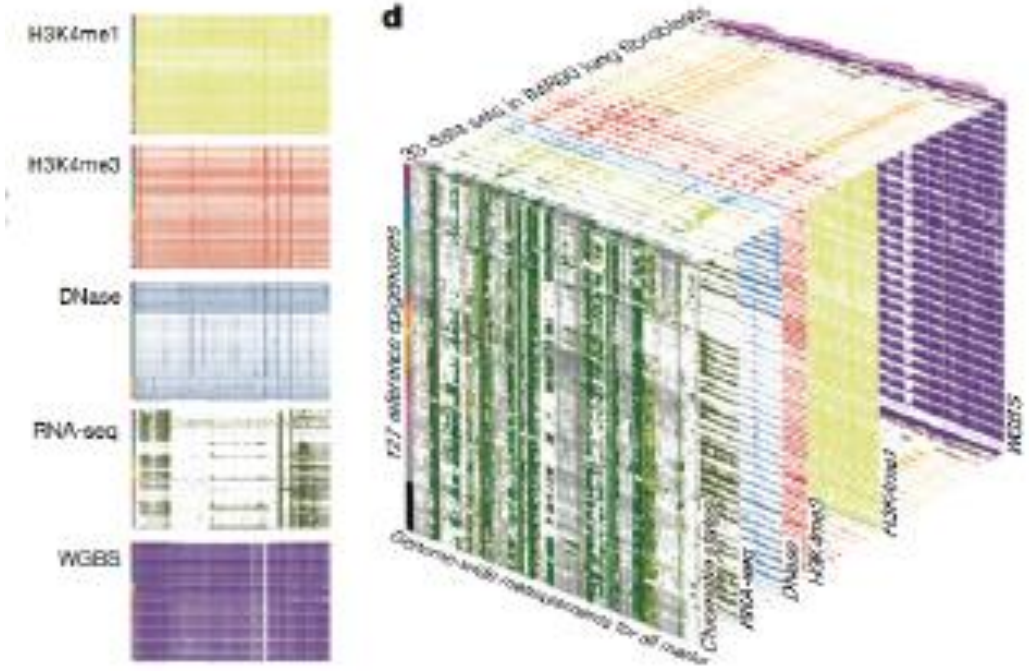
A landscape of pharmacogenomic interactions in cancer Iorio *et al.*, (2016). Cell, Volume 166, Issue 3, 740 - 754 (PMID:27397505)

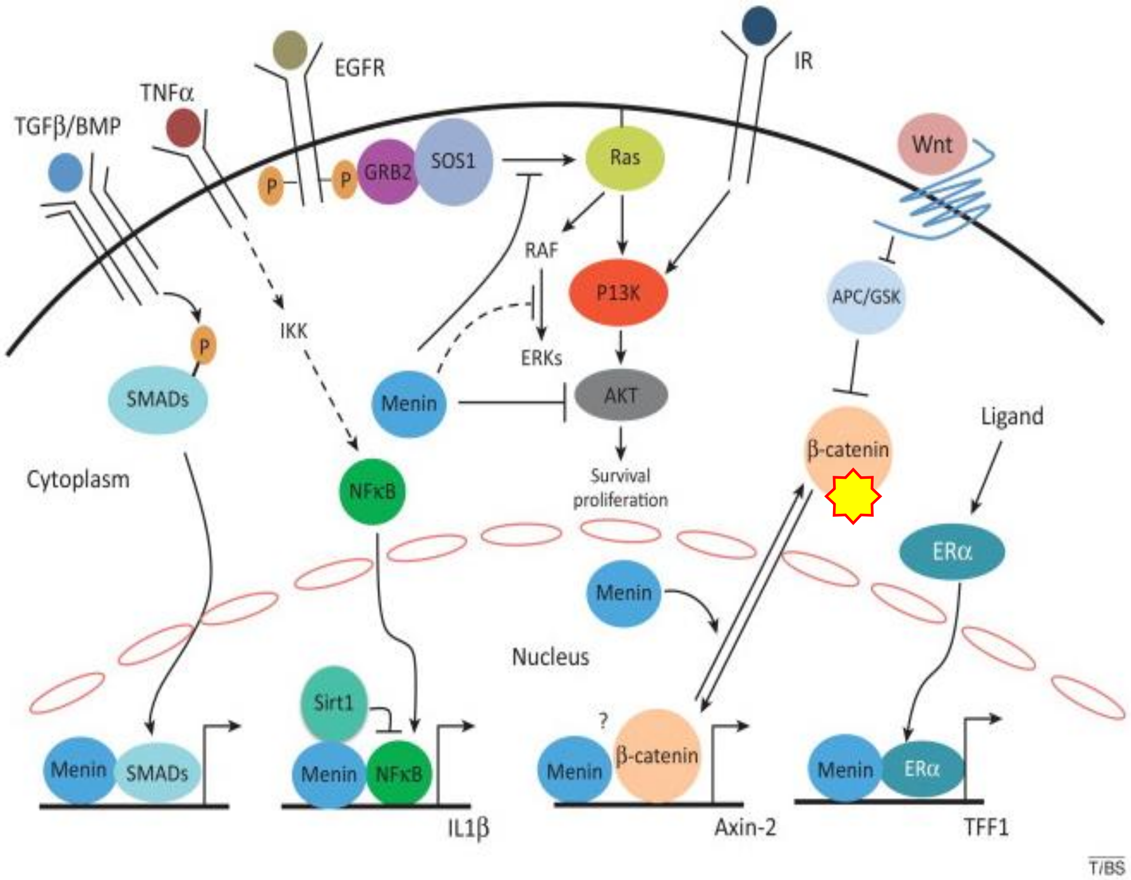
Systematic identification of genomic markers of drug sensitivity in cancer cells Garnett et al., (2012) Nature volume 483, pages 570 – 575 (PMID:27397505)

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

To infer mechanisms from omics data we need networks and models

IHEC
International Human Epigenome Consortium

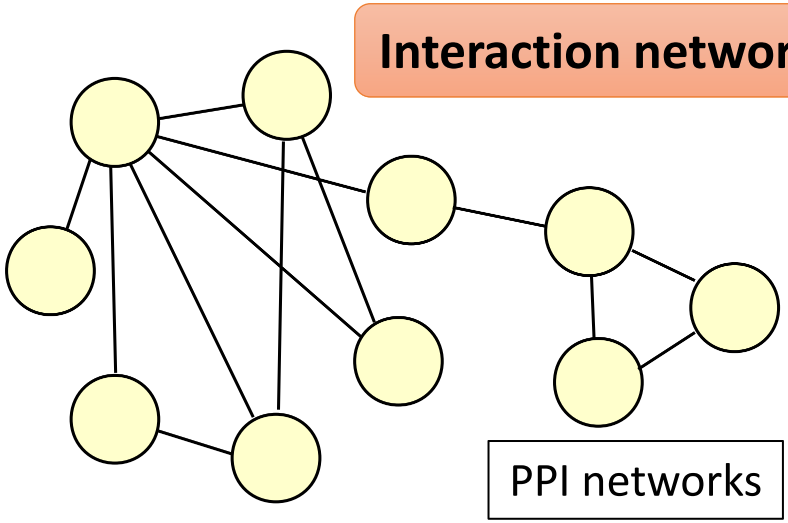




Data leads to Structure … and Structure leads to Modelling

Le Novere (2015), *Nat Rev Genet* 16(3):146-58.

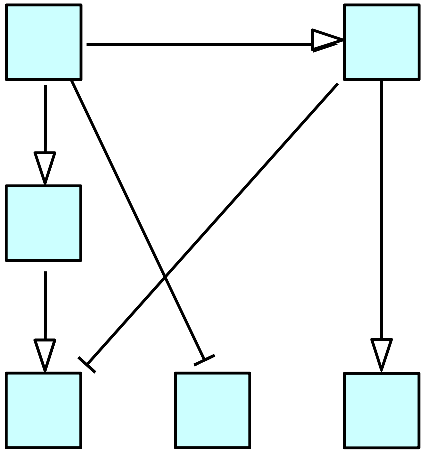
Interaction networks



PPI networks

Statistical modelling

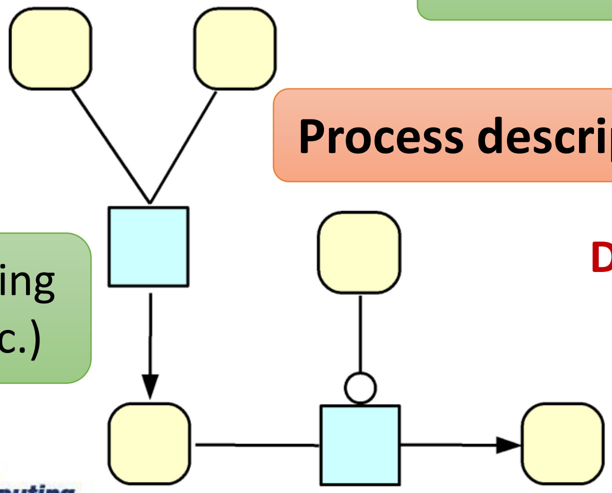
Activity flows



Signalling pathways, gene regulatory networks

Logical modelling

Process descriptions

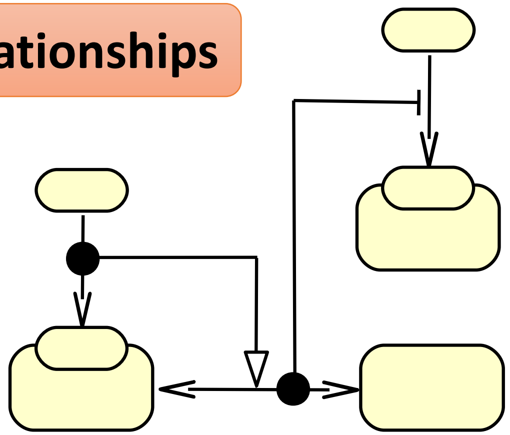


Process modelling (FBA, ODEs, etc.)

Metabolic networks, reaction networks

Different biological knowledge ↔ different types of networks ↔ different types of modelling

