

Ersilia Open Source Initiative: AI for Global Health

15th March 2023

Gemma Turon, PhD, @TuronGemma, gemma@ersilia.io
Ersilia Open Source Initiative, @ersiliaio, hello@ersilia.io
<https://ersilia.io>



Ersilia

About the Ersilia Open Source Initiative

Towards a hub of machine learning models

Open Source, why?

Current projects

About the Ersilia Open Source Initiative

Towards a hub of machine learning models

Open Source, why?

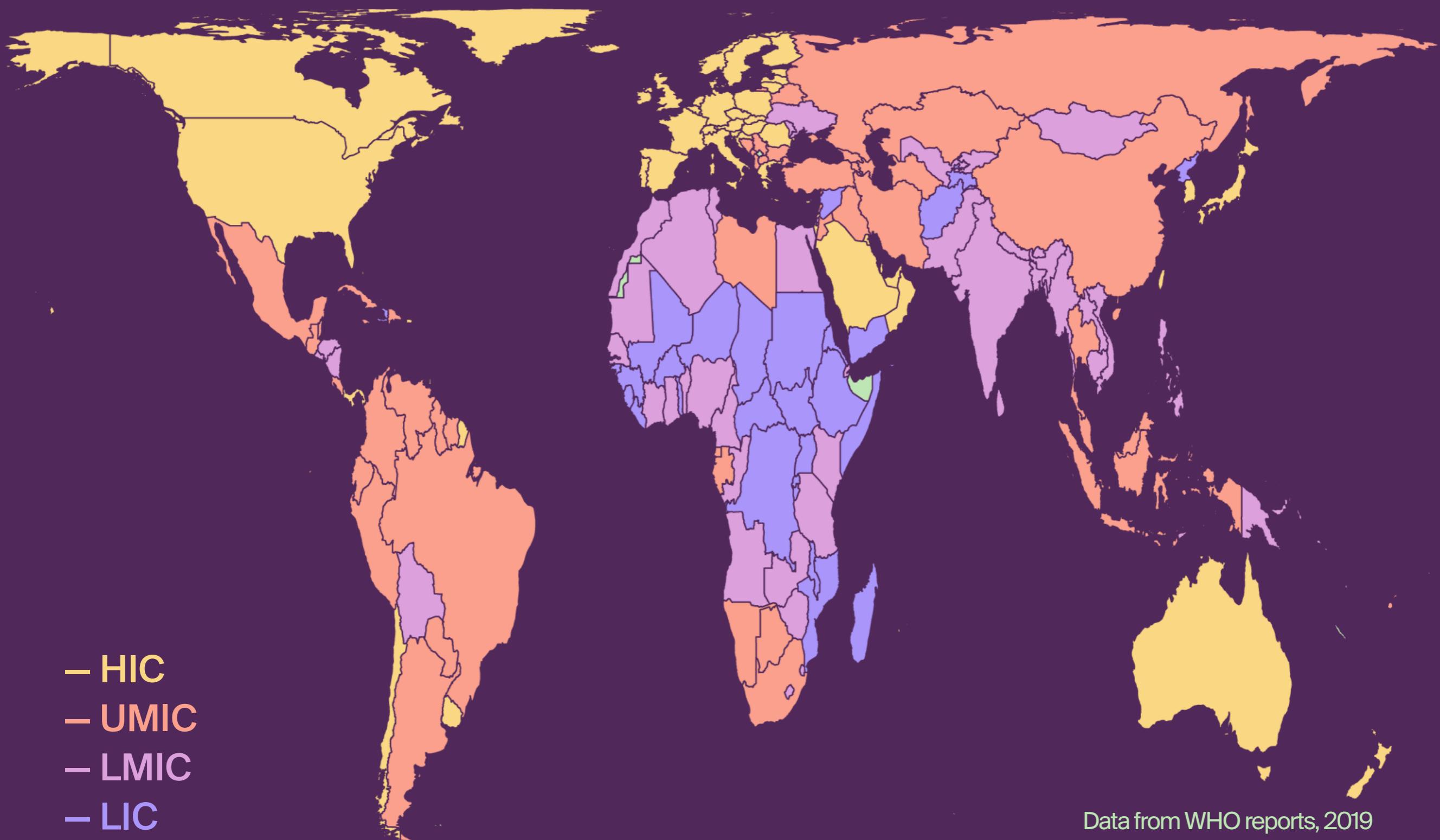
Current projects



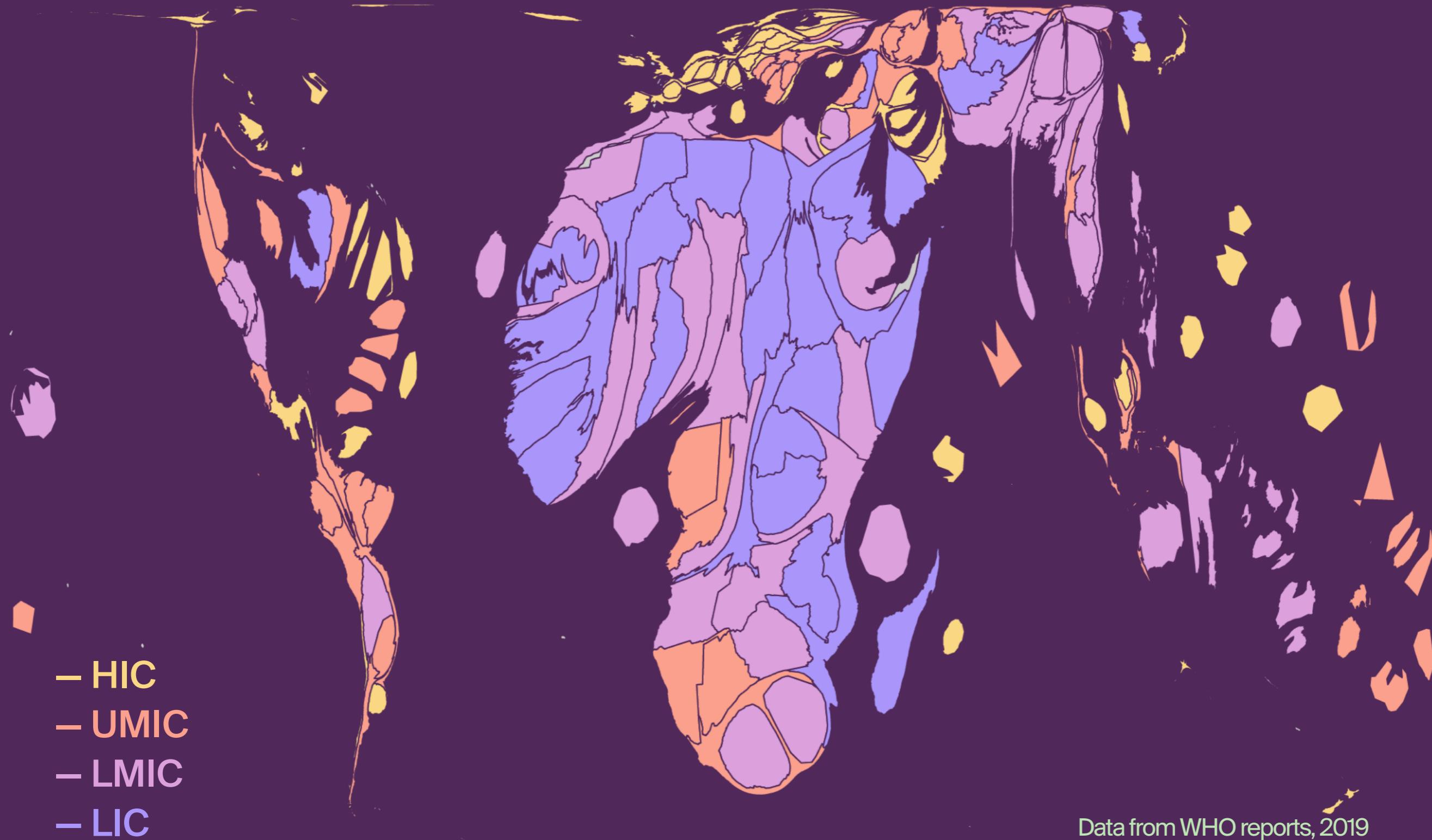
We are a small and young tech non-profit aimed at reducing inequality in global health.

We equip laboratories in LMICs with artificial intelligence tools for infectious disease research.

