Cloud-hosted mathematical models: links between education, research, and outreach

Paul Macklin, Ph.D.

Intelligent Systems Engineering Indiana University

July 25, 2019

Acknowledgements: Partners

- ► Colon cancer metabolism:
 - Stacey Finley (USC, U01 Contact PI)
- ► Colon cancer organoids:
 - Shannon Mumenthaler (USC, U01 PI)
- ► HPC via EMEWS:
 - Jonathan Ozik, Nicholson Collier, Justin Wozniak, Charles Macal (Argonne National Lab), Chase Cockrell,
 - Gary An (University of Vermont)
- Cloud-deployed PhysiCell models:
 - Gerhard Klimeck, Lynn Zentner, Martin Hunt, Steve Clark, others (NanoHUB Cyberplatform at Purdue)
 - Geoffrey Fox (IU PI, nanoBIO Node)
- ► PhysiCell:
 - Randy Heiland (IU)
 - alumni: S.H. Friedman (OKSI), A. Ghaffarizadeh (USC)

- PhysiCell performance refinements
 - GPU computing: Sunita Chandrasekaran (U. Delware)
 - RepastHPC extensions: Ozik, Collier (ANL)

► IU PhD students:

- John Metzcar (hypoxia, invasion),
- Yafei Wang (liver mets, nanotherapy)
- Furkan Kurtoglu (metabolism, hypoxia)
- Aneequa Sundus (cyanobacteria, synthetic multicellular systems)
- IU Undergraduates:
 - Metastasis: B. Fischer, D. Murphy, K. Konstantinopoulos, B. Duggan
 - Nanotherapy: T. Mahjan
 - Jupyter GUIs: E. Bower, D. Mishler, T. Zhang
 - PhysiCell technologies: E. Freeman

Acknowledgements: Funders

- ► NIH (current)
 - NIH CSBC U01 (1U01CA232137), Pls Finley* / Macklin / Mumenthaler
 - High-End instrumentation grant (1S10OD018495-01), PI Foster
 - HuBMAP CCF contract (OT2 OD026671), PI Börner
- ▶ NIH (past support that helped PhysiCell)
 - Provocative Questions grant (1R01CA180149), Pls Agus / Atala / Soker
 - NIH PS-OC center grant (5U54CA143907), Pls Agus / Hillis

► NSF:

- Engineered nanoBIO Hub (1720625), PI Fox. Co-PIs Douglas, Glazier, Macklin, Jadhao
- Cyanobacteria / Synthetic Biology (1818187), PI Kehoe, Co-PI Macklin
- ► Breast Cancer Research Foundation & JKTGF, PI Macklin
 - projects with PIs Agus, Gilkes, Peyton, Ewald, Newton, Bader









And thanks to IU Students!



About my lab's research

Multicellular systems problems

Interconnected systems and processes:

- ▶ Single-cell behaviors
- ▶ Cell-cell communication
- ► Physics-imposed constraints (e.g., diffusion)
- Systems of systems (e.g., immune system)

In disease, these systems fall out of balance.

Treatments target parts of these systems.

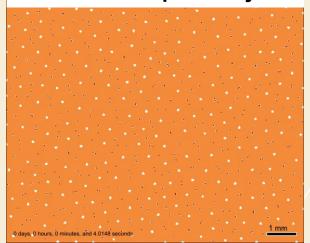
As with any complex system, changing one part can have surprising effects!

Modeling can help **understand** this system. This is **multicellular systems biology**.

If we can **control** these systems, we've arrived at **multicellular systems engineering**.