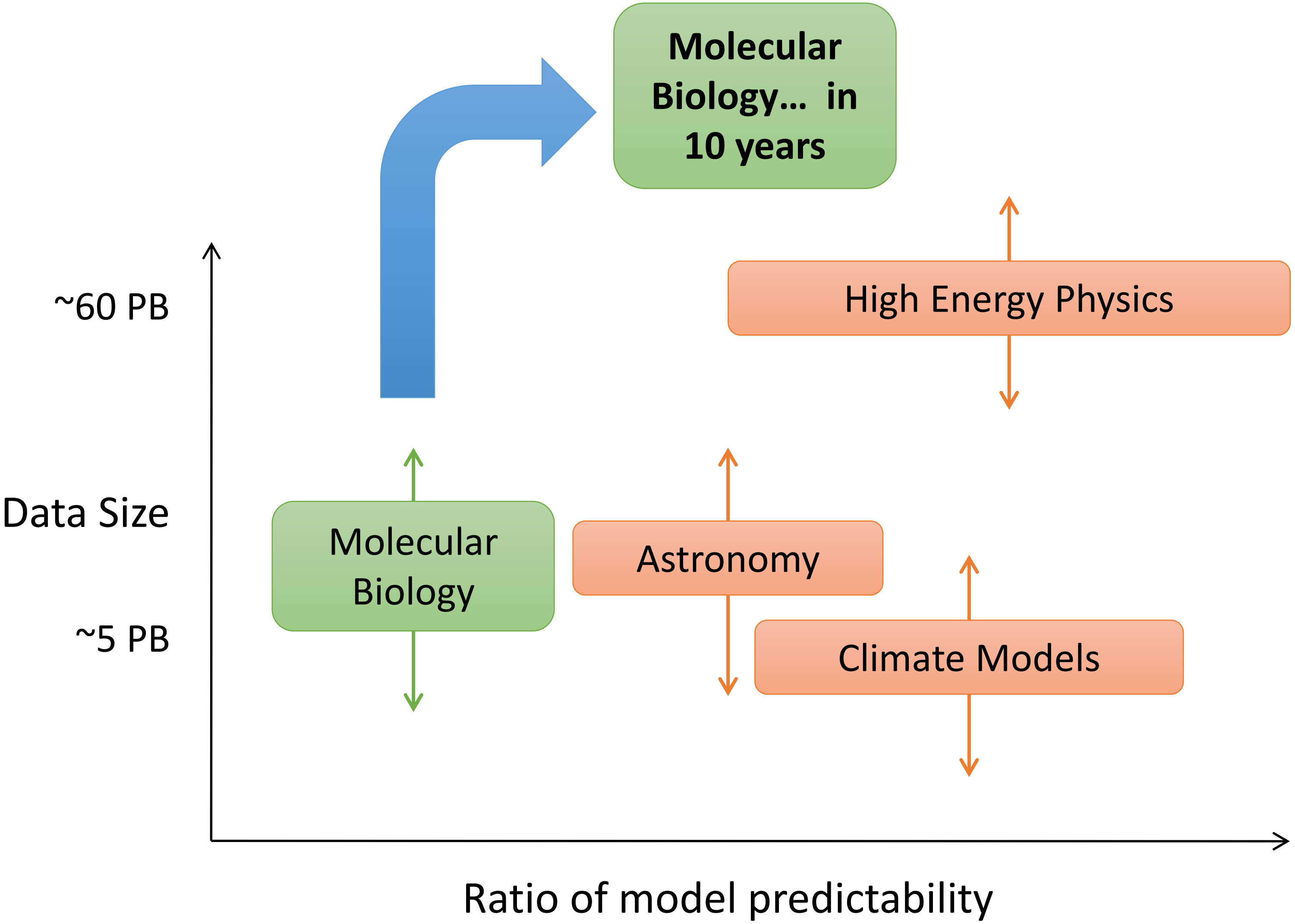
Entity relationships



Rule-based modelling

Molecular interactions, reaction networks

Modern Biology means Big Data

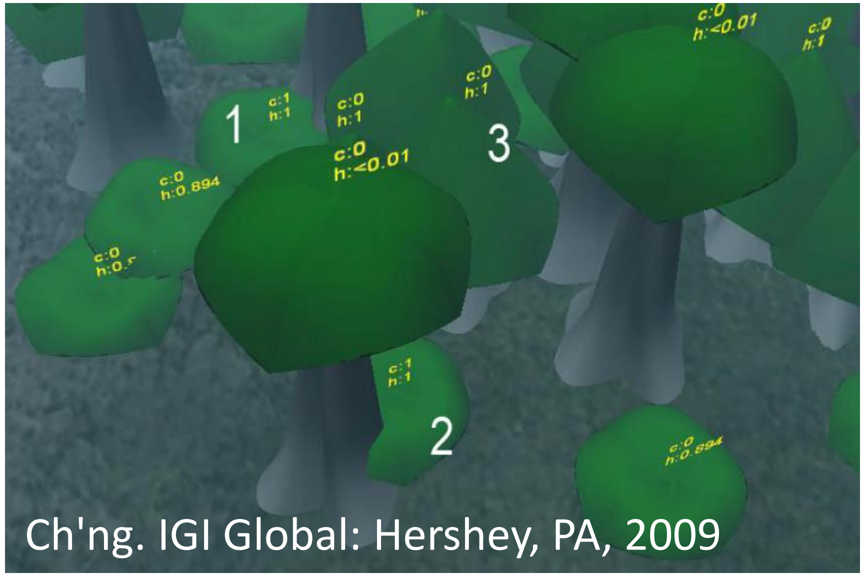


Slide from Ewan Birney, EBI

Agent-based modelling

Agent-based models are composed of:

- numerous agents;
- decision-making heuristics;
- an interaction topology; and
- a description of the environment.

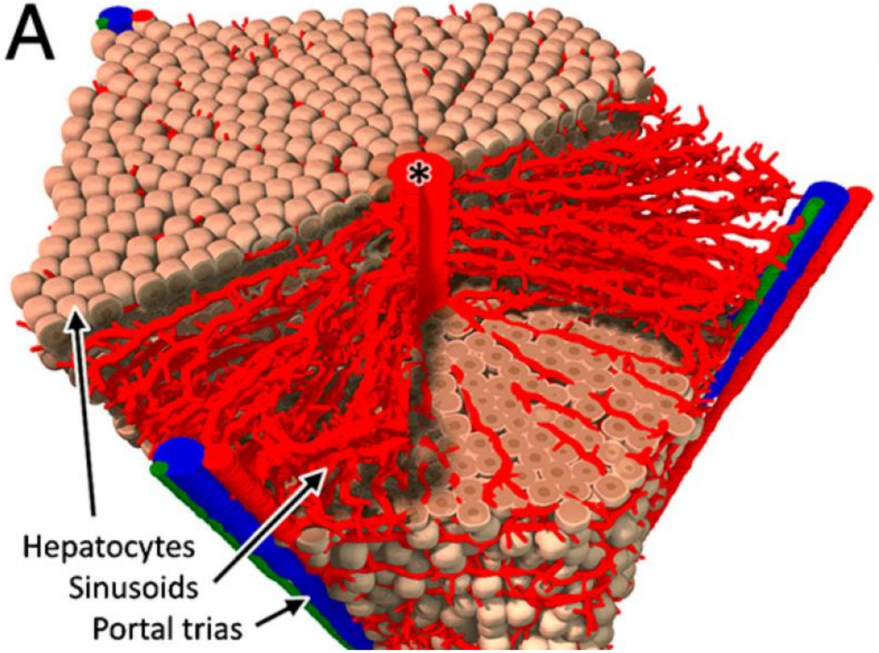


Ch'ng. IGI Global: Hershey, PA, 2009

- Examples:
  - Ecology
  - Environmental Science
  - Artificial Intelligence
  - Tissue Biology

Iteration 60

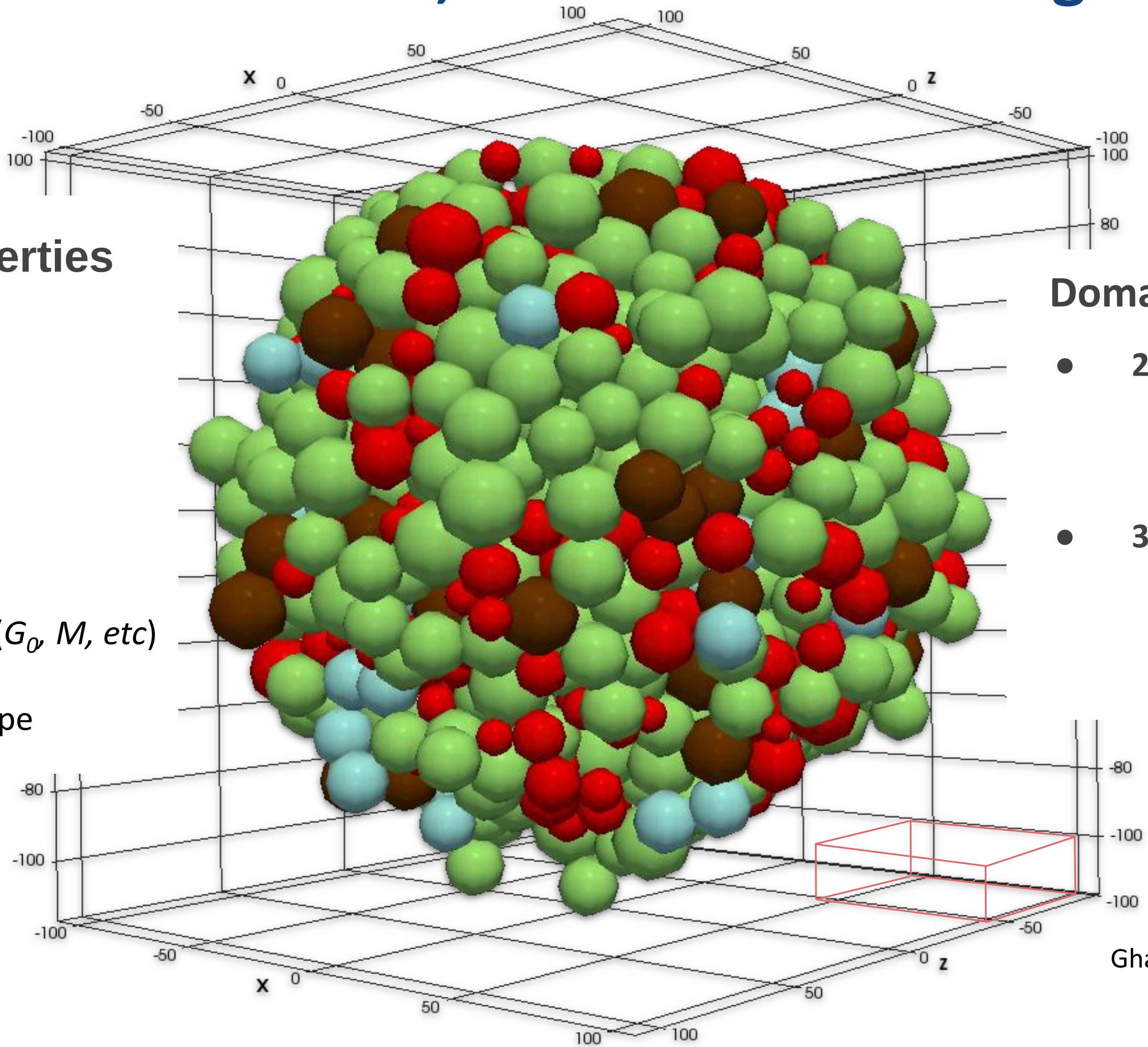
Situngkir, arXiv, 2004



Hoehme et al, *PNAS*, 2010

Osborne et al, *PLOS Comp Bio,* 2017
Metzcar et al, *JCO Clinical Cancer Informatics*, 2019

Agent-based is a flexible, multiscale modelling framework



Cell agent properties

- **Cell Volume**
  - nucleus
  - cytoplasm
- **Position** (x, y, z)
  - Neighborhood
  - Environment
- **Cell internal state**
  - Cell cycle phase (*G_0, M, etc*)
  - Growth rate
  - Custom phenotype

Domain = Voxels' grid

- **2D-Monolayers**
  - petri-dish
  - epitelia
  - bio-film
- **3D-Shapes**
  - spheroid
  - ductal
  - more complex shapes

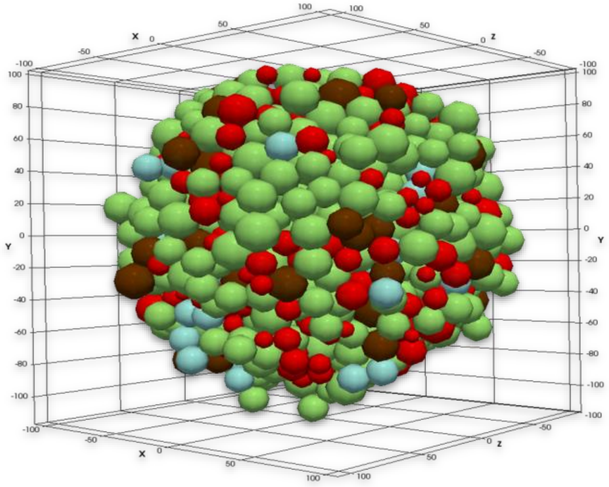
PhysiCell, PhysiBoSS

Ghaffarizadeh et al, *PLOS Comp Biol.*, 2018

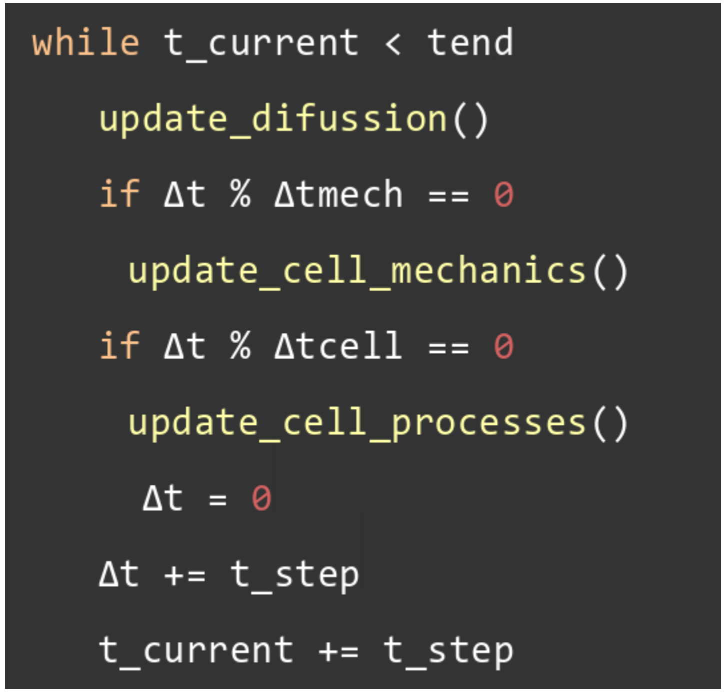
Ponce-de-Leon et al. (2022) *bioRxiv*, 2022.01.06.468363.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Multiscale because we consider different time scales



