

Drug repurposing for COVID-19 using explainable machine learning and mechanistic models of signal transduction circuits related to SARS-CoV-2 infection with real world data validation

Carlos Loucera

Bioinformatics Area, Andalusian Public Foundation Progress and Health-FPS
Institute of Biomedicine of Seville, IBiS, University Hospital Virgen del Rocío/CSIC/University of Seville
41013 Sevilla, Spain

FDA/NCTR-MAQC 2022 Conference
2022 September 26



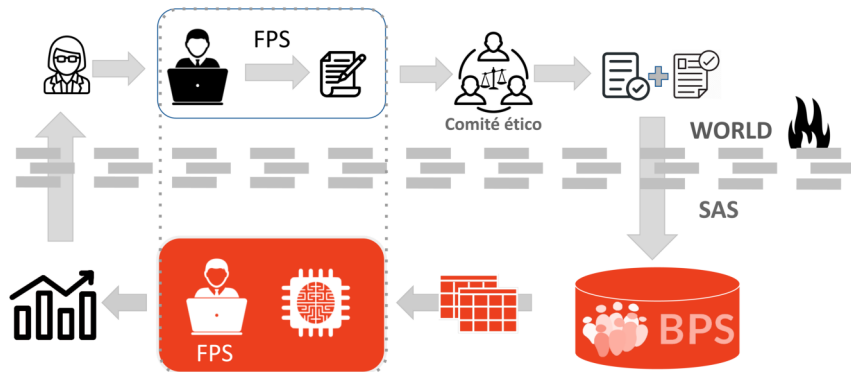
Junta
de Andalucía

Consejería de Salud
y Consumo

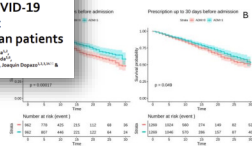
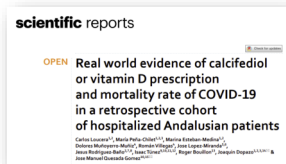
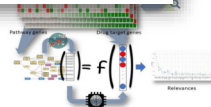
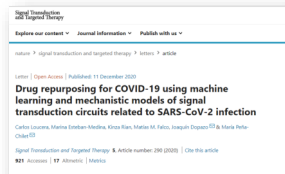
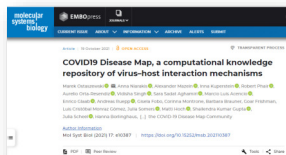
Fundación
Progreso y Salud



Infrastructure for secure generation of Real World Evidence from Real World Data from the Andalusian Health Population Database



Knowledge paradigm



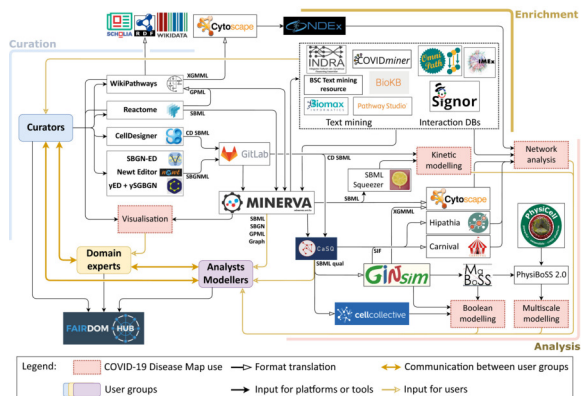
Knowledge discovery

Living Contradiction is a fascinating, honest examination of that genuine contradiction faced by teachers in reconciling the effort made to encourage young people towards independent critical thinking, with the simultaneous sense of responsibility to instruct and insist on a particular behavior.