Land area



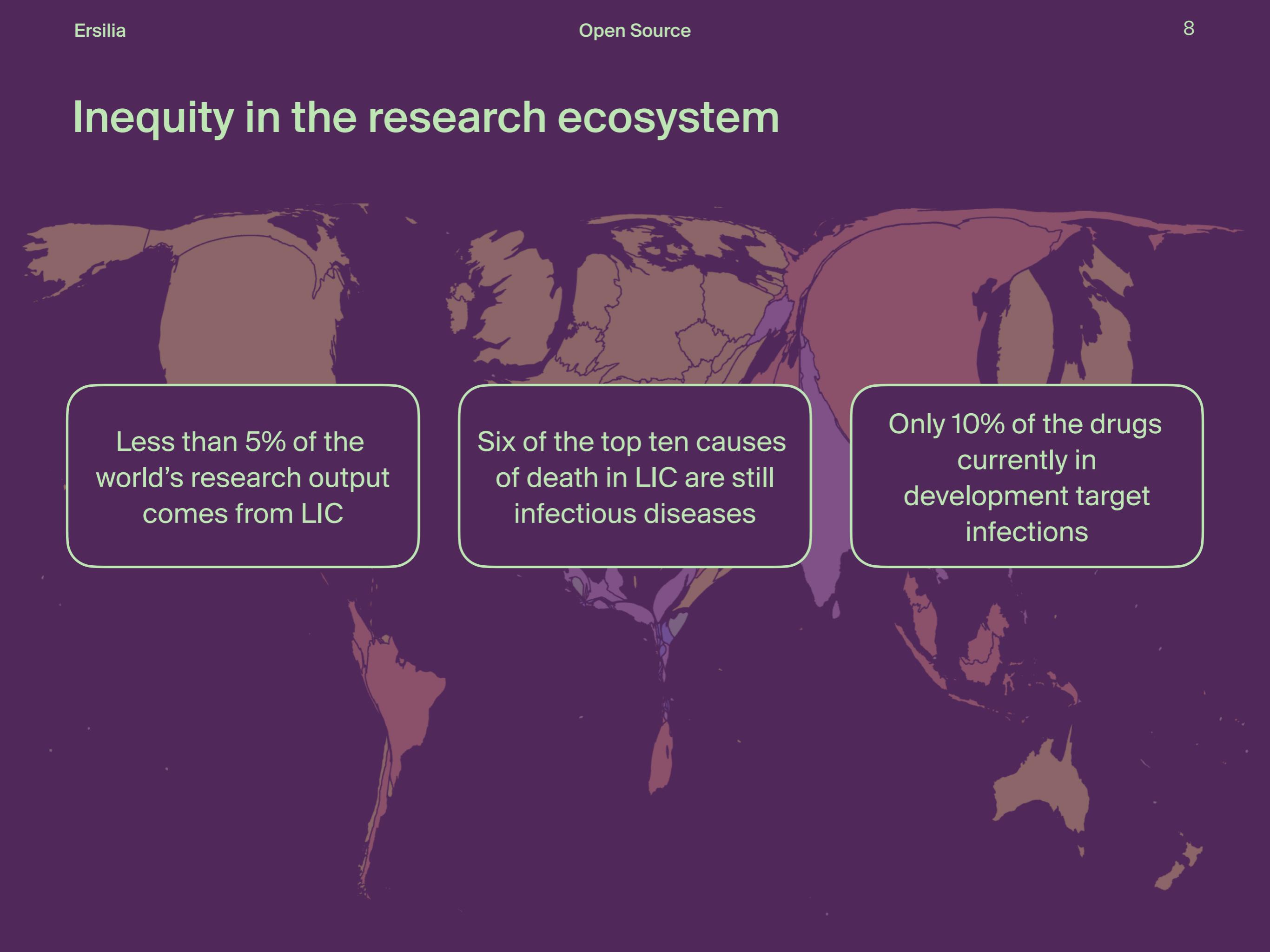
Communicable disease DALY



Science & engineering publications



Inequity in the research ecosystem



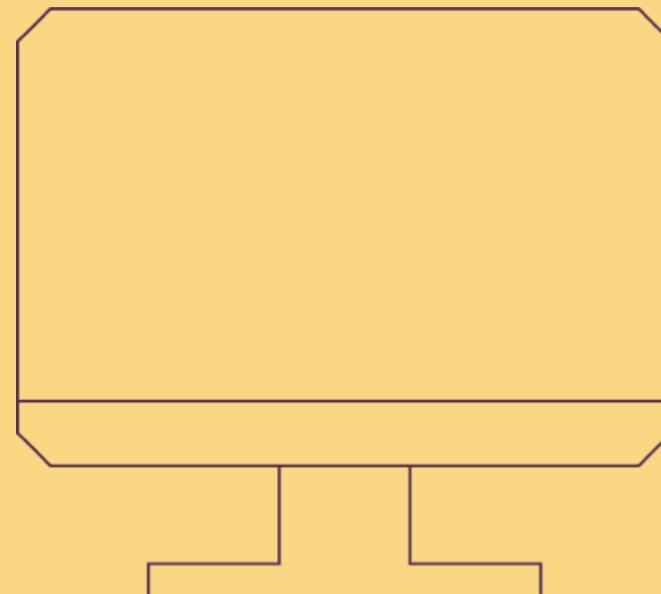
Less than 5% of the world's research output comes from LIC

Six of the top ten causes of death in LIC are still infectious diseases

Only 10% of the drugs currently in development target infections

Free & Open Source