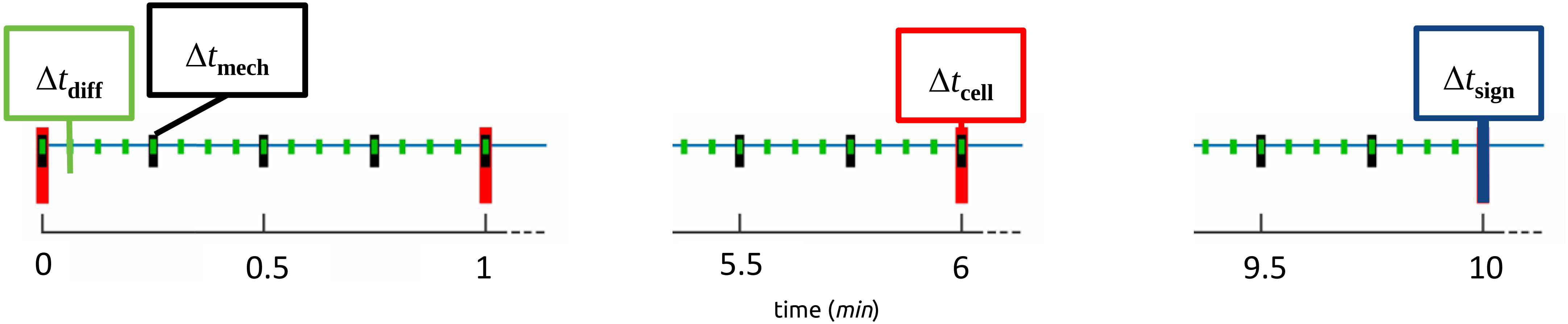
Simulation’s main loop



```
while t_current < tend
   update_difussion()
   if Δt % Δtmech == 0
    update_cell_mechanics()
   if Δt % Δtcell == 0
    update_cell_processes()
     Δt = 0
   Δt += t_step
   t_current += t_step
```

Time scales

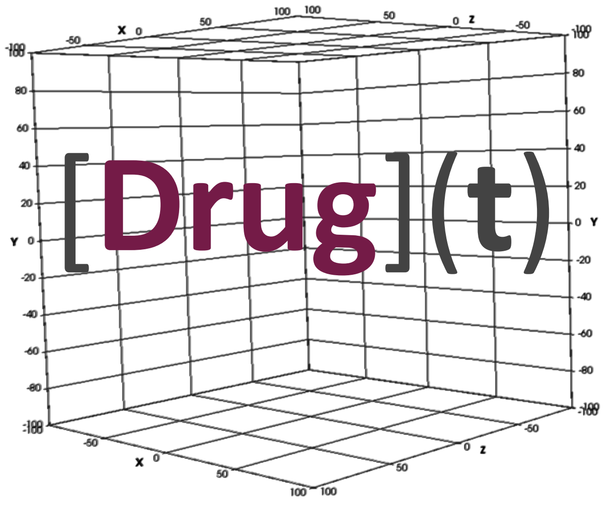
- $\Delta t_{\mathbf{diff}}$: (diffusion/transport): 0.01 min
- $\Delta t_{\mathbf{mech}}$: (cell movement): 0.1 min
- $\Delta t_{\mathbf{cell}}$: (cell processes): 6 min
- $\Delta t_{\mathbf{signalling}}$: (Boolean simulation): 10 min

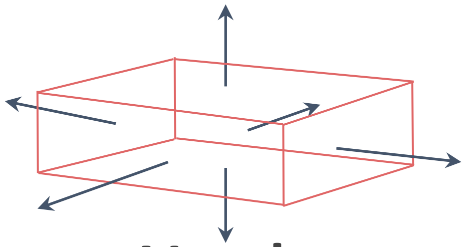


$\Delta t_{\mathbf{diff}}$
$\Delta t_{\mathbf{mech}}$
$\Delta t_{\mathbf{cell}}$
$\Delta t_{\mathbf{sign}}$
0
0.5
1
5.5
6
9.5
10
time (*min*)

Environment can be dynamic and reactive

$$\Delta t_{\textbf{diff}}$$
$$\Delta t_{\textbf{mech}}$$





Voxel

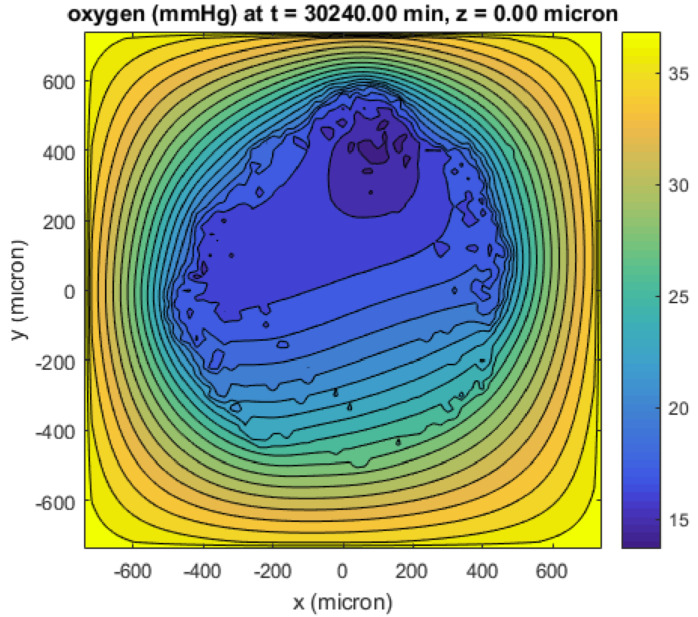
Diffusion equation

$$\frac{\partial \boldsymbol{\rho}}{\partial t} = \overbrace{\mathbf{D}\nabla^2\boldsymbol{\rho}}^{\text{diffusion}} - \overbrace{\boldsymbol{\lambda\rho}}^{\text{decay}} + \overbrace{\mathbf{S}(\boldsymbol{\rho}^* - \boldsymbol{\rho})}^{\text{bulk source}} - \overbrace{\mathbf{U}\boldsymbol{\rho}}^{\text{bulk uptake}}$$
$$+ \overbrace{\sum_{\text{cells } k} \delta(\mathbf{x} - \mathbf{x}_k) W_k [\mathbf{S}_k(\boldsymbol{\rho}_k^* - \boldsymbol{\rho}) - \mathbf{U}_k\boldsymbol{\rho}]}^{\text{sources and uptake by cells}} \text{ in } \Omega$$

System of PDEs for each molecule:

- Diffusion term
- Decay
- Uptake/Production

Gradient of chemical factors (O2)



Mechanical equation

$$\mathbf{v}_i = \sum_{j\in\mathcal{N}(i)} \left(- \overbrace{\sqrt{c^i_{\text{cca}} c^j_{\text{cca}}} \nabla \phi_{1,R_{i,A}+R_{j,A}}(\mathbf{x}_i - \mathbf{x}_j)}^{\text{cell-cell adhesion}} - \overbrace{\sqrt{c^i_{\text{ccr}} c^j_{\text{ccr}}} \nabla \psi_{1,R_i+R_j}(\mathbf{x}_i - \mathbf{x}_j)}^{\text{cell-cell repulsion}} \right).$$
$$- \overbrace{c^i_{\text{cba}} \nabla \phi_{1,R_{i,A}}(-d(\mathbf{x}_i)\mathbf{n}(\mathbf{x}_j))}^{\text{cell-BM adhesion}} - \overbrace{c^i_{\text{cbr}} \nabla \psi_{1,R_i}(-d(\mathbf{x}_i)\mathbf{n}(\mathbf{x}_j))}^{\text{cell-BM repulsion}} + \mathbf{v}_{i,\text{mot}}$$

Surrounding physical environment

Surrogate for extra-cellular matrix

- Field with densities that can be produced & consumed
- Inert agents that can be moved

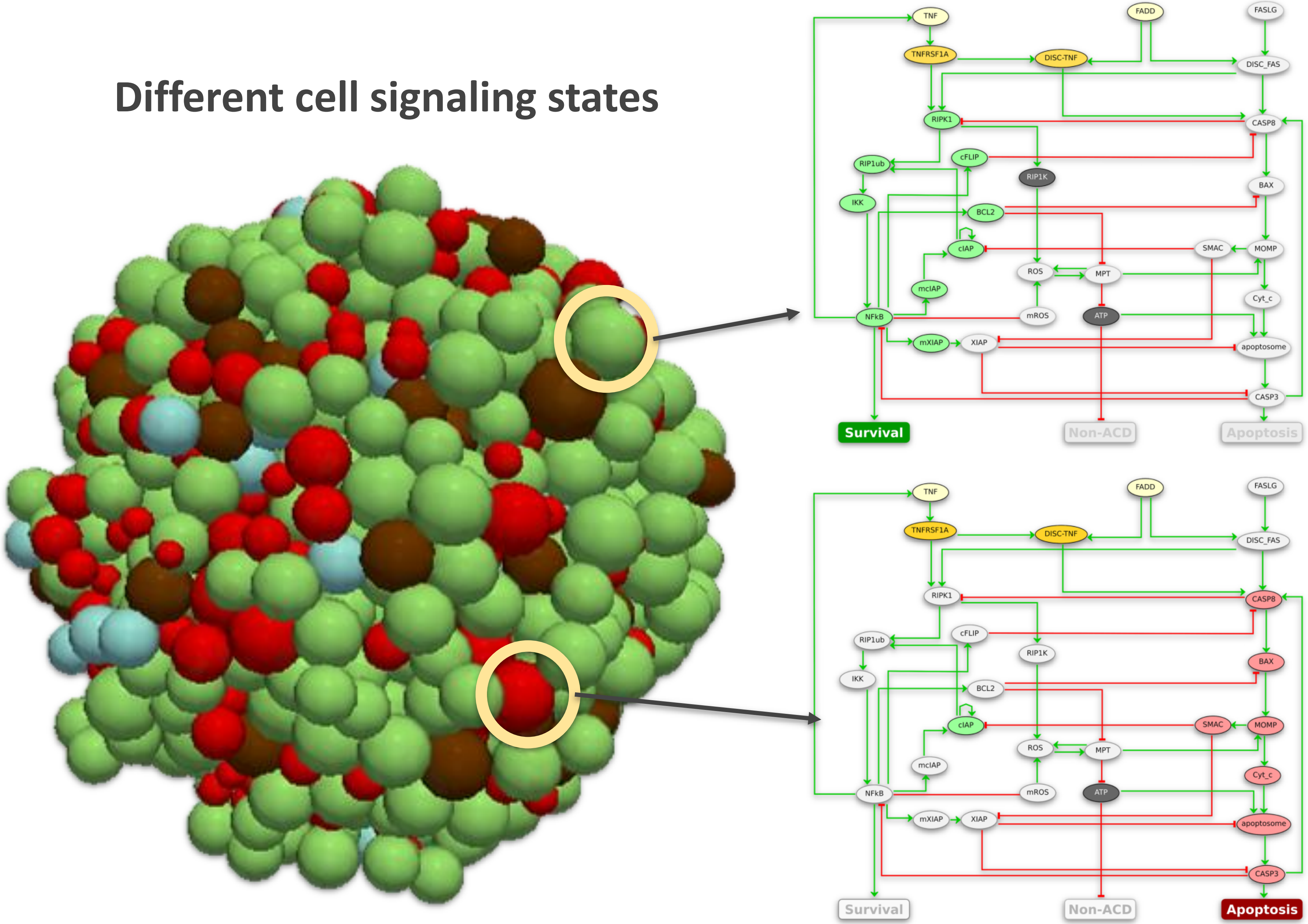
BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Cells have different phenotypes depending on their genes’ activation

Cell Cycle Phase

- 🟩 Premitotic
- 🟦 Postmitotic
- 🟨 Ki67 negative
- 🟥 Apoptotic
- 🟧 Necrotic
- 🟫 Necrotic (swelling)
- ⬜ Necrotic (lysis)

Different cell signaling states

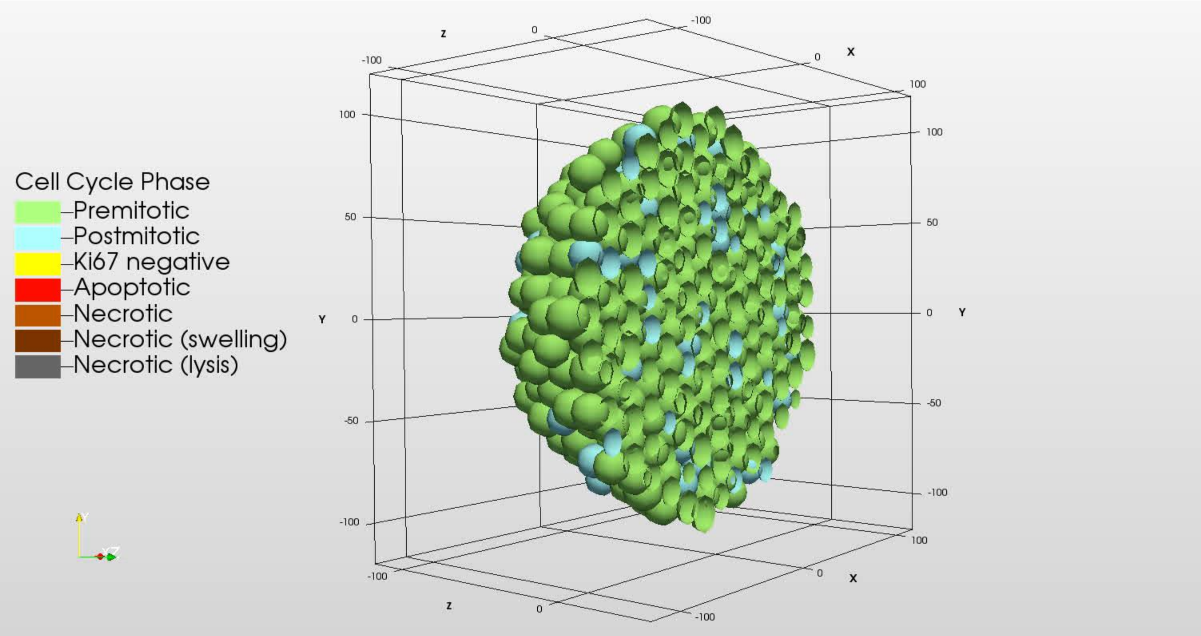


Time scales

- $\Delta t_{\mathbf{diff}}$ / $\Delta t_{\mathbf{mech}}$/ $\Delta t_{\mathbf{cell}}$
- $\Delta t_{\mathbf{signalling}}$