Trustworthiness

The **WHY** is as important as the **WHAT**

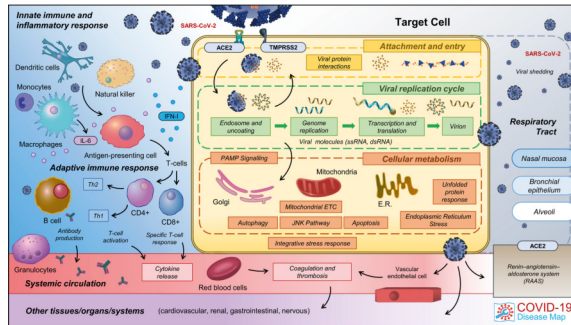
Building a COVID-19 Disease Map



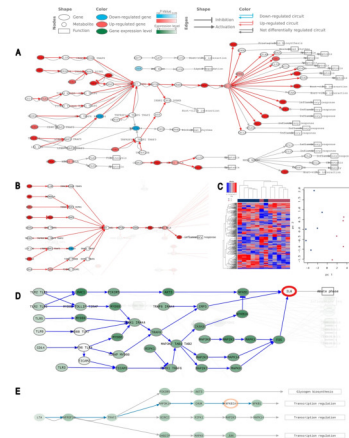
COVID-19 Disease Map¹

¹Marek Ostaszewski et al. (2021). "COVID19 Disease Map, a computational knowledge repository of virus–host interaction mechanisms." In: *Molecular systems biology* 17.10, e10387

Building a COVID-19 Disease Map



COVID-19 Disease Map¹

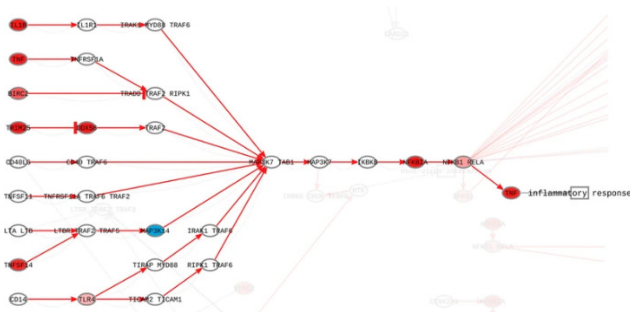


(Simplified) COVID-19 Disease Map²

¹Marek Ostaszewski et al. (2021). "COVID19 Disease Map, a computational knowledge repository of virus–host interaction mechanisms." In: *Molecular systems biology* 17.10, e10387

²Kinza Rian et al. (2021). "Mechanistic modeling of the SARS-CoV-2 disease map." In: *BioData Mining* 14.1, pp. 1–8

Mechanistic modeling



High throughput estimation of functional cell activities reveals disease mechanisms and p relevant clinical outcomes

[PDF](#) | [HTML](#) | [Supplementary Files](#) | [How to cite](#) | [Order a Reprint](#)
Oncotarget. 2017; 8:5100-5178. <https://doi.org/10.18632/oncotarget.14107>

Metrics: PDF 1955 views | HTML 4565 views

Marta R. Hidalgo¹, Cankut Cubuk¹, Alicia Amadoz^{1,2}, Francisco Salavert^{1,3}, José Carbonell-Caballero¹, Joaquín Dopazo^{1,2,3}

¹Computational Genomics Department, Centro de Investigación Principe Felipe (CIPF), Valencia, 46012, Spain

²Functional Genomics Node (INB-ELIXIR-es), Valencia, 46012, Spain

³Bioinformatics in Rare Diseases (BiER), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Valencia, 46012, Spain

Correspondence to:

Joaquín Dopazo, email: jdopazo@cipf.es

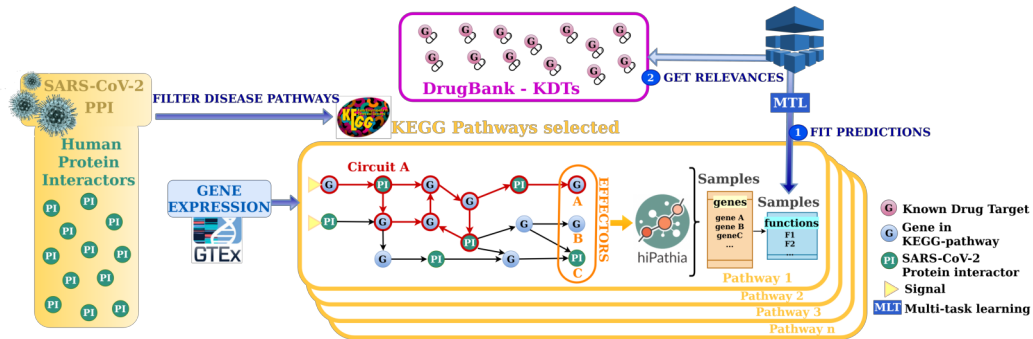
Keywords: **signaling pathway, disease mechanism, prognostic, survival, biomarker**