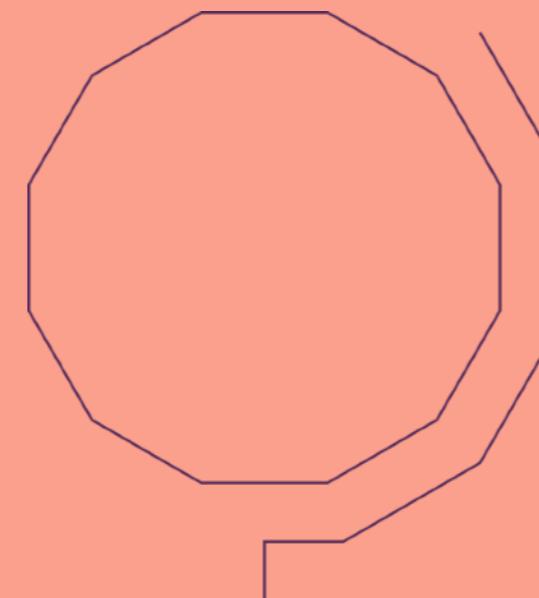
Public code
Open access licenses
No patents



Open Source

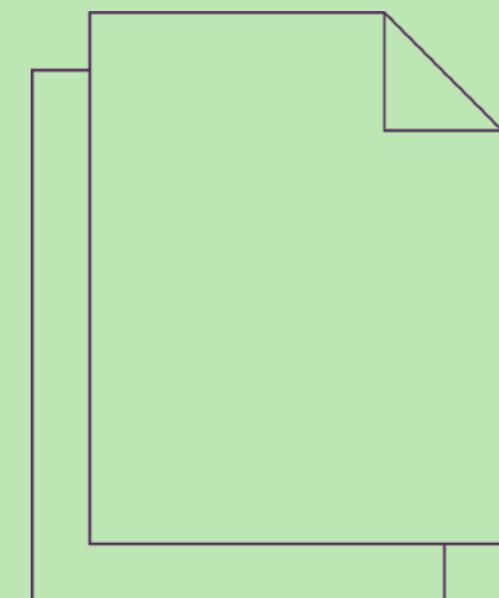
In-Country Research

Science led by local institutes
Implementation *in situ*



Sustainable Collaborations

Capacity building workshops
Train local scientists
AI with low resources



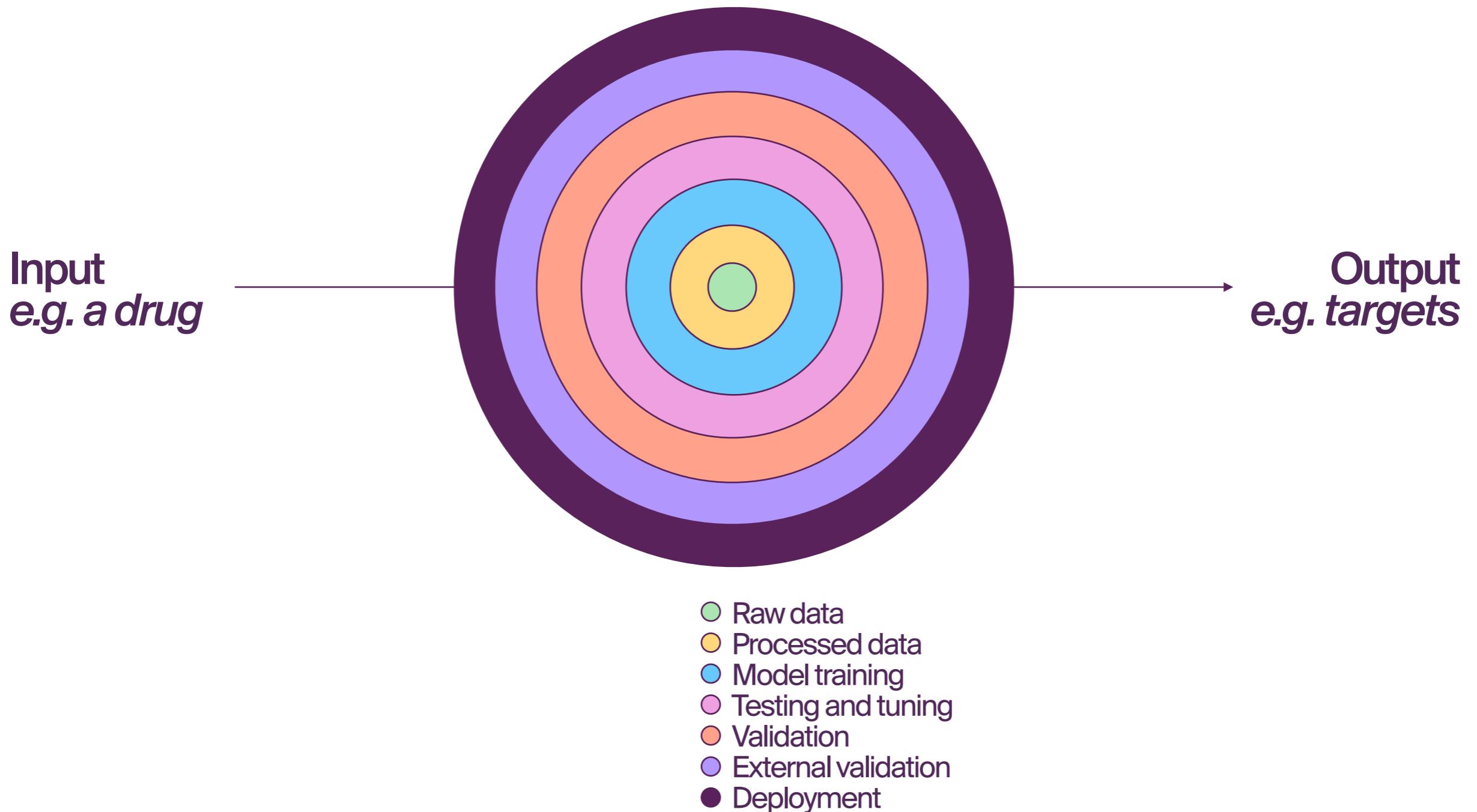
About the Ersilia Open Source Initiative