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

PhysiBoSS allows for finding optimal drug regimes



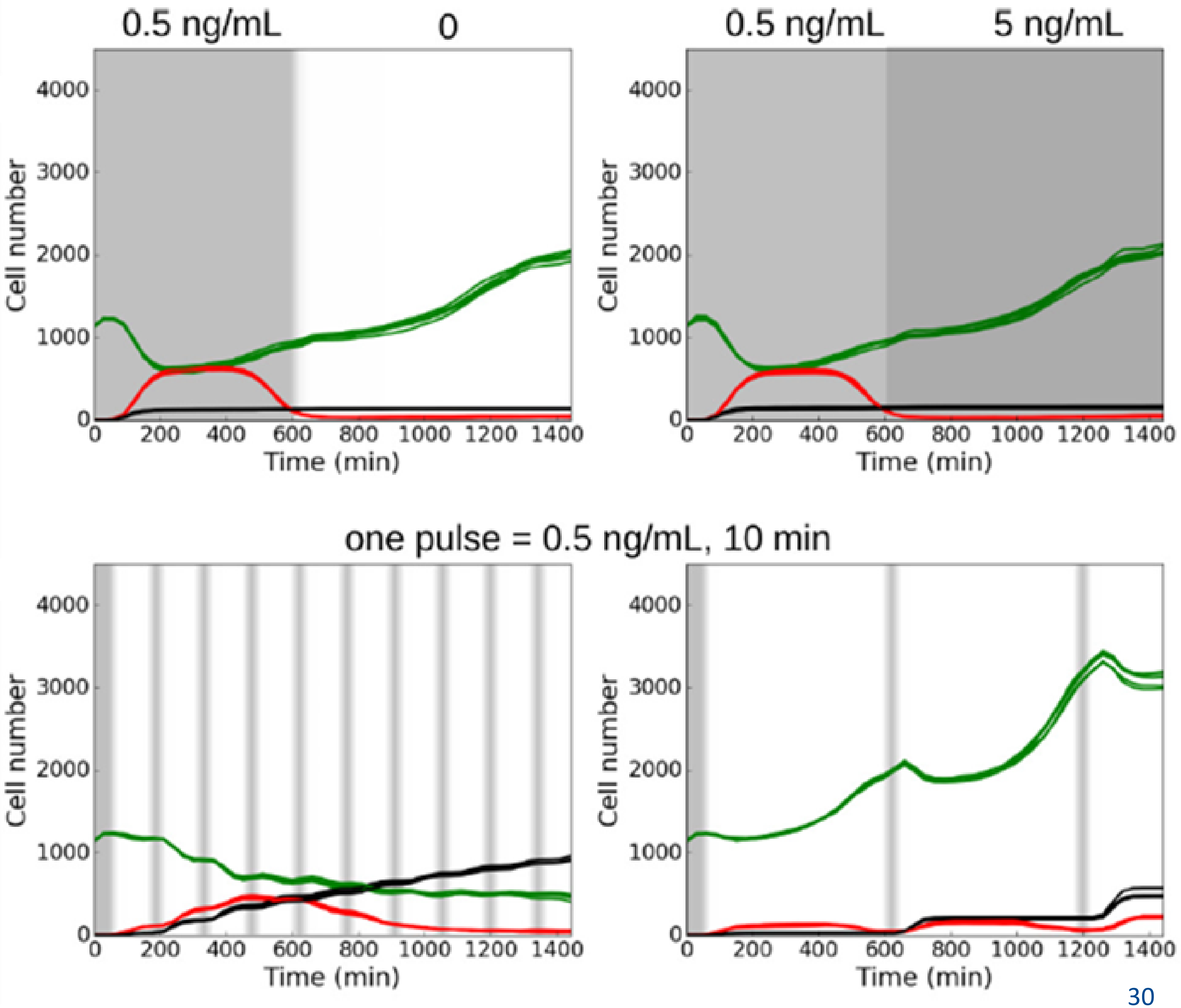
~48 h simulation time, 30 min wall time
~2500 cells

Figure from Miguel Ponce-de-Leon, BSC

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Ponce-de-Leon et al. (2022) *bioRxiv*, 2022.01.06.468363
Ponce-de-Leon et al. (2022) *Frontiers in Molecular Biosciences, 9*

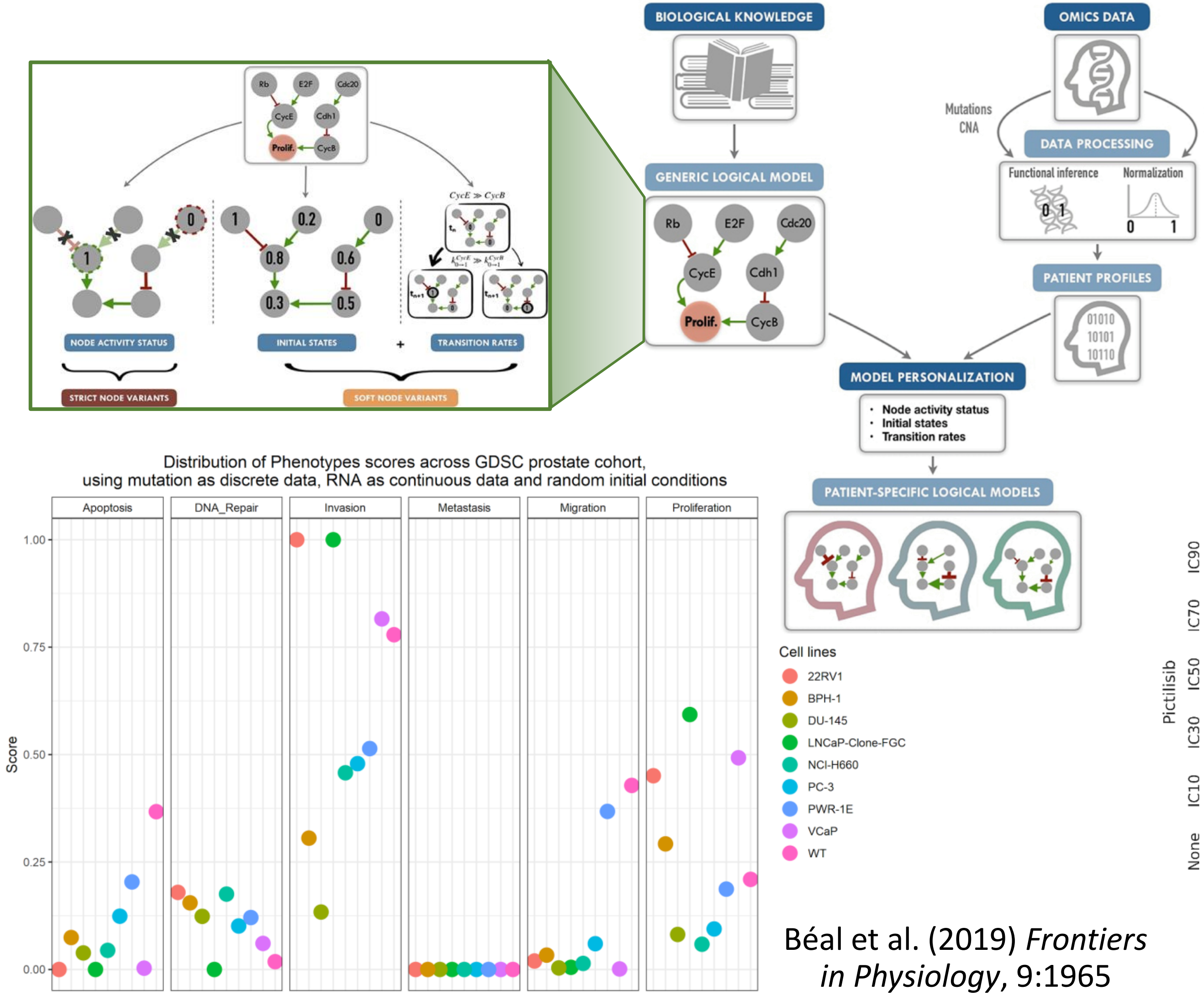
Proliferation Apoptosis Necrosis



Letort et al. (2018), *Bioinformatics*, bty766

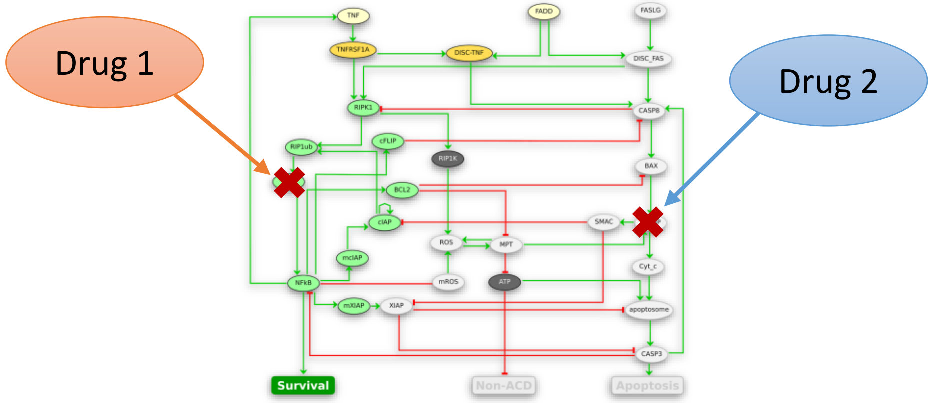
PhysiBoSS allows for personalised drug studies