Metastatic seeding in 1 cm² of liver parenchyma



BioFVM: Simulating 3-D biotransport

Design goal: Simulate multiple diffusing substrates in 3D with desktops or single HTC/HPC nodes

Typical use: pO₂, glucose, metabolic waste, signaling factors, and a drug, on 10 mm³ at 20 μm resolution

Features:

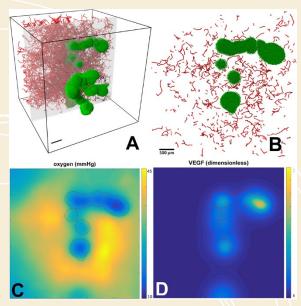
- ▶ Off-lattice cell secretion and uptake
- ▶ 2nd-order accurate (space), 1st-order accurate (time), numerically stable

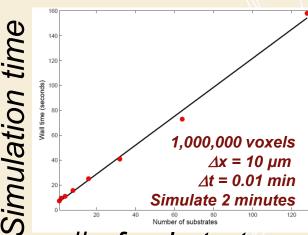
Method:

- Operator splitting, LOD, customized Thomas solvers, etc.
- ▶ Standard C++11, cross-platform
- ▶ OpenMP parallelization
- ▶ O(n) cost scaling in # substrates, # voxels
- ► Easy to simulate 5-10 substrates on 10⁶ voxels

Reference: Ghaffarizadeh et al., Bioinformatics (2016)

DOI: 10.1093/bioinformatics/btv730





PhysiCell: A multicellular framework

Design goal: Simulate 10⁶ or more cells in 2D or 3D on desktops or single HPC nodes

Features:

- ► Couple to 5+ 3D diffusion equations
- ▶ Off-lattice cell positions
- Mechanics-based cell movement
- ► Cell processes (cycling, motility, ...)
- Signal-dependent phenotype
- Can dynamically attach custom data and functions on a cell-by-cell basis
- Deployed from Raspberry Pi to Crays

Method:

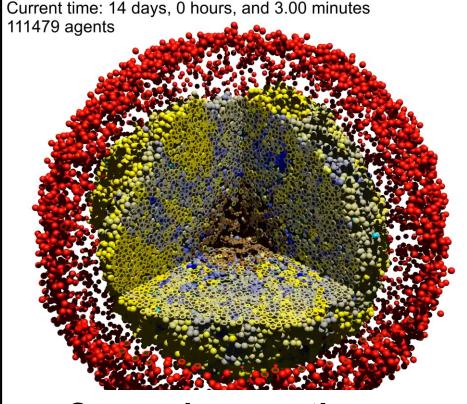
- ► Standard C++11, cross-platform
- OpenMP parallelization
- ightharpoonup O(n) cost scaling in # cells

Reference: Ghaffarizadeh et al., PLoS

Comput. Biol. (2018)

DOI: 10.1371/journal.pcbi.1005991

2019 PLoS
Computational Biology
Research Prize for
Public Impact



Cancer immunotherapy

[View on YouTube (4K)]



Try this model yourself! (2D)

nanohub.org/tools/pc4cancerimmune

Evolving lab structure

Goals for undergraduate research

- Students should be involved in the main research program
- ► Student involvement should enable new research directions
- ▶ Older students should help mentor younger students
 - sustainable research "ecosystem"
- ▶ Research results should feed back into education and outreach

Evolving lab structure & practices

► **Version 1** (Fall 2017)

5 undergrads + 1 PhD student

Format: Weekly coding session

- Mostly live coding to introduce modeling and programing
- Graduate student attended sessions and helped troubleshoot (like a teaching assistant)

What worked:

Students gained experience in ODEs, agent-based models, C++

What didn't work:

- Students needed hands-on time to learn and contribute individually
- ► Version 2 (Spring 2018)

6 undergrads

Format:

- Divide into teams of 2-3
- Meet with each team to instruct, mentor, set goals
- Students worked on their own between each weekly meeting

What worked:

Students started making individual contributions

What didn't work:

- Students felt isolated and wanted more interaction between groups
- Not scalable for me to meet each team each week

Evolving lab structure & practices

- Version 3 (Fall 2018-present) 10 undergrads + 3 PhD students + 1 staff Format:
 - Teams of 2-4 undergrads + 1 PhD student (or staff) co-mentor
 - Each week:
 - Early: I meet with co-mentors to set and assess goals
 - Mid: Co-mentors meet with their teams to guide the week
 - Late / weekend: Undergrads, co-mentors work on their own towards goals
 - All-hands "hackathon" each Friday:
 - One student group presents results (promotes cross-pollination)
 - Unstructured mentoring time by me for rest of session