Received: September 01, 2016 Accepted: November 21, 2016 Published: December 22, 2016

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



Drug repurposing schema



ML pipeline

Our **FS** model makes decisions based on the **SHAP** ranks

We want to be **true to the data**

Avoid overconfidence

Draw from a background distribution

Compute the explanations using a different subset



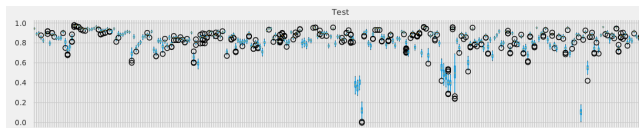
(Nogueira) Build a population (repeat the whole procedure **100 times**)

Make it faster: **GPU-CUDA**

Data-driven Validation (Performance + Explanations)

Explanation stability

Robustness Vs Stochastic, Noise, Sampling
100 Holdout splits of half the sample size
Split *training* into *learning* and *validation*
Check unbiased performance over *test*
Compute Nogueiras¹ statistic test and CI



R^2 score distribution over the test splits

Mean R^2 : 0.82 ± 0.02
N-Stat CI: (0.729, 0.735)

¹Sarah Nogueira et al. (2017). "On the stability of feature selection algorithms." In: [The Journal of Machine Learning Research](#) 18.1, pp. 6345–6398

SHapley Additive exPlanations

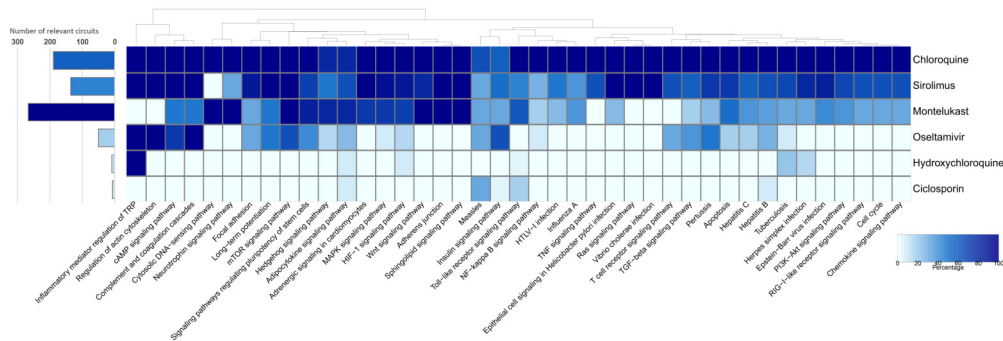
Fair feature responsibility attribution.
Dis-aggregated by circuit by construction
Additive (use biologically-relevant groups)

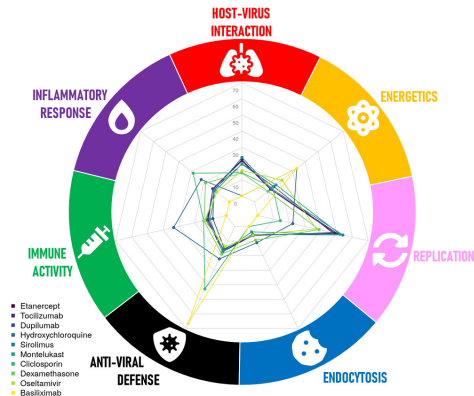
Template patterns

Conform a series of templates, e.g.:

Affect massively almost all MAP

Affect a few very specific circuits





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

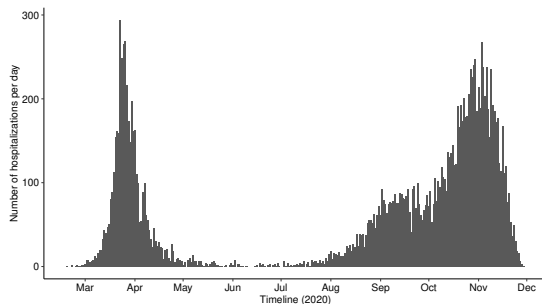
Data facts

Andalusian public health system

Year 2020, no COVID-vaccine

15968 COVID-19 hospitalized patients

End point: COVID-19 certified death (28 days)