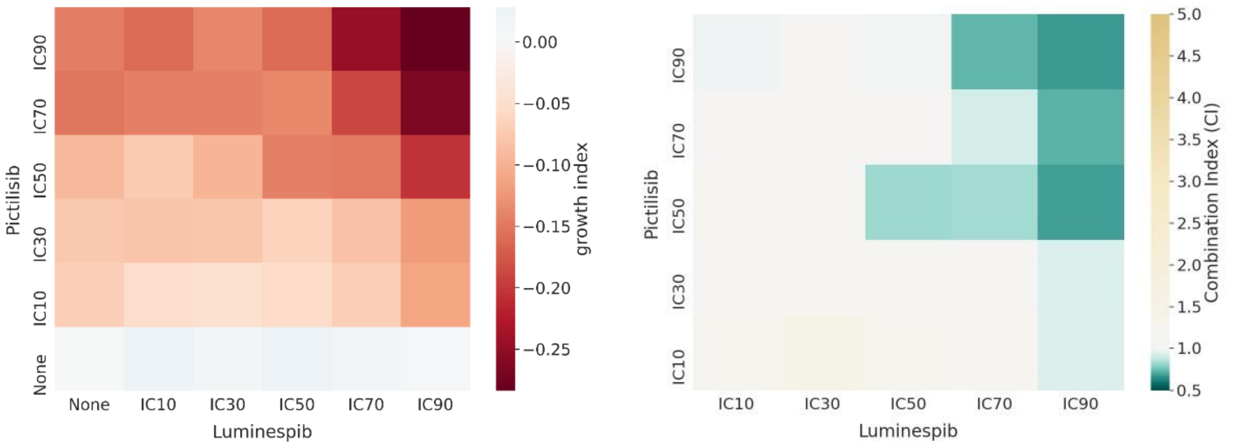
Personalisation of intracellular models



Béal et al. (2019) *Frontiers in Physiology*, 9:1965

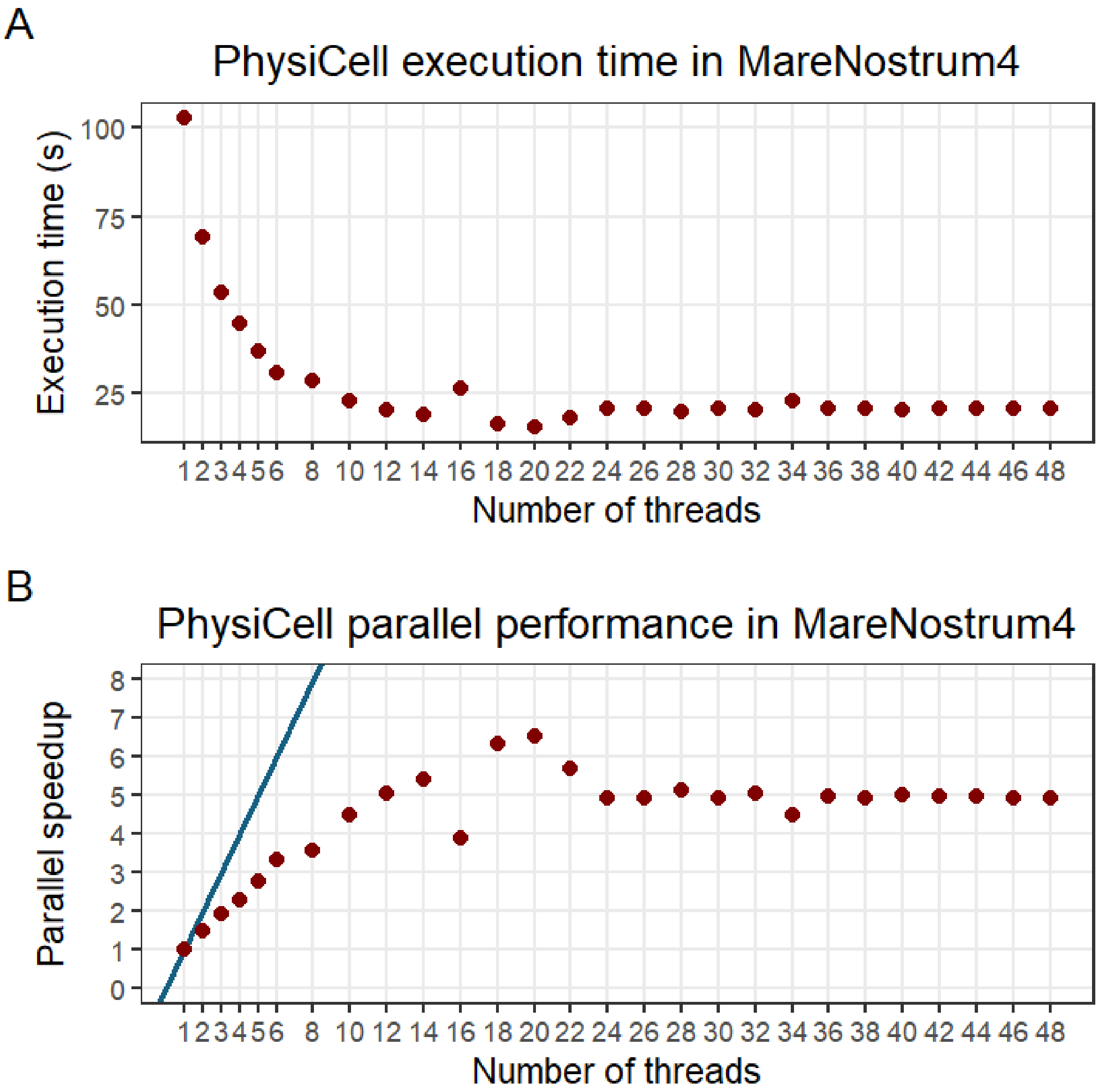
Different combinations of drugs





Montagud et al. (2022) *eLife* 2022;11:e72626

PhysiCell scalability analysis stagnates at ~8 cores

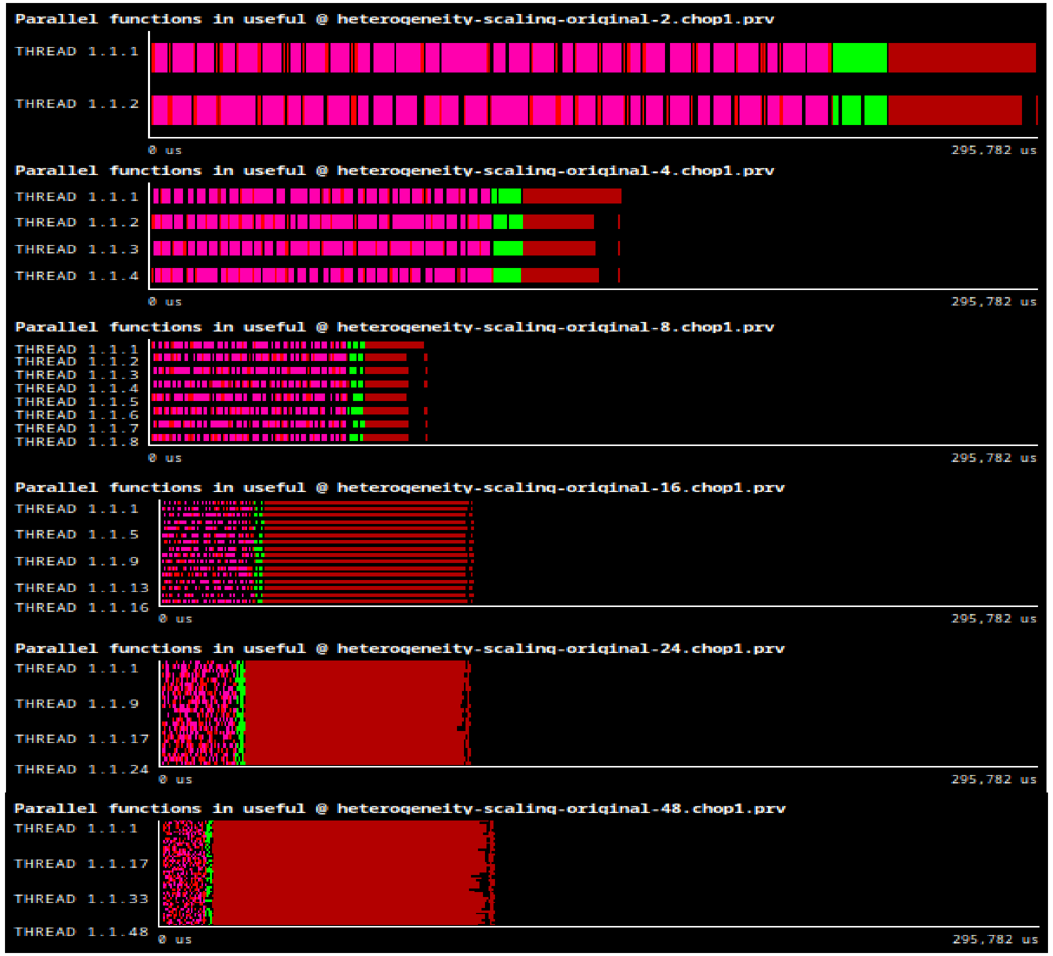


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Optimisation in MN4 : Improving PhysiCell's scalability –memory allocation problem

PhysiCell vanilla *malloc* library

PhysiCell using *jemalloc* library



2 threads

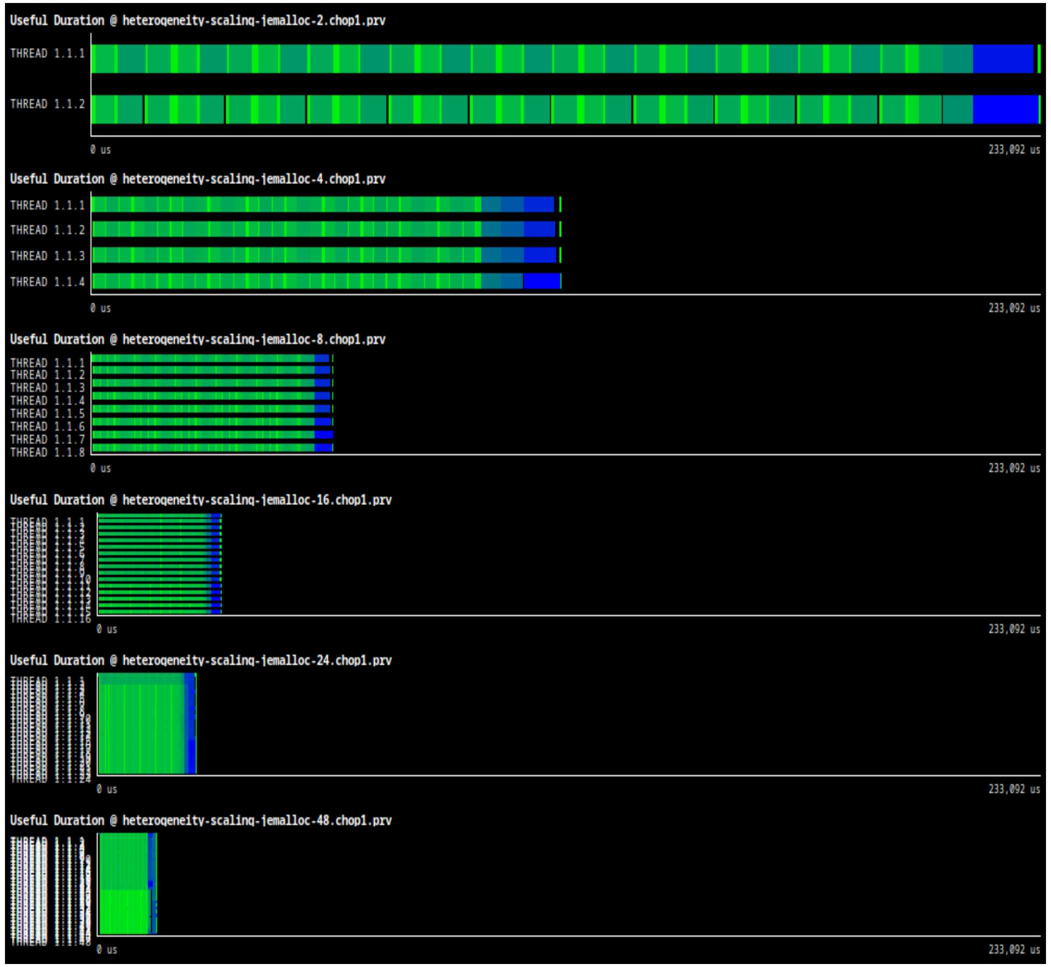
4 threads

8 threads

16 threads

24 threads

48 threads

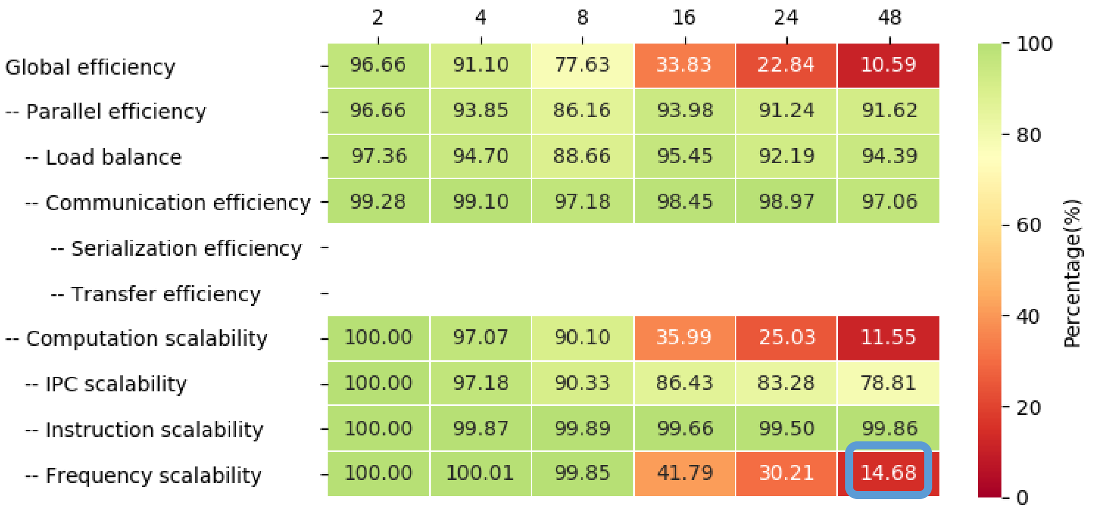


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Slide from Marc Clascà,
Marta Garcia-Gasulla, BSC