What worked:

- Undergrads made individual contributions
- PhD students learned team management, gained "firepower"
- Undergrads practiced communication and presentation skills
- All lab members saw their part of the "big picture"

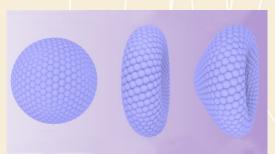
What didn't work:

- Students can't always perceive their learning progress. Need better metrics.
- Need more coordination time with co-mentors

case study: xml2jupyter

Project background

- ▶ NSF-funded nanoBIO project (NSF Award 1720625, PI Fox):
 - Create cloud-hosted tools for nanodesign in biology
 - nano: Simulate self-assembling nanoparticles (Co-PI Jadhao)
 - micro: Simulate cell-nanoparticle interactions (Co-PI Glazier)
 - macro: Simulate nanoparticle effects in multicellular systems (Co-Pl Macklin)
 - Host tools in nanoHUB.org
- ▶ Original macro goal: make a few feature-rich applications over 5 years

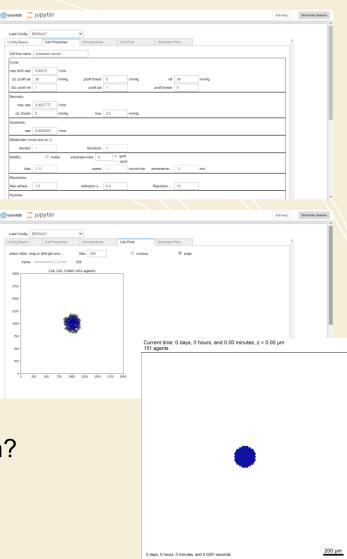


Simulated soft NP shape change

Vikram Jadhao lab Indiana University Jadhao et al. PNAS (2011)

Prior to undergraduate work

- ► Year 1: built a first prototype application
 - PhysiCell simulator (the "engine")
 - Command-line simulation
 - XML-based configuration file
 - outputs include SVG files (graphics) and matlab data
 - Jupyter notebook (the GUI)
 - Use widgets to set parameter values
 - Write the XML parameter file
 - C++ PhysiCell simulation reads XML and starts
 - C++ PhysiCell simulation saves data
 - Jupyter notebook reads and displays results.
 - Took about 4 months to prototype
- ► This works, but it's very manual.
 - Can we generalize it to any PhysiCell simulation?
 - Can we automate the process?
 - Could undergraduates play a key role?
 - What are the implications?



xml2jupyter: automating the process

Undergraduate team:

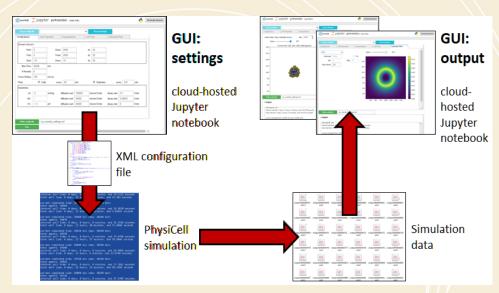
- ▶ Randy Heiland (Research Associate) lead team of 3 undergraduates, using "Version 3" mentoring
 - Daniel Mishler, Tyler Zhang, Eric Bower

Result: generalize to a toolkit with xml2jupyter. Build apps in hours instead of months.