covariate	survival	death	p-value
Total N	13116	2678	
Sex (female)	6024 (45.9)	1129 (42.2)	<0.001
Flu vaccine	5465 (41.7)	1746 (65.2)	<0.001
Pneumococcal vaccine	3441 (26.2)	1111 (41.5)	<0.001
Diabetes	3856 (29.4)	1167 (43.6)	<0.001
Circulatory diseases	8111 (61.8)	2261 (84.4)	<0.001
Cancer	1550 (11.8)	545 (20.4)	<0.001
Respiratory diseases	2896 (22.1)	828 (30.9)	<0.001
Dementia	964 (7.3)	536 (20.0)	<0.001
Other mental diseases	1764 (13.4)	407 (15.2)	0.018
Anxiety and mood disorders	3382 (25.8)	784 (29.3)	<0.001
Age			<0.001
18_41	1399 (10.7)	20 (0.7)	
41_68	5971 (45.5)	380 (14.2)	
68_99	5746 (43.8)	2278 (85.1)	

Methods Facts

We Only include properly balanced treatments¹

964 treatments found

122 were eligible

HR Closed-form variance estimator for Weighted Propensity Scores²

Lymphocyte count registered up to 14 days since hospitalization begins

Covariate-adjusted linear mixed effect model to test Lymphocyte progression trends

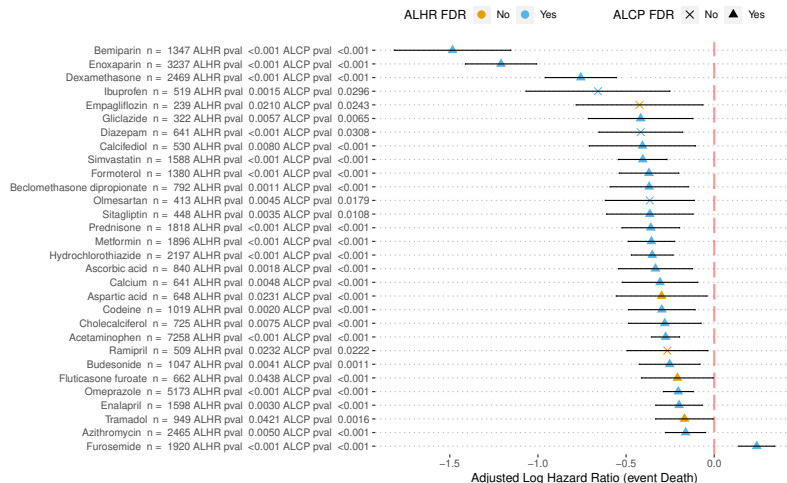
30 treatments are significant for both tests

22 after FDR adjustment

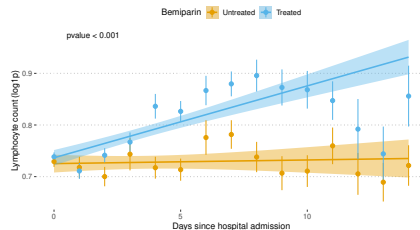
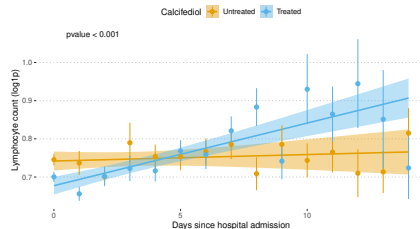
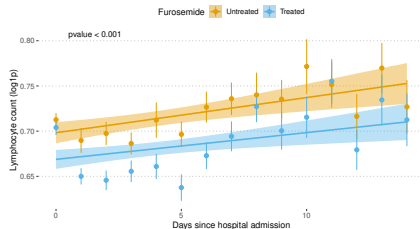
¹Elizabeth A Stuart et al. (2013). "Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research." In: Journal of clinical epidemiology 66.8, S84–S90

²David Hajage et al. (2018). "Closed-form variance estimator for weighted propensity score estimators with survival outcome." In: Biometrical Journal 60.6, pp. 1151–1163