Towards a hub of machine learning models

Open Source, why?

Current projects

Our goal: to provide ready-to-use AI models



AI from the literature

- Ersilia ‘bundles’ a good model developed by others

Cell

Article

A Deep Learning Approach to Antibiotic Discovery

Jonathan M. Stokes,^{1,2,3} Kevin Yang,^{3,4,10} Kyle Swanson,^{3,4,10} Wengong Jin,^{3,4} Andres Cubillos-Ruiz,^{1,2,5} Nina M. Donghai,⁶ Crisostomo Machair,⁷ Shawn French,⁸ Lindsey A. Carras,⁹ Zohar Barak-Cohen,^{1,7} Victoria M. Tran,² Ayush Chaudhary-Pepe,⁸ Alvin H. Badran,⁸ Ian W. Andrews,^{1,2,8} Emma J. Chory,^{1,2} George M. Church,^{6,7,8} Eric D. Brown,⁹ Tomasz S. Jakubowski,^{1,4} Regina Barzilay,^{3,4,8,10} and James J. Collins^{1,2,3,6,8,10}

¹Department of Biological Engineering, Synthetic Biology Center, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
²Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
³Machine Learning for Pharmaceutical Discovery and Synthesis Consortium, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
⁴Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
⁵Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA
⁶Department of Biochemistry and Biomedical Sciences, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON L8N 3Z5, Canada
⁷Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
⁸Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA 02139, USA
⁹Abdul Latif Jameel Clinic for Machine Learning in Health, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
¹⁰These authors contributed equally
¹¹Lead Contact
^{Correspondence:} regina@csail.mit.edu (R.B.), jmcj@mit.edu (J.J.C.)
<https://doi.org/10.1016/j.cell.2020.01.021>

SUMMARY

Due to the rapid emergence of antibiotic-resistant bacteria, there is a growing need to discover new antibiotics. To address this challenge, we trained a deep neural network capable of predicting molecules with antibacterial activity. We performed predictions on multiple chemical libraries and discovered a molecule from the Drug Repurposing Hub—halicin—that is structurally divergent from conventional antibiotics and displays bactericidal activity against a wide phylogenetic spectrum of pathogens, including *Mycobacterium tuberculosis* and carbapenem-resistant Enterobacteriaceae. Halicin also effectively treated *Clostridioides difficile* and pan-resistant *Acinetobacter baumannii* infections in murine models. Additionally, from a discrete set of 23 empirically tested predictions from >107 million molecules curated from the ZINC15 database, our model identified eight antibacterial compounds that are structurally distinct from known antibiotics. This work highlights the utility of deep learning approaches to expand our antibiotic arsenal through the discovery of structurally distinct antibacterial molecules.

INTRODUCTION

Since the discovery of penicillin, antibiotics have become the cornerstone of modern medicine. However, the continued effi-

cacy of these essential drugs is uncertain due to the global dissemination of antibiotic-resistance determinants. Moreover, the decreasing development of new antibiotics in the private sector that has resulted from a lack of economic incentives is exacerbating this already dire problem (Brown and Wright, 2016; PEW Trusts, 2019). Indeed, without immediate action to discover and develop new antibiotics, it is projected that deaths attributable to resistant infections will reach 10 million per year by 2050 (O’Neill, 2014).

Historically, antibiotics were discovered largely through screening soil-dwelling microbes for secondary metabolites that prevented the growth of pathogenic bacteria (Clardy et al., 2006; Wright, 2017). This approach resulted in the majority of clinically used classes of antibiotics, including β -lactam, aminoglycosides, polymyxins, and glycopeptides, among others. Semi-synthetic derivatives of these scaffolds have maintained a viable clinical arsenal of antibiotics by increasing potency, decreasing toxicity, and sidestepping resistance determinants. Entirely synthetic antibiotics of the pyrimidine, quinolone, oxazolidinone, and sulfa classes have also found prolonged clinical utility, and continue to be optimized for the same properties.