- Heavy undergraduate researcher involvement
- nanoHUB workflow
 - xml2jupyter: ~seconds
 - parses XML config file to create Jupyter widgets
 - Boolean, integer, string, double parameters
 - additional Python scripts: ~minutes
 - convert PhysiCell + Jupyter to nanoHUB structure
 - User + nanoHUB staff: ~hours
 - Create GitHub repo and nanoHUB project
 - Build and deploy new app on nanoHUB

Advantages:

- No modification to simulation C++
- No modification to simulation XML required
- Optional "hinting" in XML attributes refines GUI
- Jupyter and Python mesh with Student interest
- xml2jupyter refinements help all Macro apps



XML+Jupyter architecture

Paper:

Heiland et al. (2019)

First paper at IU with undergrad authors!



xml2jupyter: Mapping parameters between XML and Jupyter widgets

Randy Heiland¹, Daniel Mishler¹, Tyler Zhang¹, Eric Bower¹, and Paul Macklin¹

1 Intelligent Systems Engineering, Indiana University

DOI: 10.21105/joss.01408

Software

- Review c
- Repository @
- Archive @

Submitted: 08 April 2019 Published: 01 July 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC-BY).

Summary

Jupyter Notebooks (Kluyver et al., 2016, Perkel (2018)) provide executable documents (in a variety of programming languages) that can be run in a web browser. When a notebook contains graphical widgets, it becomes an easy-to-use graphical user interface (GUI). Many scientific simulation packages use text-based configuration files to provide parameter values and run at the command line without a graphical interface. Manually editing these files to explore how different values affect a simulation can be burdensome for technical users, and impossible to use for those with other scientific backgrounds. xml2jupyter is a Python package that addresses these scientific bottlenecks. It provides a mapping between configuration files, formatted in the Extensible Markup Language (XML), and Jupyter widgets. Widgets are automatically generated from the XML file and these can, optionally, be incorporated into

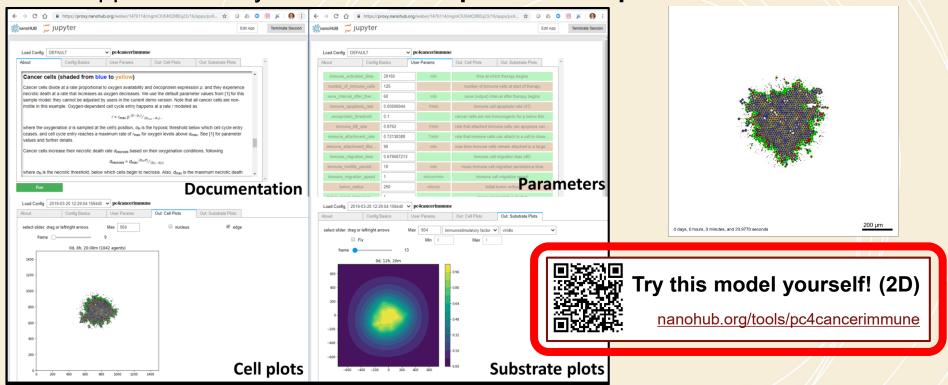
Benefits to Research and Outreach

Some early research benefits

- xml2jupyter increased research productivity for the NSF nanoBIO project:
 - Released 6 new simulation apps in Year 2
 - Opened the door for community contributions
 - Adrianne Jenner (University of Sydney) contributed a new app on virotherapy
- ► Cloud-hosted math models speed up multidisciplinary work
 - Biologist collaborators can try the models and give earlier feedback
 - "Close the loop" faster
- ► Cloud-hosted math models can help advertise your work
 - Post "try my model" link on Twitter, link to your paper
 - Attract new collaborators who can discover and try your work
- Ability to share models offers exciting possibilities for public outreach
 - Easy element for NSF "Broader Impact" section

Use case: Publication companion app

- ▶ We ran more over 10⁵ PhysiCell simulations in Ozik et al. (2019)
- ► We created a **publication companion app** for the **method section** to illustrate the model for reviewers and readers.
- We're now making this standard practice for all our modeling papers.
- ▶ The apps can easily be included in posters and presentations.



Future use case(?): math communication

- ▶ If it only takes a couple hours to make and deploy an app:
 - 1. Biologist collaborators can try the models are they're being built.
 - Do the rules make sense?
 - Can our collaborators suggest new rules?
 - 2. Outreach for published papers.
 - Share Companion Apps on social media to bring new readers to a paper.
 - 3. Modelers could share ideas and try them out
 - Here's what this model does. Any ideas?
 - Could it be a dialog? Model, and counter-model?
 - Could supplement current dialog as
- We're still working through the potential impact of fast, cheap, easy cloudhosted mathematical models.
- In a "post-scarcity" world, how could they be used?