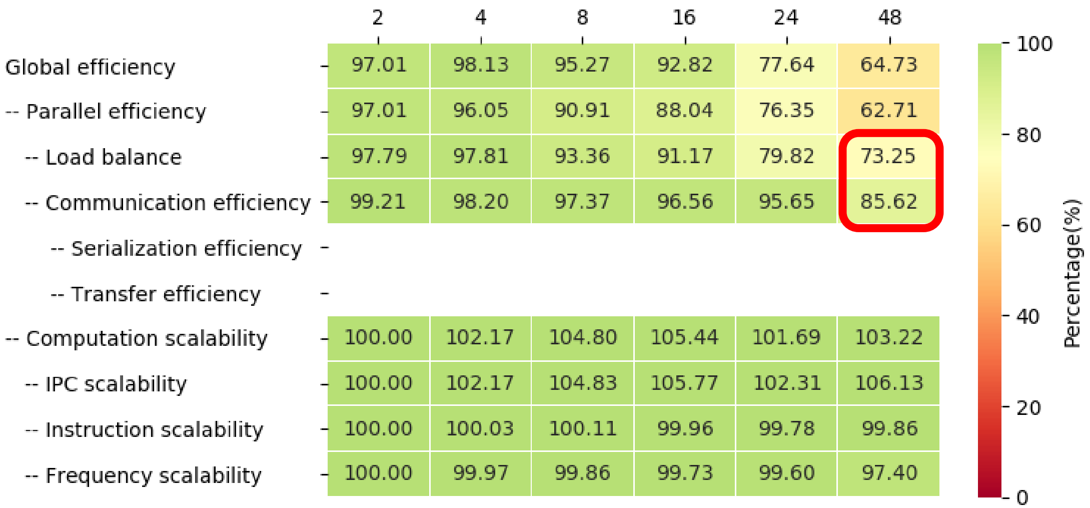
Benchmarking in MN4: Changing the library improves PhysiCell's scalability

PhysiCell vanilla *malloc* library

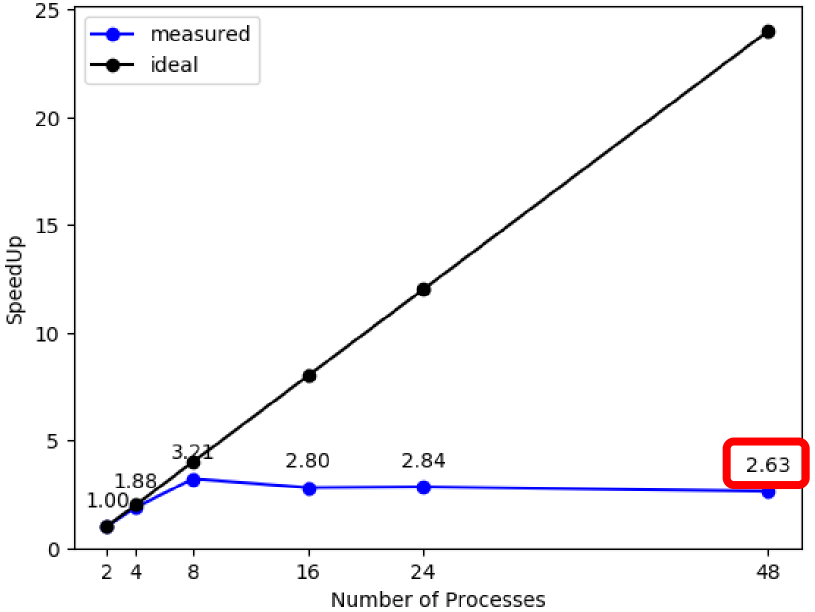


	2	4	8	16	24	48
Global efficiency	96.66	91.10	77.63	33.83	22.84	10.59
-- Parallel efficiency	96.66	93.85	86.16	93.98	91.24	91.62
-- Load balance	97.36	94.70	88.66	95.45	92.19	94.39
-- Communication efficiency	99.28	99.10	97.18	98.45	98.97	97.06
-- Serialization efficiency						
-- Transfer efficiency						
-- Computation scalability	100.00	97.07	90.10	35.99	25.03	11.55
-- IPC scalability	100.00	97.18	90.33	86.43	83.28	78.81
-- Instruction scalability	100.00	99.87	99.89	99.66	99.50	99.86
-- Frequency scalability	100.00	100.01	99.85	41.79	30.21	14.68

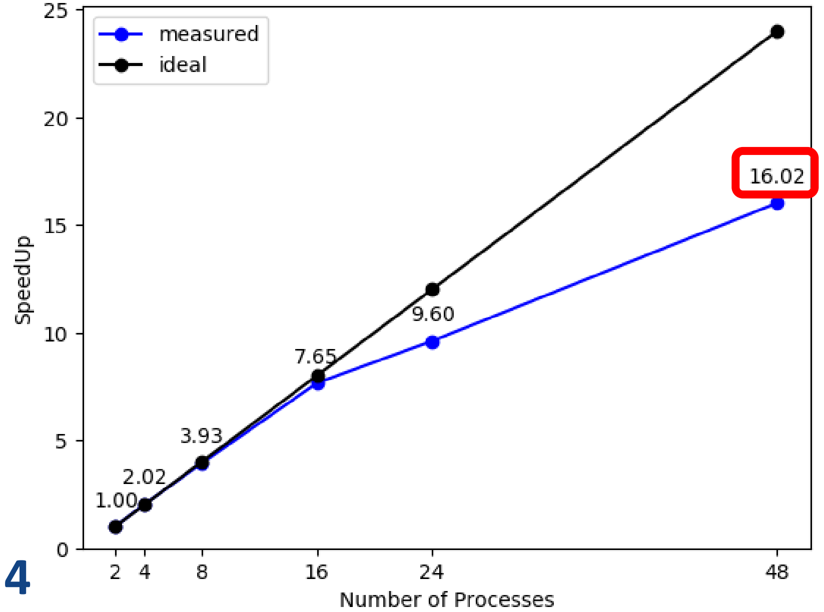
PhysiCell using *jemalloc* library



	2	4	8	16	24	48
Global efficiency	97.01	98.13	95.27	92.82	77.64	64.73
-- Parallel efficiency	97.01	96.05	90.91	88.04	76.35	62.71
-- Load balance	97.79	97.81	93.36	91.17	79.82	73.25
-- Communication efficiency	99.21	98.20	97.37	96.56	95.65	85.62
-- Serialization efficiency						
-- Transfer efficiency						
-- Computation scalability	100.00	102.17	104.80	105.44	101.69	103.22
-- IPC scalability	100.00	102.17	104.83	105.77	102.31	106.13
-- Instruction scalability	100.00	100.03	100.11	99.96	99.78	99.86
-- Frequency scalability	100.00	99.97	99.86	99.73	99.60	97.40



Strong scaling speedup with **2 processes as baseline**



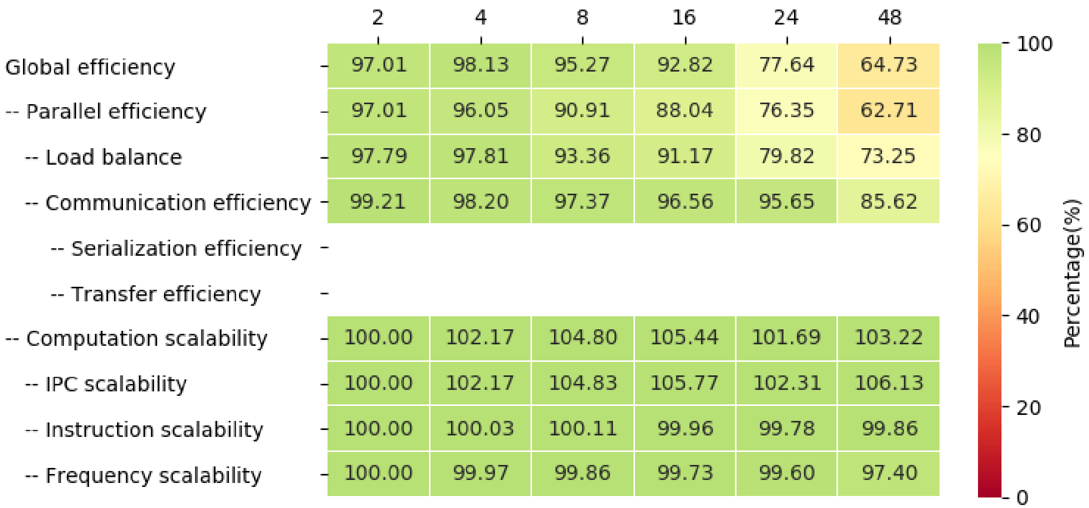
6x faster!

Slide from Marc Clascà, Marta Garcia-Gasulla, BSC

MareNostrum 4

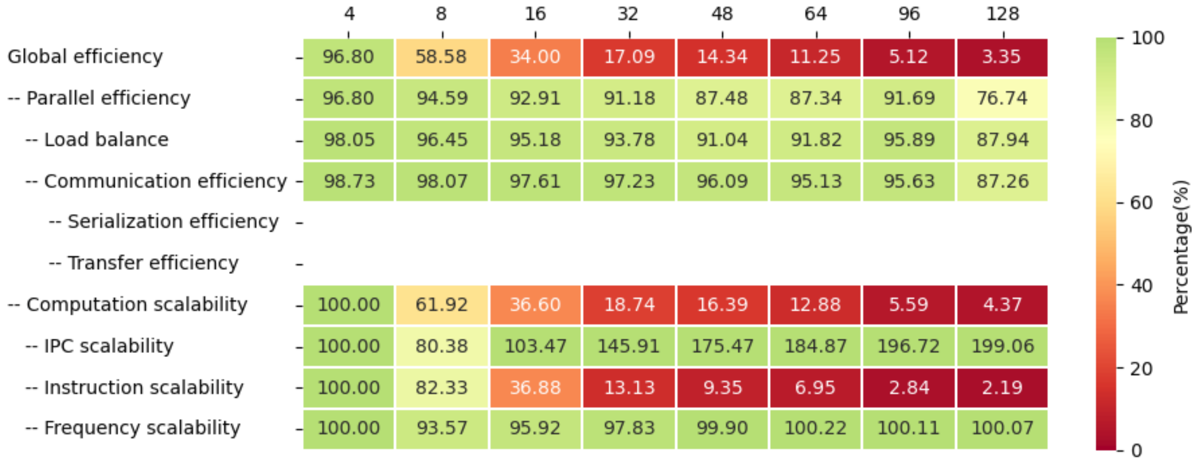
Benchmarking in MN4 vs Kupeng

PhysiCell using *jemalloc* library in **MN4**



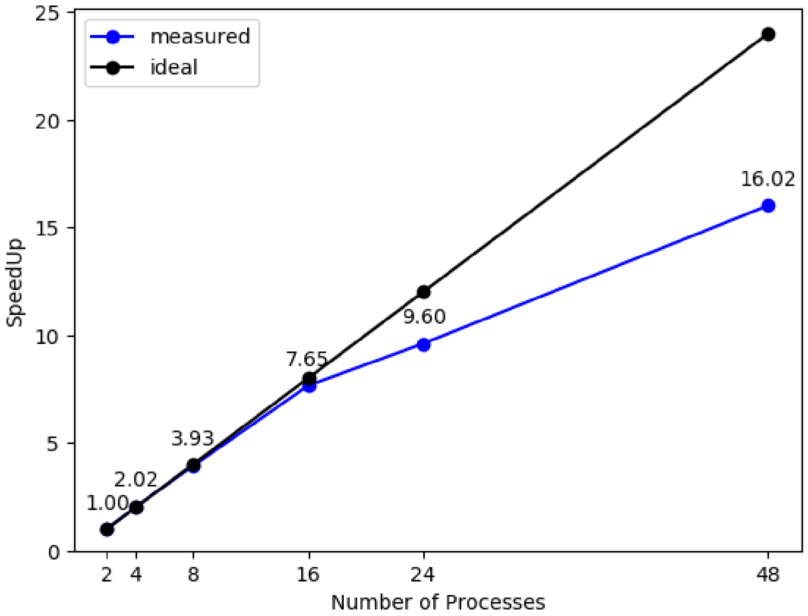
	2	4	8	16	24	48
Global efficiency	97.01	98.13	95.27	92.82	77.64	64.73
-- Parallel efficiency	97.01	96.05	90.91	88.04	76.35	62.71
-- Load balance	97.79	97.81	93.36	91.17	79.82	73.25
-- Communication efficiency	99.21	98.20	97.37	96.56	95.65	85.62
-- Serialization efficiency						
-- Transfer efficiency						
-- Computation scalability	100.00	102.17	104.80	105.44	101.69	103.22
-- IPC scalability	100.00	102.17	104.83	105.77	102.31	106.13
-- Instruction scalability	100.00	100.03	100.11	99.96	99.78	99.86
-- Frequency scalability	100.00	99.97	99.86	99.73	99.60	97.40