Unfortunately, the discovery of new antibiotics is becoming increasingly difficult. Natural product discovery is now plagued by the dereplication problem, wherein the same molecules are being repeatedly discovered (Cox et al., 2017). Moreover, given the rapid expansion of chemical spaces that are accessible by the derivatization of complex scaffolds (Ortholand and Ganeshan, 2004), engineering next-generation versions of existing antibiotics results in substantially more failures than leads. Therefore, many antibiotic discovery programs have turned to screening large synthetic chemical libraries (Tomasini et al., 2019). However, these libraries, which can contain hundreds of thousands

688 Cell 180, 688–702, February 20, 2020 © 2020 Elsevier Inc.

Antibiotic activity *E.coli*
Stokes et al., 2020

Halicin

Active!

In-house AI

- Ersilia ‘trains’ a model based on good data

RESEARCH

RESEARCH ARTICLE

ANTIMALARIALS

Open-source discovery of chemical leads for next-generation chemoprotective antimalarials

Yevgeniya Antonova-Koch¹, Stephan Meister^{1*}, Mathew Abraham¹, Madeline R. Lath¹, Sabine Ottile², Amanda K. Lukens^{2,3}, Tomoya Sakata-Kato², Mann Vanarschot⁴, Edward Owen⁵, Juan Carlos Jado⁶, Steven P. Maher^{6,7}, Jason Calla⁸, David Plouffe⁹, Yang Zhong⁹, Kaiheng Chen⁹, Victor Chaumeau^{9,10}, Amy J. Conway^{6,11}, Case W. McNamara¹², Maureen Ibanez¹², Kerstin Gagaring¹², Fernando Neria Serrano¹³, Korina Ezrhi¹⁴, Cullin McLean Taggard¹⁴, Andrea L. Cheung¹⁴, Christie Lincoln¹⁴, Biniam Ambachew¹⁵, Melanie Koullier¹², Dionicio Siegel¹⁵, François Nosten^{10,16}, Dennis E. Kyle^{6,7}, Francisco-Javier Gamez¹², Yingyao Zhou⁸, Mannel Llina^{5,14}, David A. Fidock⁴, Dyan F. Wirth^{2,3}, Jeremy Burrows¹², Brice Campu¹², Elizabeth A. Winzeler^{1,12,13}

To discover leads for next-generation chemoprotective antimalarial drugs, we tested more than 500,000 compounds for their ability to inhibit liver stage development of luciferase-expressing *Plasmodium* spp. parasites (681 compounds showed a half-maximal inhibitory concentration of less than 1 micromolar). Cluster analysis identified potent and previously unreported scaffold families as well as other series previously associated with chemoprophylaxis. Further testing through multiple phenotypic assays that predict stage-specific and multispecies antimalarial activity distinguished compound classes that are likely to provide symptomatic relief by reducing asexual blood stage parasitemia from those which are likely to only prevent malaria. Target identification by using functional assays, *in vitro* evolution, or metabolic profiling revealed 58 mitochondrial inhibitors but also many chemotypes possibly with previously unidentified mechanisms of action.

Primary screening results

Previous high-throughput screens for antimalarial compounds have generally focused on the *ABC* sector, which can be defined as either the asexual blood stage or the liver stage (6–8). To obtain hits with possible protective activity, blood stage active compounds have been further retested against malaria hepatic stages, often using sporozoites from the rodent malaria species *Plasmodium yoelii* or *Plasmodium berghei* (8,10). Because there are no suitable methods for asexually culturing sporozoites (which are responsible for liver-stage infection), this testing has required the production of infected laboratory-reared mosquitoes and hand dissection of the sporozoite-infected salivary glands from mosquito thoraces. Despite this complex challenge, thousands of compounds have been examined by using this general workflow progression, leading to previously unidentified chemoprotective candidates, including KAF158 (12), KDU691 (12), DDD107498 (13), and BHD3444 (14). However, this approach would not reveal compounds that might protect from infection without affecting

Malaria remains a devastating disease, with 216 million annual cases and 443,000 deaths, primarily in children under 5 years old. Malaria treatment relies primarily on drugs that target the disease-causing asexual blood stages (ABS) of the parasite *Plasmodium falciparum*. These drugs include the 4-aminoquinolines piperaquine and amodiaquine, the antifolates pyrimethamine and sulfadiazine, and the endoperoxides artemisinin and its derivatives artesunate, artemether, and dihydroartemisinin (1). Artemisinin-based combination therapies (ACTs, such as artemether-lumefantrine) are being used worldwide as a well-established tool to prevent malaria caused by different *Plasmodium* spp. To prevent malaria, travelers may take oral atovaquone plus proguanil, artemisinins, and ACT treatment failures are rising (2). In anticipation of eventual widespread ACT failure, there has been a focused and coordinated effort to place new antimalarial candidates into the drug development pipeline (www.mnv.org/research-development/rd-portfolio) (3). However, in vitro resistance can be generated for most of these new classes in ABS parasites (4), suggesting that the antimalarial pipeline will require continuous replenishment.

An alternative approach is to create drugs that prevent malaria by inhibiting parasites during their initial stage of development in the liver before their first symptom of infection (symptomatic blood-stage infection). Using chemotherapy for prophylaxis is a well-established tool to prevent malaria caused by different *Plasmodium* spp. To prevent malaria, travelers may take oral atovaquone plus proguanil, artemisinins, and ACT treatment failures are rising (2). In anticipation of eventual widespread ACT failure, there has been a focused and coordinated effort to place new antimalarial candidates into the drug development pipeline (www.mnv.org/research-development/rd-portfolio) (3). However, in vitro resistance can be generated for most of these new classes in ABS parasites (4), suggesting that the antimalarial pipeline will require continuous replenishment.

Although estimated malaria mortality rates have decreased by 47% worldwide since 2000 (2), resistance has emerged to the first-line treatments (2).