Benefits to Education

Use case: Educational microapps

- Advanced undergraduate / graduate class in multicellular systems biology
- ▶ Students had difficulty intuiting biased random migration:
 - $\mathbf{v}_{\text{motility}} = b\mathbf{v}_{\text{bias}} + (1-b)\mathbf{v}_{\text{random}}$

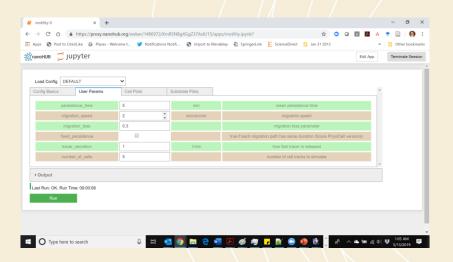


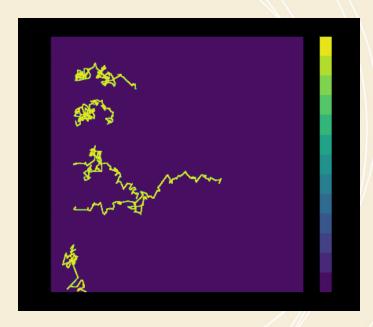
Try this model yourself! (2D)

https://nanohub.org/tools/motility

▶ Built microapp to create and visualize cell tracks with student-selected

parameters. Deployed by next lecture.

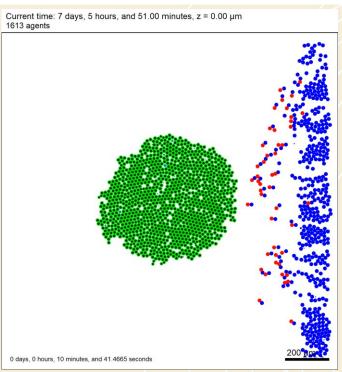




Use case: Research in the classroom

▶ Because it's easy to convert a research-grade model to a graphical app, we can quickly integrate new research into the classroom.

Synthetic cell-based therapy

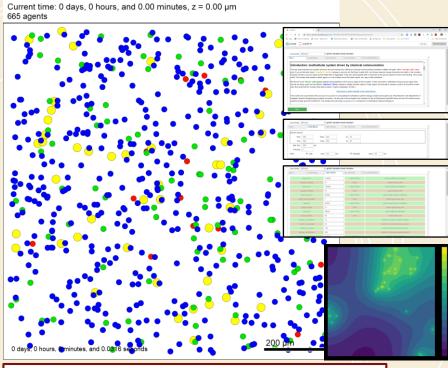




Try this model yourself!

https://nanohub.org/tools/pc4cancerbots

Chemokine demo





Try this model yourself!

https://nanohub.org/tools/pcisa

Use case: student project demos

- ▶ Student teams create simulation models in upper-level courses.
- ▶ We integrated xml2jupyter into the course in Spring 2019
 - Each team creates a cloud-hosted version of their model
 - Teams live demo their models in final presentations
 - Students can now try all the models
 - Bonus 1: Students get experience in cloud computing
 - Bonus 2: Students can use these as permanent portfolio pieces
- ▶ 4 **student portfolio apps** as part of ENGR-E 399/599 (Spring 2019)
 - Impact of "off-screen" immune expansion on immunologic systems
 - Biofilm model
 - Engineering cell rules for Turing-like pattern emergence
 - Wound healing model with mechanosensing

Takeaway messages

- ▶ Inclusion of undergraduates opened up new research possibilities
- ▶ Research-grade tools can be rapidly integrated into education
- ► Scalable undergraduate research requires:
 - Experimentation
 - Observation of the progress
 - Student feedback
 - Continuous refinement
- ► There are untapped opportunities for collaboration between "traditional" and educational research labs
- Accepted education paper on this work:
 - A. Madamanchi, R. Heiland, P. Macklin, and A.J. Magana. Students' Use of Metacognitive Skills in Undergraduate Research Experiences in Computational Modeling. *IEEE Frontiers in Education (FIE) Conference Cincinnati, OH USA*, 2019 (accepted).

