Drug repurposing for SARS-CoV-2 infection using machine learning and mechanistic models of the COVID-19 disease map

Carlos Loucera[1,2], Marina Esteban-Medina[1,2], Kinza Rian[1], Matias Falco[1,2,3], Joaquin Dopazo[1,2,3,4], and María Peña-Chilet[1,2,3]

[1]*Clinical Bioinformatics Area, Fundación Preogreso y Salud* (FPS)
[2]*Computational Systems Medicine, Institute of Biomedicine of Seville* (IBIS)
[3]*Bioinformatics in Rare Diseases* (BiER), *Centro de Investigación Biomédica en Red de Enfermedades Raras* (CIBERER)
[4]FPS/ELIXIR-es, *Hospital Virgen del Rocío, Seville, Spain*

5th 2020 Disease Maps Community Meeting, 12-November

Introduction

In this work, we use an innovative methodology that combines mechanistic modeling of the signal transduction circuits related to SARS-CoV-2 infection (the COVID-19 disease map) with a machine learning algorithm that learns potential causal interactions between proteins, already targets of drugs, and specific signaling circuits in the COVID-19 disease map, to suggest potentially *repurposable* drugs.

(a0) Build the Signalization Space

- Over **10k samples** across 53 non-diseased tissue sites
- A **COVID-19 Disease Map** is transformed into ...
- **277 circuits** across **47** signalization **pathways**.

(a1) Hipathia Signal Transduction

$$S_n = v_n \left(1 - \prod_{s_a \in A} (1 - s_a)\right) \prod_{s_i \in I} (1 - s_i)$$

(a2) Connect the Dots

- The **Map** results in 277 activities values for 10k samples.
- Over **2k KDT** targeted by **>2k drugs**
- Most **KDT** lie outside the **Disease Map**.

(a3) MultiTask Machine Learning

Can **Map Activity** be predicted from (Drugbank) **KDT**? Can we infer what KDT are important to regulate the map. *Learn* KDT-Signalization *relationships*.

(a4) Model Explanations

- Useful but too broad.
- They (only) speak us about the whole map.

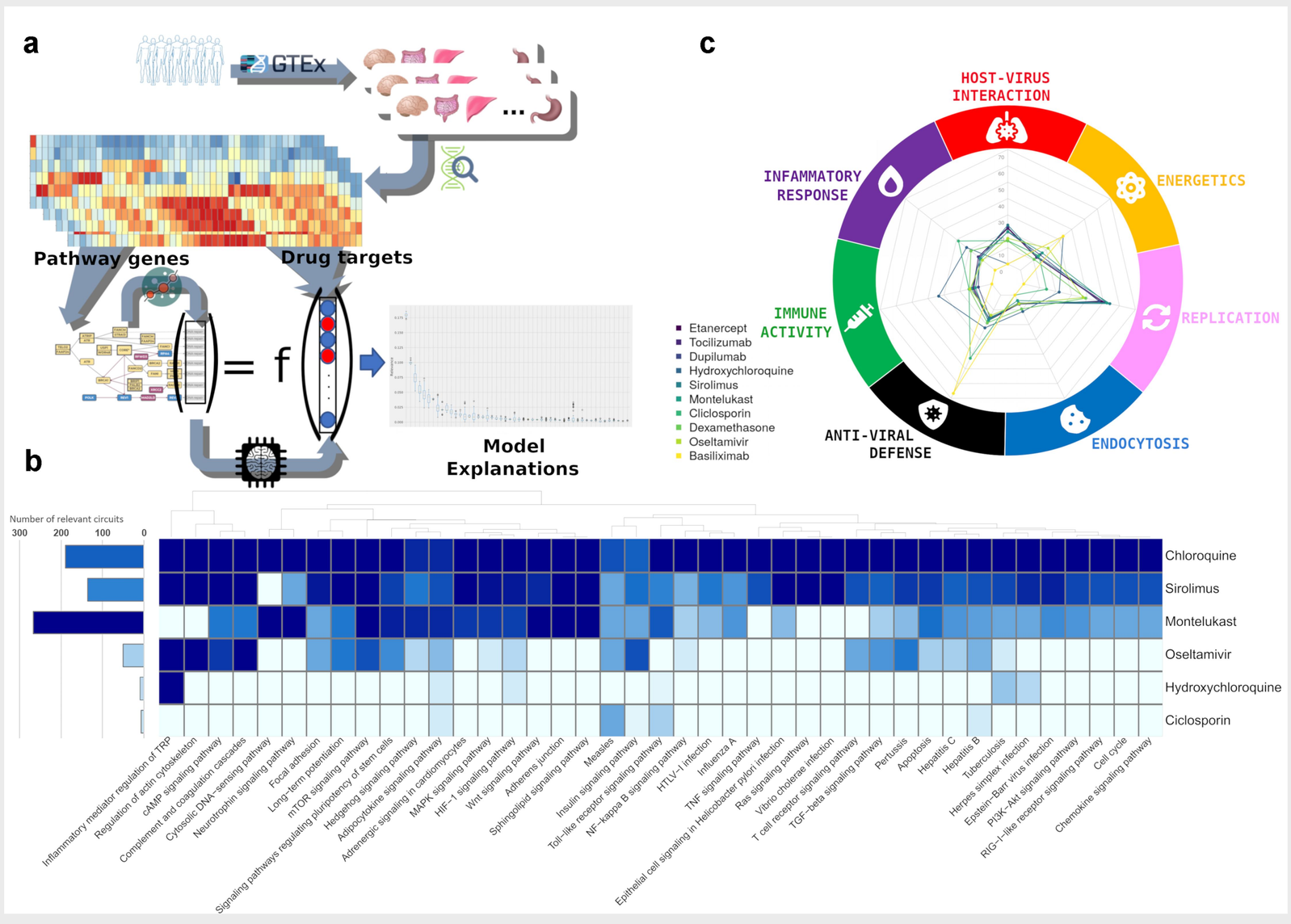
(a5) SHapley Additive exPlanations

- Fair feature and sample wise contributions.
- Disaggregated by circuit by construction.
- Additive (aggregated by biologically-relevant groups)

(a6) Data driven Validation

- Optimization + Performance with Repeated Nested CV
- Explainability tested with novel methods (Nogueira)

A Modular Methodology



(b, c) Findings

- 380 KDTs (targeted by 679 drugs) have direct influence over the whole or partial parts of the map.
- The GO biological processes enriched are mostly related to immune activity (T-cell, inflammatory response)
- The COVID-19 Hallmarks are represented.
- The disaggregated attributions conform a series of templates, e.g.:
  - Chloroquine affects massively almost all the Disease Map
  - Ciclosporin affect a few very specific circuits
- Several of the predicted drugs are under clinical trial

Future Works

- Speed the computation (GPU extensions).
- Develop *sign* like SHAP aggregations.
- Aggregations in *task*-space.
- Use directly the COVID-19 Disease Map (near done! see Kinza's talk, and our ML framework is DB-agnostic (**modular!**))
- Build the model with the updated Maps
- Incorporate more databases
- End the COVID-19 pandemic!

Funding

Fundación BBVA

GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL

Co-funded by the Horizon 2020 programme of the European Union

- This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813533. -

Speaker Affiliations

JUNTA DE ANDALUCIA
Fundación Progreso y Salud
CONSEJERÍA DE SALUD Y FAMILIAS

IBIS INSTITUTO DE BIOMEDICINA DE SEVILLA