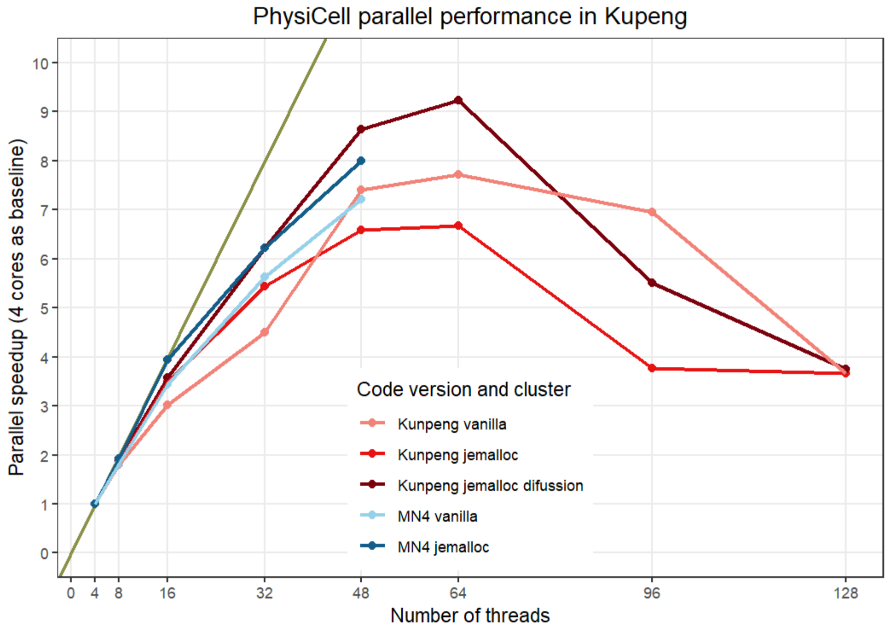
PhysiCell using *jemalloc* library in **Kupeng**



	4	8	16	32	48	64	96	128
Global efficiency	96.80	58.58	34.00	17.09	14.34	11.25	5.12	3.35
-- Parallel efficiency	96.80	94.59	92.91	91.18	87.48	87.34	91.69	76.74
-- Load balance	98.05	96.45	95.18	93.78	91.04	91.82	95.89	87.94
-- Communication efficiency	98.73	98.07	97.61	97.23	96.09	95.13	95.63	87.26
-- Serialization efficiency								
-- Transfer efficiency								
-- Computation scalability	100.00	61.92	36.60	18.74	16.39	12.88	5.59	4.37
-- IPC scalability	100.00	80.38	103.47	145.91	175.47	184.87	196.72	199.06
-- Instruction scalability	100.00	82.33	36.88	13.13	9.35	6.95	2.84	2.19
-- Frequency scalability	100.00	93.57	95.92	97.83	99.90	100.22	100.11	100.07

Strong scaling speedup with **2 processes as baseline**

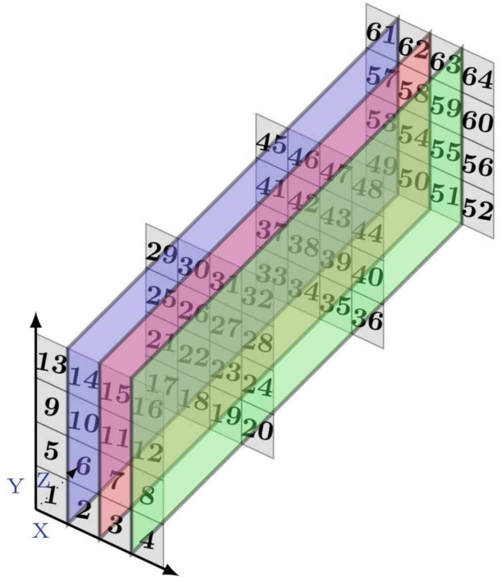


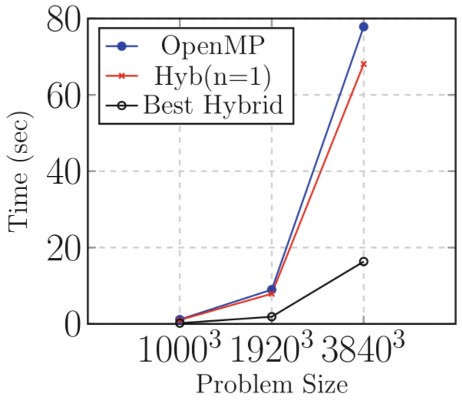


Strong scaling speedup with **4 processes as baseline**

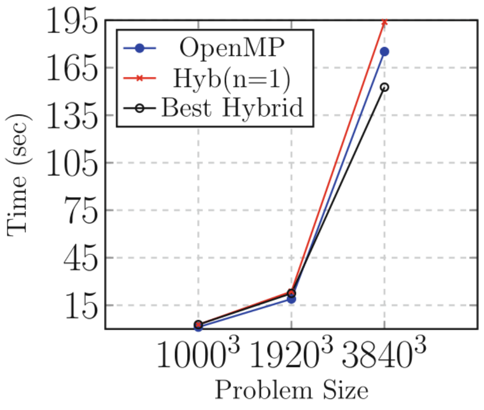
Work from José Estragués, BSC

PhysiCell-X: extending PhysiCell with MPI





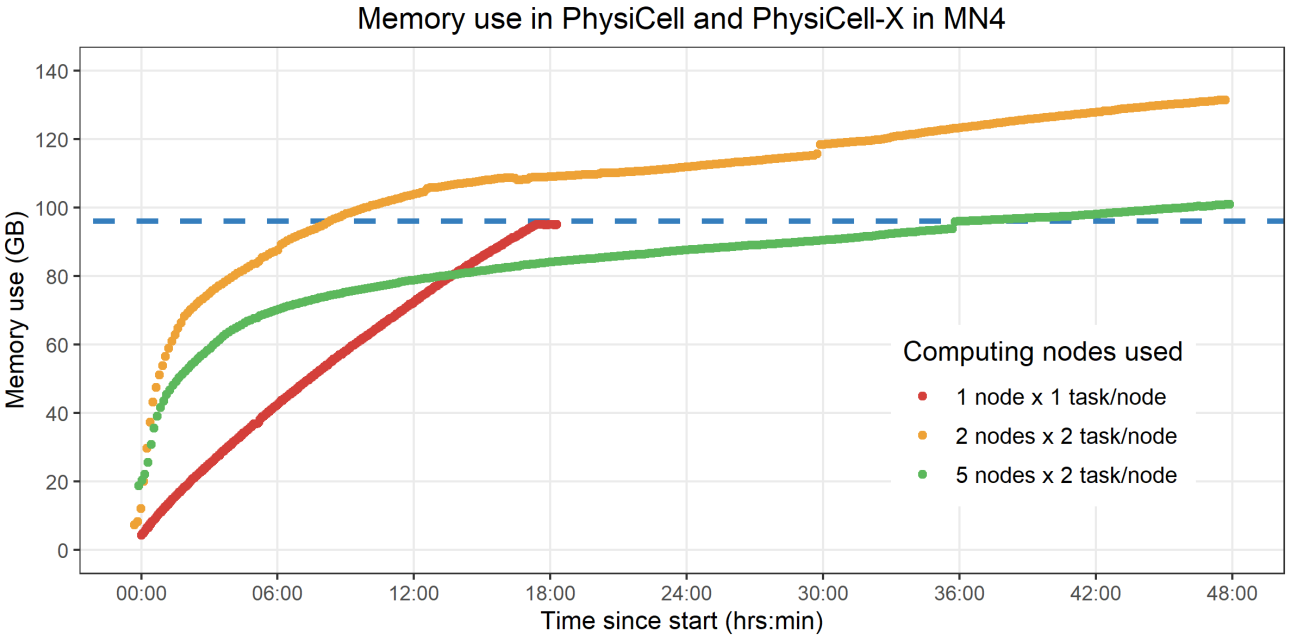
(a) μ-environment



(e) Thomas solver

- Re-factored the diffusion solver
    - Lower scale of multiscale
- Allows to simulate bigger setups
    - Needed to reach huge, complex simulations
- Still efficient vs serial in smaller setups

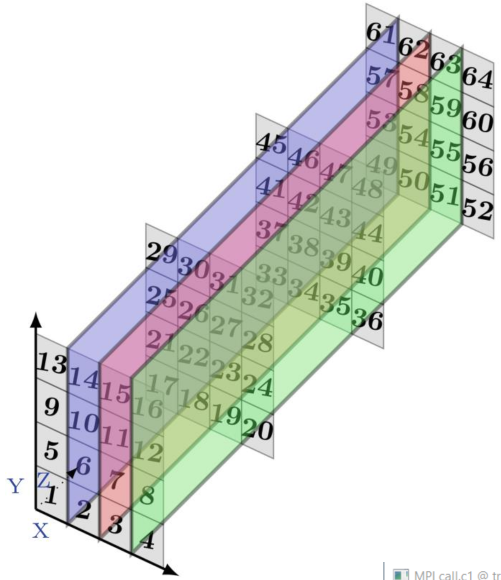
Domain: 7680x7680x7680 ≈ 0.5 billion voxels, 1 substrate	OpenMP	Hybrid (n=4)	Hybrid (n=8)
Build Microenvironment	x	141.98 s	67.81 s
Gaussian profile	x	0.92 s	0.45 s
File I/O	x	7.30 s	7.40 s
Agent Generation	x	0.11 s	0.0023 s
Source/Sink Diffusion Solver	x	1109.69 s	1210.41 s


Memory use in PhysiCell and PhysiCell-X in MN4

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

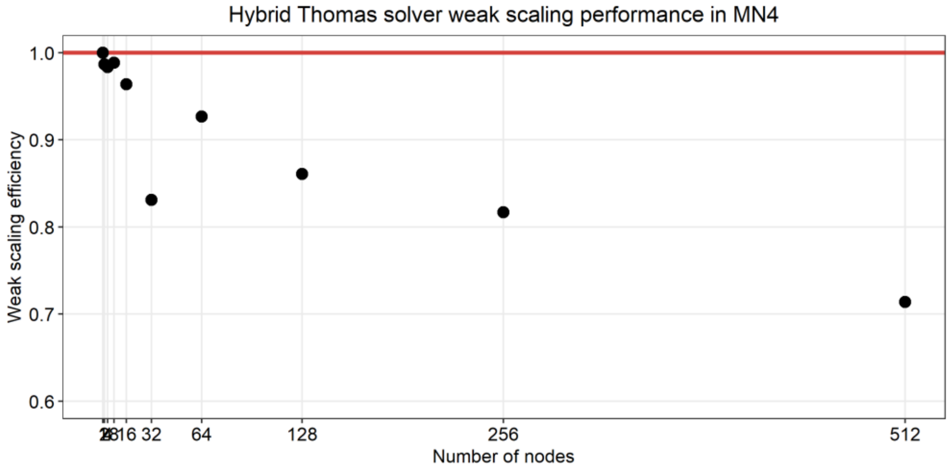
Saxena, G. *et al.* (2021) BioFVM-X: An MPI+OpenMP 3-D Simulator for Biological Systems., *Computational Methods in Systems Biology*, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 266–279.

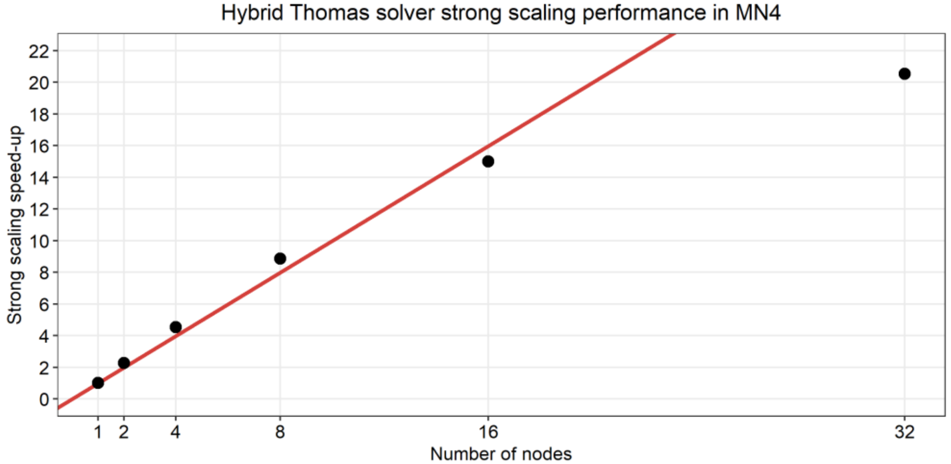
Towards a flexible exascale multiscale framework using PhysiCell-X



Limitations:

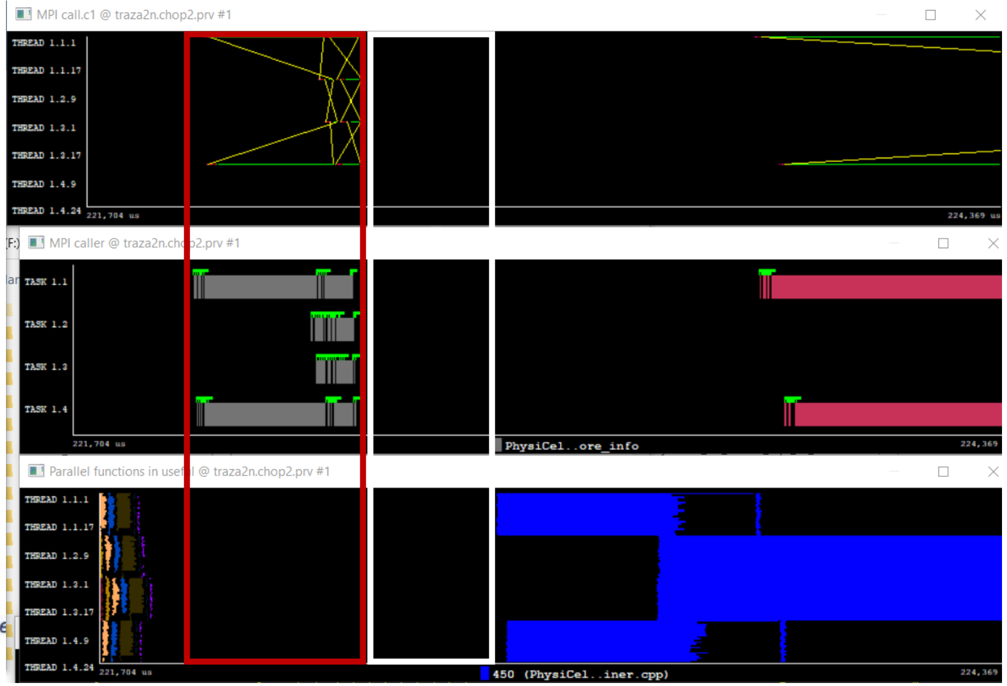
- MPI domain decomposition is 1D
- Diffusion solver in BioFVM-X is serial on the X-axis
- Working on a Parallel Modified (hybrid) Thomas solver





MPI send

MPI unpack



Testing the performance of PhysiCell-X

- Overhead of MPI messaging

Barcelona Supercomputing Center
Centro Nacional de Supe

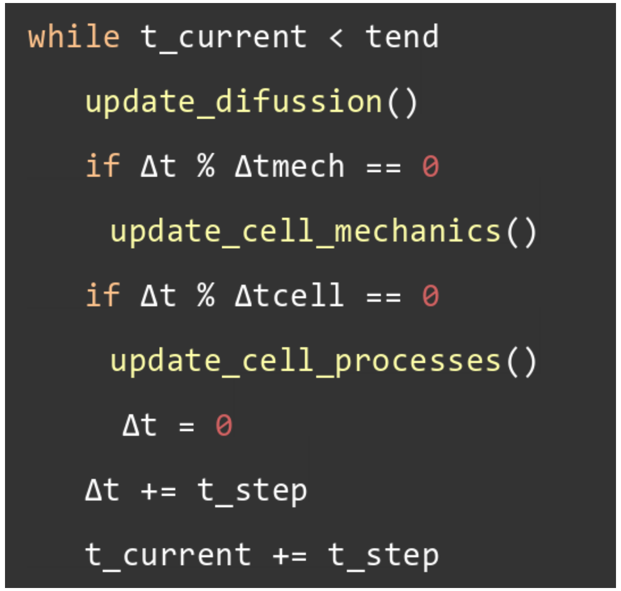
Further work: study how to migrate to hybrid architectures

- Combine CPU / GPU
    - Lower loops in GPU
    - Tools already published:

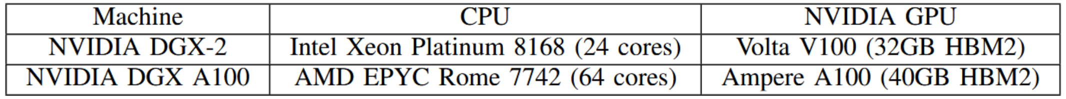
Time scales

- $\Delta t_{\mathbf{diff}}$: (diffusion/transport): 0.01 min ← PhysiCell-OpenACC
- $\Delta t_{\mathbf{mech}}$: (cell movement): 0.1 min ← FLAME-GPU
- $\Delta t_{\mathbf{cell}}$: (cell processes): 6 min
- $\Delta t_{\mathbf{signalling}}$: (Boolean simulation): 10 min

Simulation’s main loop

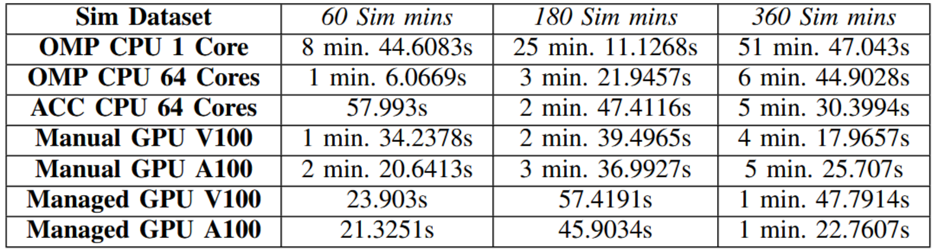


```
while t_current < tend
   update_difusion()
   if Δt % Δtmech == 0
    update_cell_mechanics()
   if Δt % Δtcell == 0
    update_cell_processes()
     Δt = 0
   Δt += t_step
   t_current += t_step
```



Machine	CPU	NVIDIA GPU
NVIDIA DGX-2	Intel Xeon Platinum 8168 (24 cores)	Volta V100 (32GB HBM2)
NVIDIA DGX A100	AMD EPYC Rome 7742 (64 cores)	Ampere A100 (40GB HBM2)

Table 1: Specifications of the nodes in the two systems



Sim Dataset	*60 Sim mins*	*180 Sim mins*	*360 Sim mins*
OMP CPU 1 Core	8 min. 44.6083s	25 min. 11.1268s	51 min. 47.043s
OMP CPU 64 Cores	1 min. 6.0669s	3 min. 21.9457s	6 min. 44.9028s
ACC CPU 64 Cores	57.993s	2 min. 47.4116s	5 min. 30.3994s
Manual GPU V100	1 min. 34.2378s	2 min. 39.4965s	4 min. 17.9657s
Manual GPU A100	2 min. 20.6413s	3 min. 36.9927s	5 min. 25.707s
Managed GPU V100	23.903s	57.4191s	1 min. 47.7914s
Managed GPU A100	21.3251s	45.9034s	1 min. 22.7607s

Table 2: Results Table

Richmond *et al.* (2010) *Briefings in Bioinformatics*, **11**, 334–347

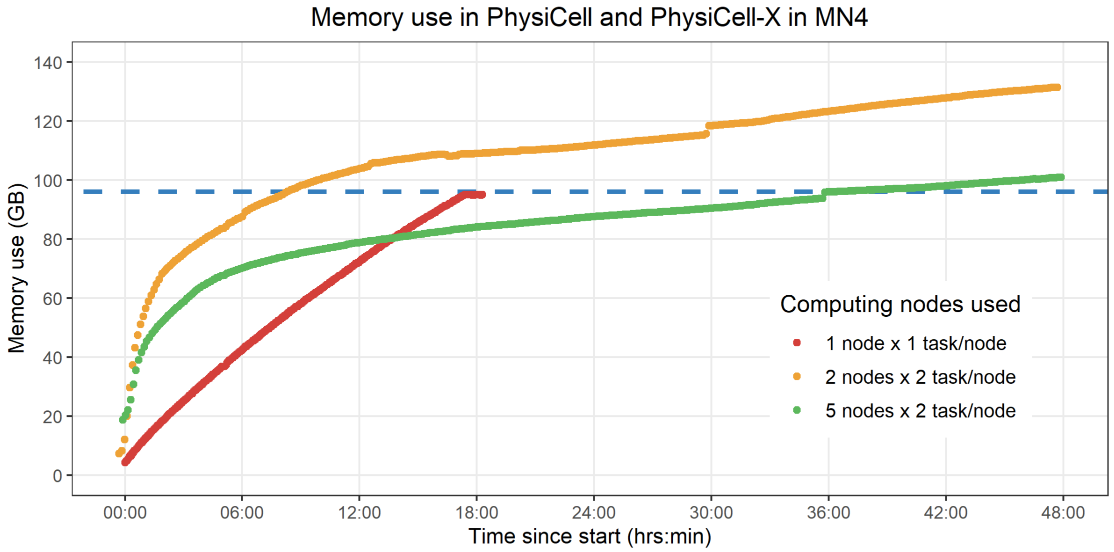
Stack *et al.* (2021) *arXiv:2110.13368 [cs]*

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Further work: prepare set-up to many patients and many replicates

- We would like to scale-up simulations
  - Currently running jobs of 10^4 cells
  - Looking forward to 10^9 cells

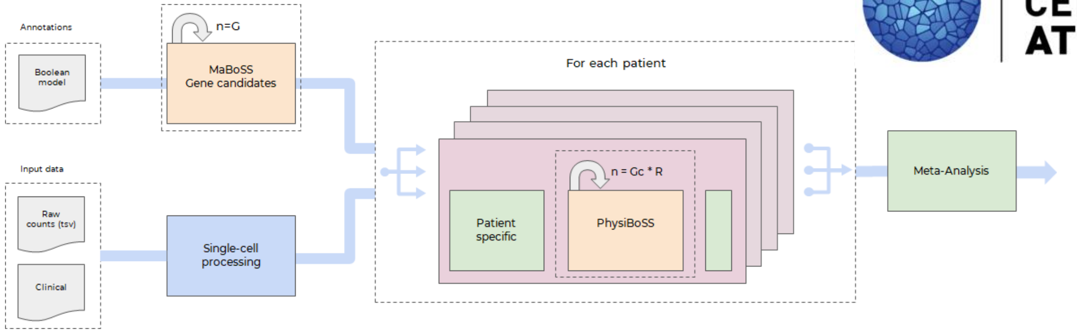
- Obstacles and potential solutions:
  - Memory-limited multiscale tools
    - MPI implementation



- We would like to simulate many more patients at once
  - Big datasets for each patient:
    - Storage & security problems

- Obstacles and potential solutions:
  - Different tools and codes
    - Use of pipelines and orchestrators
  - Homogeneous 3Dal set-up
    - Use of clinical images as initial set-ups
  - Parameter fitting evaluation
    - Model exploration techniques

HUMAN CELL ATLAS

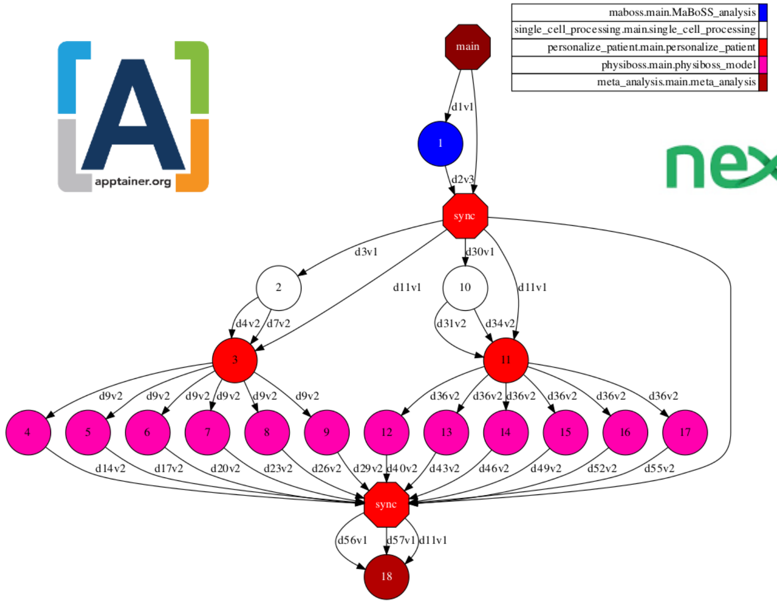


Further work: prepare infrastructure that fills the community needs

- PerMedCoE's initiatives open to the Computational Biology community:
  - Scaling-up success cases
  - Observatory of tools
  - Benchmark of similar tools
  - Repository of building blocks and workflows

- Obstacles and potential solutions:
  - Long-term maintenance costs
    - National funds? EU?

- Further outreach to the community:
  - Extensive use of pipelines and orchestrators
  - Offer alternatives to cloud computing
  - Offer training of Life scientists in HPC
    - Simulation tools
    - HPC performance tools
    - Offer testbeds for users' tools



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Arnau Montagud
Jose Carbonell
Miguel Ponce de León
Gaurav Saxena
Alfonso Valencia

Collaborators

Paul Macklin (Indiana U)
Laurence Calzone (Institut Curie)
Rosa Maria Badia (BSC)
Julio Sáez-Rodríguez (Heidelberg U)
Tommi Nyrönen (CSC)

Per Med CoE

HPC/Exascale Centre of Excellence in Personalised Medicine

www.permedcoe.eu

Follow us in social media:

www.linkedin.com/company/permedcoe
@permedcoe

The PerMedCoE project has received funding from the European Union’s Horizon 2020 research and innovation programme under the grant agreement Nº951773