ARTICLE IN CONTENTS

¹School of Medicine, University of California, San Diego, 9500 Gilman Drive 0720, La Jolla, CA 92093, USA. ²Harvard T. H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA. ³Department of Microbiology and Immunobiology, Division of Immunobiology and Molecular Medicine, Harvard University, Cambridge, MA 02137, USA. ⁴Department of Biochemistry and Molecular Biology and Center for Malaria Research, New York, NY 10032, USA. ⁵Department of Biochemistry and Molecular Biology and Center for Malaria Research, New York, NY 10032, USA. ⁶Center for Tropical and Emerging Global Diseases, University of Georgia, 550 D. T. Byrd Drive, Athens, GA 30602, USA. ⁷Department of Global Health, University of South Florida, 3720 Spectrum Boulevard, Tampa, FL 33632, USA. ⁸The Genomics Institute of the Novartis Research Foundation, 10075 John J. Hopkins Drive, San Diego, CA 92121, USA. ⁹Stichter Malaria Research Unit, Mahidol University, Mahidol University, Bangkok, Thailand. ¹⁰Centre for Tropical Diseases, Mahidol University, Bangkok, Thailand. ¹¹Department of Clinical Pathology, Mahidol University, Bangkok, Thailand. ¹²Department of Microbiology and Immunobiology, Harvard University, Cambridge, MA 02137, USA. ¹³Department of Biochemistry and Molecular Biology, University Park, PA 16802, USA. ¹⁴Department of Chemistry and Center for Infectious Disease Research, Pennsylvania State University, University Park, PA 16802, USA. ¹⁵Present address: BioSero, San Diego, CA 92122, USA. ¹⁶Present address: California Institute for Biomedical Research, La Jolla, CA, USA.
¹⁷*Corresponding author. Email: ewinzer@hsous.edu

Antonova-Koch et al., *Science* 362, eaat9446 (2018) 7 December 2018

Downloaded from <http://science.sciencemag.org/> on November 28, 2020

1 of 8

Chemoprotective antimalarials
Antonova-Koch et al., 2018

Atovaquone

Active!

AI in collaboration



Ersilia ‘trains’ a model based on partner’s data



Natural product activity
University of Buea

Humimycin A

Active!

Welcome to the Ersilia Model Hub!

Work in progress... 90 models added since the summer

<https://ersilia.io/model-hub>

 Type to search model...

Tags

Tox21

Toxicity

MoleculeNet

Grover

Graph Transformer

Output

Antibiotic activity

Toxicity

Synthetic accessibility

Antiviral activity

Target

Mode

Pretrained

Retrained

In-house

Online

License

Carcinogenic potential of metabolites and small molecules

eos1579 metabokiller

Carcinogenicity is a result of several potential effects on cells. This model predicts the carcinogenic potential of a small molecule based on their potential to induce cellular proliferation, genomic instability, oxidative stress, anti-apoptotic responses and epigenetic alterations.

Metabokiller uses the Chemical Checker signaturizer to featurize the molecules, and the Lime package to provide interpretable results.

Using Metabokiller, the authors screened a panel of human metabolites and experimentally demonstrated two of the predicted carcinogenic metabolites induced carcinogenic transformations in yeast and human cells.

Molecular maps based on broadly learned knowledge-based representations

eos6m4j bidd-molmap

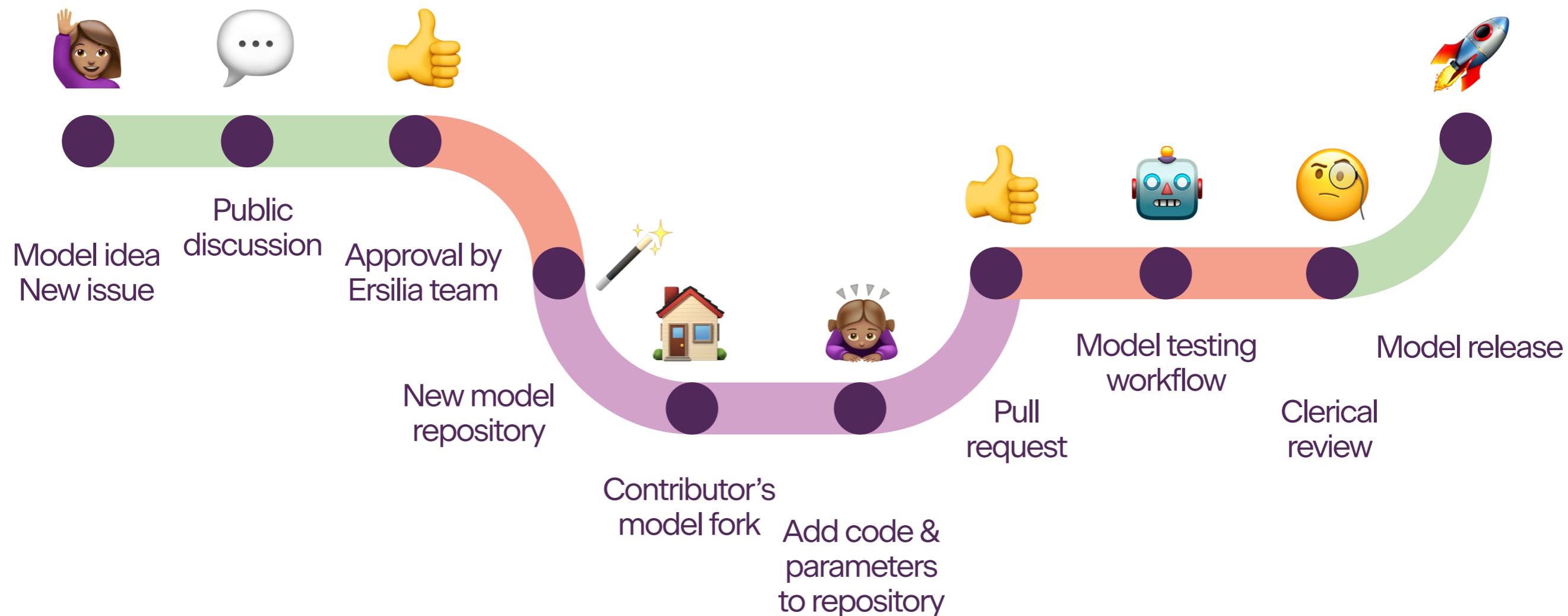
Descriptor-based or fingerprint-based molecular maps (images) are created. Typically, the goal is to use these images as inputs for an image-based deep learning model such as a convolutional neural network

SMILES transformer descriptor

eos2lm8 smiles-transformer

Molecular fingerprint based on natural language processing. It converts SMILES into fingerprints using an unsupervised model pre-trained on a very large SMILES dataset. The transformer is particularly well-suited for low-data drug discovery

An open-source community effort



With the support of GitHub, Atlassian, Outreachy and Harvard University

About the Ersilia Open Source Initiative

Towards a hub of machine learning models

Open Source, why?