Future Work

- ▶ Improved metrics and learning assessment for undergrad research
- ► Making multicellular modeling easier
 - Developing support for SBML in PhysiCell
 - Planning graphical model editing tools
- Growing the PhysiCell community.
 - We're happy to help you adapt PhysiCell (+HPC, +nanoHUB) to projects
 - We write letters of support for travel fellowships to host IU visitors:
 - Learn PhysiCell, write models together, and share interactive models on nanoHUB.
 - So far, Ph.D. and postdoctoral visitors from Australia and Europe
 - 2020: Hackathon and funds for ~2 long-term PhD student visitors
- ► Working to make HPC-powered models available to all
 - Using xml2jupyter to create interactive training materials
 - Graphical design of molecular-scale models
 - Improving performance and visualizations
 - Linking apps to HPC so the public can use cutting-edge 3-D models

Mathematical-Oncology.org/MathOnc20

#MathOnc20



- World Class Plenaries
- Tutorial Sessions
- Abstract Sessions
- Junior Investigator Leadership Training

March 11-14, 2020



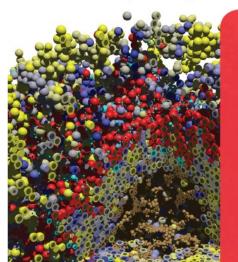
Data-Driven Multicellular Systems Biology

is a new thematic series and now accepting submissions.

Guest Editor: Paul Macklin

Given that decades of research has shed light and led to the characterization of most molecular parts of the cell, including the human genome - interest has expanded towards cellular dynamics and interactions within a cell, as in systems biology. However, many regulatory circuits go beyond a single cell and involve dynamic, complex crosstalk between multiple cells and their environments including tissue biomechanics - an extension of systems biology that is multicellular systems biology. With increasingly sophisticated experiments, coupled with large-scale computational models, the community has the potential to systematically test biological hypotheses to further our understanding of multicellular systems; however, challenges surrounding reproducibility exist.

GigaScience is proud to present a new thematic series, "Data-Driven Multicellular Systems Biology". This comprehensive and forward-looking series will highlight new advances, applications, and challenges, and serve to promote and improve data sharing and reproducibility within the multicellular systems biology field and beyond.



We encourage the submission of Research Articles and Technical Notes, as w as Data Notes, which are papers that focus on the description of interesting datasets, curated and hosted in our database, GigaDB. We also consider thought-provoking Commentary and Reviews in this area. Topics include, but are not limited to:

- Modelling and simulation utilizing multicellular data
- High-throughput microscopy (e.g., multiwell plates) and associated image processing
- Mathematical models applied to multicellular systems
- Machine learning and computational analysis techniques and results for big multicellular data
- New open tools that aim to improve reproducibility and interoperability
- Data standards including ontologies and curation.
- New big multicellular datasets (e.g., multiplex image data)
- Open repositories for multicellular systems data
- Synthetic multicellular systems biology and engineering
- Virtual machines and Containers



link

Looking for papers!

data-driven multicellular systems biology

in GigaScience

Macklin lab

MathCancer.org 🏏 @MathCancer



Postdoctoral Research Fellow Position Indiana University, Bloomington

Dr. Paul Macklin, the leader of a dynamic laboratory for open source computational biology and oncology, seeks an interdisciplinary postdoctoral fellow for a project in computational data-driven modeling of breast cancer metastasis, jointly with experimental collaborators at Johns Hopkins University. Work in the MathCancer lab includes agent-based modeling of multicellular systems, multi-substrate diffusion solutions, cancer modeling, high-throughput computing, machine learning, and digital pathology, with a heavy emphasis in building open source technologies that benefit the entire scientific community. Learn more at http://MathCancer.org

