Open Source AI/ML for infectious disease research

Open Research Forum, University of Reading
09.03.2022

Gemma Turon, gemma@ersilia.io
Ersilia Open Source Initiative
https://ersilia.io
I’m here today...

- **Software Sustainability Fellowship**
  - Improve computational practices in research software
    
    *Reusability of AI/ML models for biomedical research*

- **Open Life Sciences Program**
  - Mentoring for Open Science Ambassadors
    
    *Improve the documentation and accessibility of Ersilia’s tools*

- **Digital Infrastructure Incubator**
  - Supporting open source project leaders implementing best practices in sustainability, governance, and community health
    
    *Community building tools and governance models*
Edoardo Gaude, PhD
Co-founder & Trustee
Trained as molecular biologist at Cambridge University, UK
Co-founder of PockIt

Miquel Duran-Frigola, PhD
Co-founder & CSO
Trained as a computational chemist at IRB Barcelona, Spain

Gemma Turon, PhD
Co-founder & CEO
Trained as molecular biologist at IRB Barcelona, Spain
Our Mission

*Strengthen the research capacity in Low and Middle Income Countries*
Land area

Data from WHO Reports 2019
DALY – Communicable Diseases

Data from WHO Reports 2019
Scientific Publications

Data from WHO Reports 2019
Western bias in Biomedical Research

Disease burden

Drugs in development

Non-communicable
Communicable, maternal & neonatal
Injuries
Other
Western bias in Biomedical Research

Malaria

• Causes 0.5 million deaths a year (mostly amongst children)
• 95% of the new malaria cases are detected in Africa
• Resistance to front-line treatment (Artemisinin-combination therapies) is widespread in South-East Asia and first reports in Africa

a) Malaria case incidence (per 1000 population at risk) b) mortality rate (deaths per 100000 population at risk)  
WHO Malaria report 2021
Western bias in Biomedical Research

Tuberculosis

- Leading cause of death from a single infectious agent (before COVID) – around 1.1.5 million deaths per year (including HIV+ patients)
- 85% of the deaths occur in the WHO Africa and South-East Asia regions
- TB death incidence is back to levels of 2017

Top causes of death worldwide in 2019

Deaths from TB among HIV-positive people are shown in grey.

- Ischaemic heart disease
- Stroke
- Chronic obstructive pulmonary disease
- Lower respiratory infections
- Neonatal conditions
- Trachea, bronchus, lung cancers
- Alzheimer disease and other dementias
- Diarrhoeal diseases
- Diabetes mellitus
- Kidney diseases
- Cirrhosis of the liver
- Road injury
- Tuberculosis
- Hypertensive heart disease
- Colon and rectum cancers
- Stomach cancer
- Self-harm
- Falls
- HIV/AIDS
- Breast cancer

Global trend in case notifications of people newly diagnosed with TB, 2016–2020

WHO Tuberculosis Report 2021
Free & Open Source
Real-time code sharing
Permissive licenses
No patents
Reproducibility

In-Country Research
Avoid “helicopter research”
Science led by local institutes
Implementation in situ

Sustainable Collaborations
Capacity building activities
Identify & train local champions
AI/ML with low resources
Ersilia’s life cycle

Open Source tools for biomedical research

Researchers in the field

Data collection
Project support 

in situ implementation

Open Data and Open Source

Use Ersilia’s infrastructure

Broader scientific community

Apply Ersilia’s assets

User-friendly deployment
Our goal: ready to use AI

Input: e.g. natural products

Output: e.g. activity against a pathogen

- Raw data
- Processed data
- Model training
- Testing and tuning
- Validation
- External validation
- Deployment
AI/ML from the literature