Current projects

Open Source

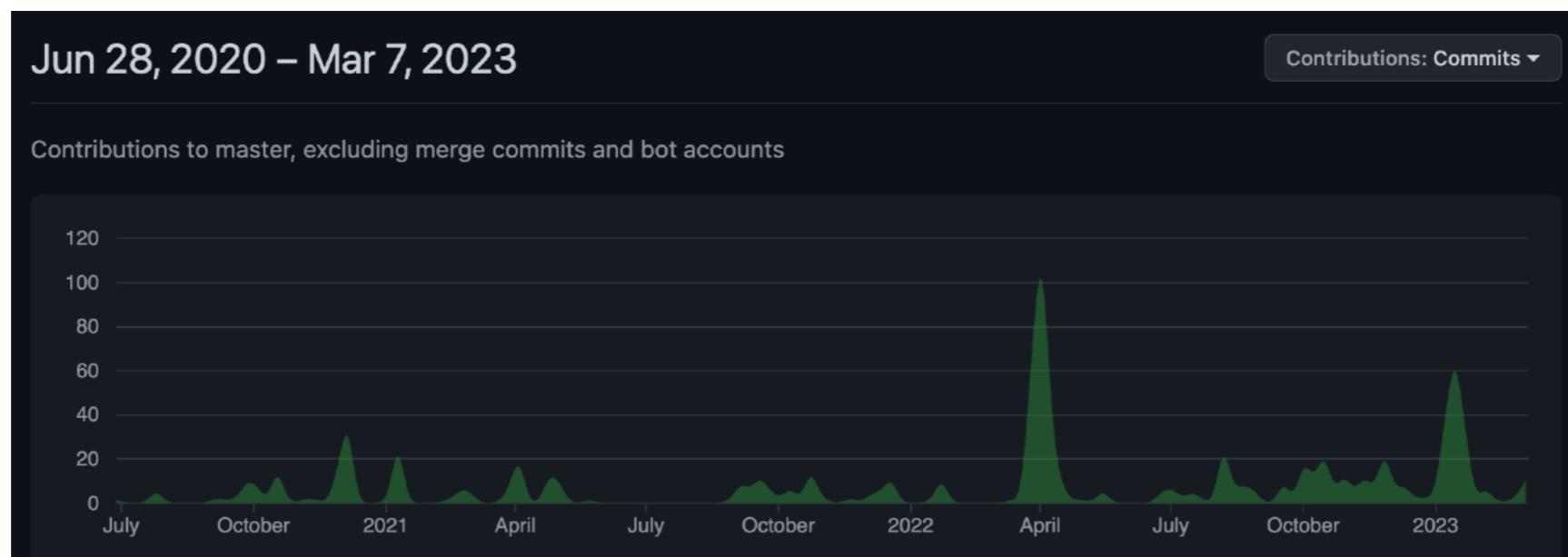


Why?

- Free to use
- Extensive testing and review
- Transparency
- Easy to onboard new contributors
- Enhances and facilitates collaborations
- Great community!

⭐ 101 Stars

✍ 1113 Commits



Ersilia's code is released under an OS License:

GNU General Public License v3 (GPL v3)

Free to:

- Run
- Study
- Share
- Modify

Purposes:

- Commercial
- Non-commercial

**Any derivative work must remain equally free to
use**

Our Open Source Community

Corporate volunteers



Academic collaborators



Students and interns



About the Ersilia Open Source Initiative

Towards a hub of machine learning models

Data-driven drug discovery

Current projects

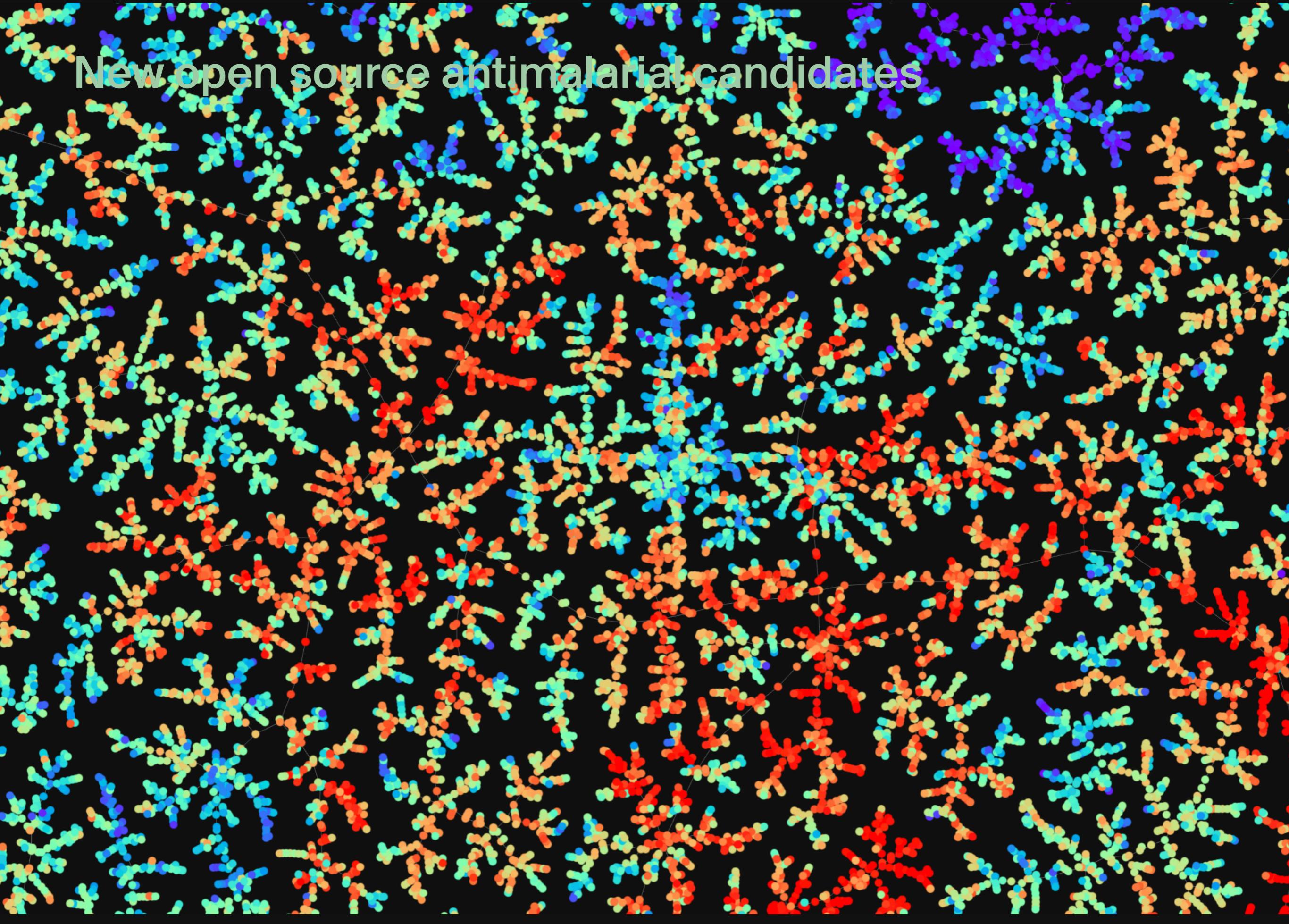
Antimalarial and antituberculosis drug discovery