The fellow will join a dynamic, interdisciplinary, and collaborative research environment with ample opportunities for career development and mentoring. IU SICE is first of its kind and among the largest in the country, with more than 125 full-time faculty and more than 1000 graduate students. SICE has received recognition as a "top-ten program to watch" (Computerworld) thanks to its excellence and leadership in academic programs, interdisciplinary research, placement, and outreach. IUB is also renowned for its top-ranked music school, performing and fine arts, and historic campus. Bloomington is located in the wooded rolling hills of southern Indiana, 50 miles south of Indianapolis. It is a culturally thriving college town with a moderate cost of living, cycling traditions, active lifestyle, and natural beauty.

Minimum Qualifications: A Ph.D. in Mathematics, Computer Science, Informatics, Engineering, or related discipline is required, with strong background in mathematical biology and strong C/C++ skills. Prior expertise in agent-based modeling (especially PhysiCell) is a plus. ABD (all but dissertation) candidates may apply but will need the Ph.D. prior to start. Applicants should have a record of independent research and publications along with the ability to participate in interdisciplinary collaborations.

Salary: Commensurate with qualifications and experience

Appointment: 3-year non-tenure-track appointment with potential for extension subject to performance and funding

Apply Online: https://indiana.peopleadmin.com/postings/8144

Special Instructions: Please upload a letter of interest, CV, representative publications, and contact information for three references at the application link above. Questions may be sent to Paul Macklin (macklinp@iu.edu).

For Fullest Consideration Apply By: August 31, 2019

Ideal Start Date: October 1, 2019 or sooner (negotiable)

Indiana University is an equal employment and affirmative action employer and a provider of ADA services. All qualified applicants will receive consideration for employment without regard to age, ethnicity, color, race, religion, sex, sexual orientation, gender identity or expression, genetic information, marital status, national origin, disability status or protected veteran status.



[link]

Postdoc!

PhysiCell models of metastatic breast cancer

Macklin lab

RING

Some references

- ▶ PhysiCell Method paper -- PLoS Computational Biology
 - A. Ghaffarizadeh, R. Heiland, S.H. Friedman, S.M. Mumenthaler, and P. Macklin. PhysiCell: an open source physics-based cell simulator for 3-D multicellular systems. PLoS Comput. Biol. 14(2):e1005991, 2018. DOI: 10.1371/journal.pcbi.1005991.
- ▶ PhysiCeII + High-Throughput Computing (HTC) -- BMC Bioinformatics
 - J. Ozik, N. Collier, J. Wozniak, C. Macal, C. Cockrell, S.H. Friedman, A. Ghaffarizadeh, R. Heiland, G. An, and P. Macklin. High-throughput cancer hypothesis testing with an integrated PhysiCell-EMEWS workflow. BMC Bioinformatics 19:483, 2018.
 DOI: 10.1186/s12859-018-2510-x.
- ▶ PhysiCeII + HPC + machine learning -- Molecular Systems Design and Engineering
 - J. Ozik, N. Collier, R. Heiland, G. An, and P. Macklin. Learning-accelerated Discovery of Immune-Tumour Interactions. Molec. Sys. Design Eng., 2019.
 DOI: 10.1039/c9me00036d
- ► PhysiCeII + Boolean Networks -- Bioinformatics
 - G. Letort, A. Montagud, G. Stoll, R. Heiland, E. Barillot, P. Macklin, A. Zinovyev, and L. Calzone. PhysiBoSS: a multi-scale agent based modelling framework integrating physical dimension and cell signalling. *Bioinformatics* 35(7):1188-96, 2019.
 DOI: 10.1093/bioinformatics/bty766.
- ► Automatically create Jupyter-based GUIs -- Journal of Open Source Software
 - R. Heiland, D. Mishler, T. Zhang, E. Bower, and P. Macklin. xml2jupyter: Mapping parameters between XML and Jupyter widgets. J. Open Source Software 4(39): 1408, 2019. DOI: 10.21105/joss.01408