Covariate-Adjusted LHR by Treatment



Covariate-Adjusted Lymphocyte trend



Conclusions

Bemiparin and **Enoxaparin** → Highest survival

- prevent thrombotic and thromboembolic complications

- enoxaparin has been previously reported as protective

- the protective effect is not shared by other anticoagulants

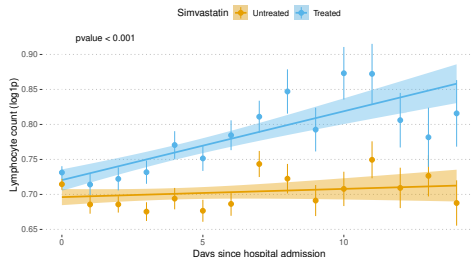
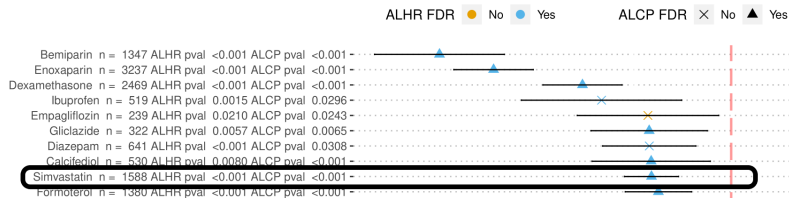
Calcifediol and **Cholecalciferol** have a protective effect

- probably due to vitamin D and its pro-immune and anti-inflammatory properties²

There is a **significant intersection** between **ML predicted drugs** and **RWE**

$$\chi^2 = 6.674, \text{ pvalue} = 0.009785$$

Simvastatin summary



Pleiotropic effect
antiinflammatory and antithrombotic
inhibiting the NF- κ B pathway
directly reduces inflammatory cytokines
(IL1, IL6, TNF- α), CRP, and neutrophils
Under heavy study

Closing remarks

It is not a competition between paradigms and methods

We aim to gain evidences from as many sources as possible

Our **xML** tools aim to help users **prioritize** their work:

- if they *look for a needle in a haystack*

- help them by

 - (**filter**) removing stacks of hay

 - prioritize** what remains

 - with **biology-driven constrains** and **explanations**

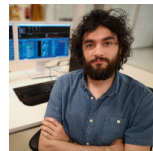
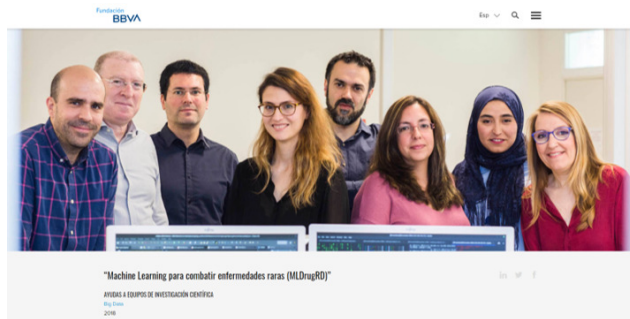
Explainability and interpretability should be **central** to ML solutions

If we include explanations we need to add a **new evaluation axis**:

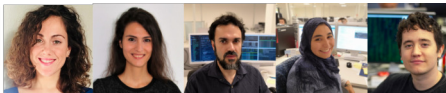
- predictive performance is not enough

BPS, the database for secondary use of clinical data of the Andalusian Public Health System, constitutes a unique resource for large-scale RWE studies

The team



Funded by the ISCIII



Personal funding. Contact: carlos.loucera@juntadeandalucia.es



Carlos Loucera postdoctoral contract PAIDI2020-D0C_00350 is funded by Junta de Andalucía and co-funded by the European Social Fund (FSE) 2014-2020





Fundación Progreso y Salud
CONSEJERÍA DE SALUD Y FAMILIAS

Clinical Bioinformatics Area Fundación Progreso y Salud, Sevilla, Spain, and...

...the INB-ELIXIR-ES, National Institute of Bioinformatics
and the BiER (CIBERER Network of Centers for Research in Rare Diseases)



<https://www.slideshare.net/xdopazo/>



Follow us on
twitter



@xdopazo

@ClinicalBioinfo

@loucerac

@CSoriguer



Fundación BBVA



Thank You!