Antimalarial and antituberculosis drug discovery

- Virtual screening cascade (15 AI/ML models) *
- End to end implementation at the H3D Centre (University of Cape Town)
- Using in-house proprietary data + open datasets
- Serving 100 scientists in South Africa

New open source antimalarial candidates



New open source antimalarial candidates

- AI-based generation of new compounds with activity against *P.falciparum*
- In collaboration with the Open Source Malaria consortium (international)
- 8 experimentally tested drugs, 4 with excellent in vitro activity
- Further development of selected candidates to optimize the TPP

African natural products as novel antivirals



African natural products as novel antivirals

- Virtual screening of ~15,000 African natural product compounds
- With University of Buea (Cameroon) & The Wistar Institute (US)
- Capacity building: ~6 months on-field
- 3 researchers from Cameroon to visit Ersilia (Barcelona Health Hub)
- 5-year project, Bill & Melinda Gates Foundation Calestous Juma

Pharmacogenomics of African ethnicities



Pharmacogenomics of African ethnicities

- In collaboration with the H3D Center, Cape Town (South Africa)
- Dose optimisation of malaria & tuberculosis treatments in Africa
- AI models to predict pharmacogenetic gene variants
- Coupling with in-house PBPK modeling tools
- Supported by GRADIENT: GSK, Novartis & SAMRC

Capacity building and training



Capacity building and training

- One-week workshop in Cape Town
- “Bringing Data Science and AI to Infectious Disease Research”
- Scientists from 7 African countries
- Partnership with the H3D Foundation
- Supported by Code For Science & Society (CS&S), and Wellcome Trust

A photograph showing a group of people in a workshop or lecture hall setting. They are seated at tables, looking at laptops and discussing. The room is dimly lit, and there are water bottles on the tables.

Training materials available at the online Ersilia Book

Remote internship programs for computer scientists



Remote internship programs for computer scientists

- Selected for the Outrechy internship programme
- Support to underrepresented minorities in computer science
- 12-week internship
- Multiple batches of Stanford University and Barcelona University interns
- New offices: Barcelona Health Hub (soon: 3 visitors from Cameroon!)

Take-home messages

- Ersilia is a **non profit organisation** aimed at developing **AI tools for drug discovery** and global health
- We combine **remote work** and **on-site implementation**
- We do **capacity building** activities
- Our products are **open-source**
- We offer support to **research institutes** in the Global South
- We are always looking for **collaborators!**

Funding

The Fore Fellowship
Event Fund (CS&S)
SeedCorn Award (Rosetrees Trust)
Biopharma Speed Grant (Merck KGaA)

In-kind support

A4ID
Airtable
Amazon Web Services
The Cranfield Trust
Atlassian Foundation
GitBook
Monday-dot-com
Outreachy
Red Española de Supercomputación
Software Sustainability Institute
Open Life Sciences
GitHub for Non-Profits
Digital Ocean
Think Evolve

Donors

Astor Foundation
Digital Lift
Black Rock Employee Award
Fast Forward
FundOSS
GSF Lab
Oakdale Trust
Okta Employee Giving
Splunk Pledge

Special thanks

Kelly Chibale
Matthew Todd
Edwin Tse
Jason Hlozek
John Woodland
Fidele Ntie-Kang
Albert Manasyan
Kati Taghavi
Julia Bohlius
Patrick Aloy
Adrià Fernández-Torras
Martino Bertoni
Aleix Gimeno
Arnau Comajuncosa
Francesco Iorio

Collaborators

University of Buea, Cameroon
H3D-UCT, South Africa
CIDRZ, Zambia
Scuola Normale Pisa, Italy
Open Source Malaria
University College London, UK
NYU Langone, US

Team & Alumni

Edoardo Gaude
Akash Rungta
Alacia Armstrong
Anna Homs
Gemma Turon
Miquel Duran-Frigola
Alice Chen
Raphael Brosular
Samuel Volk
Ifeoluwa Ojumoro
Amna Ali
Su Yen
Lazar Deretic
Tomasz Tokarczuk
Núria Camí
Marcos de la Torre
Ankur Kumar

Ersilia Open Source Initiative

Website: <https://ersilia.io>

E-mail: hello@ersilia.io

Twitter: [@ersiliaio](https://twitter.com/ersiliaio)

Gemma Turon, PhD

E-mail: gemma@ersilia.io

Twitter: [@TuronGemma](https://twitter.com/TuronGemma)



Ersilia