Ersilia “bundles” a model developed by others

In-House AI/ML

Ersilia trains an AI/ML model based on data

Antibiotic activity

E. coli

Stokes et al, 2020

Chemoprotective antimalarials

Antonova-Koch et al, 2018
AI/ML in collaboration

* Ersilia trains an AI/ML model based on partner’s data

Your awesome project
You and Ersilia, 202

Your question  Our answer
The Ersilia Model Hub

User Query
Active

Author
GitHub repository
Summary / Applicability
The Ersilia Model Hub – How to

https://github.com/ersilia-os/ersilia

1. Ersilia installation in local computer
2. Selection of model of interest:
   - 40 publicly available models – browsable catalog
3. Use a command line interface to download model from our repository
4. Select the model api (predict, calculate…) and input the molecule (or list of molecules) of interest
5. Close model

```bash
ersilia fetch chemprop-antibiotic
ersilia api predict -i "C1=C(SC(=N1)SC2=NN=C(S2)N)[N+](=O)[O-]"
ersilia close
```

*Disclaimer: the EMH is in testing mode*
Applications of the Ersilia Model Hub

Implementation of a virtual drug screening cascade

• Where: H3D Centre, Cape Town (South Africa)
• What: AI/ML modelling of drug screening assays for antimalarial and antituberculosis drug discovery

- Compound prioritization for synthesis
- Reduction of attrition rates
- De novo design of new candidates
Applications of the Ersilia Model Hub

Generation of new antimalarial leads

- Where: Open Source Malaria Consortium (Prof. Todd, UCL)
- What: computer-based optimization of a chemical series with potent activity against malaria

Chemical Series Origin:
*Pfizer → MMV → OSM*

**OSM SERIES 4**
*The Triazolopyrazine (TP) Series*
(Representative Compound)
Already investigated in Pharma/CRO
Promising PK
Possible PfATP4 activity

https://github.com/opensourcemalaria
Open Source Malaria Consortium
Ersilia’s Roadmap

November 2020
- Charity Incorporation

Today
- First release of the EMH
- Early adopters

End of 2022
- Stable version of EMH
- Users in several institutions
- OS Contributors

End of 2023
- Serving >1000 beneficiaries
- Established OS repository for biomedicine
- Application of our tools to other fields
User experience

- Testing, debugging and improving CLI
- Offer online predictions via cloud services
- Adding a user interface

Scalability

- Zairachem: an automated end-to-end ML pipeline for chemistry
- Model encryption to leverage IP-sensitive datasets (Merck funded)
- Facilitating third party model deposition

Dissemination

- Seminars and conferences
- Scientific publications
- Implementation in situ with our partners
- Training and workshops in underserved countries
Ersilia and Openness

Strengthen the research capacity in Low and Middle Income Countries by developing and deploying AI/ML tools in collaboration with scientists in low-resourced settings

• Make assets already developed more accessible
• Re-use published data & encourage sharing of private datasets
• Avoid reinventing the wheel
• Escaling collaborations between scientists from different institutes
• We are also working with low resources – Open Access, Open Source
• Find alternatives to the traditional drug discovery models, particularly in diseases with low revenue (MMV, DNDi, OSM, M4iD)


https://ersilia.io                      hello@ersilia.io                      https://medium.com/ersiliaio
Ersilia and Openness

How do we try to ensure we stay in the Open Science domain?

• Incorporated as a non-profit organisation
• We accompany our projects with training and dissemination activities to ensure open means accessible
• At the organizational level:
  • Open Code via repositories
  • Real-time financial status
  • Grant applications disclosed
  • Governance and strategic decision making – wip
  • Identifying avenues to work with proprietary data for the public benefit - wip

https://ersilia.io  hello@ersilia.io  https://medium.com/ersiliaio
Take Home Messages

• We are a **Non Profit Organisation** with the mission to democratise access to AI/ML tools for biomedical research
• We are building an **international** network of collaborators
• We combine **remote working** and **on-site** project development and capacity building
• All our assets are **open-source**
• We work at the **intersection** between academia, start-ups and pharmaceutical companies
• We **welcome** new contributors and collaborators
Open Source AI/ML for infectious disease research

Open Research Forum, University of Reading
09.03.2022

Gemma Turon, gemma@ersilia.io
Ersilia Open Source Initiative
https://ersilia.io