Accelerating drug repositioning with interpretable machine learning

JABI
JORNADAS ANDALUZAS DE BIOINFORMÁTICA
SEVILLA, 3-5 DE JULIO 2023
FIS20 (PI20/01305)

Marina Esteban-Medina[1,3], Carlos Loucera[1,3], Maria Peña-Chilet[1,2,3], Joaquín Dopazo[1,2,3]

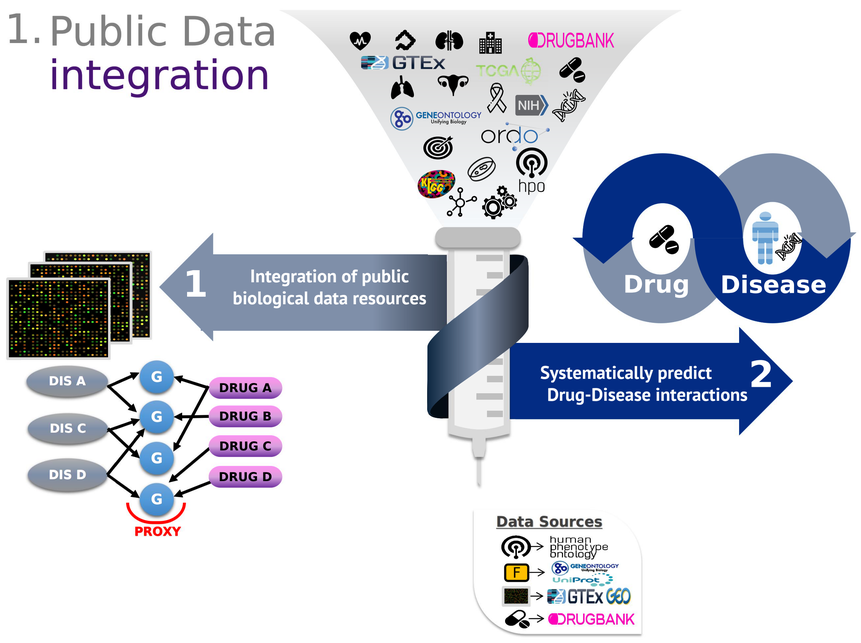
[1]Área de Bioinformática Clínica, Fundación Progreso y Salud. Sevilla, España.
[2]BieR - centro de Investigaciones Biomédicas en Red en Enfermedades Raras (CIBERER). Sevilla, España.
[3]Computational Systems Medicine, Institute of Biomedicine of Seville (IBIS), Hospital Virgen del Rocío, 41013, Sevilla, Spain.

ABSTRACT

The development of new therapies can be a slow and expensive process, particularly when it comes to rare or arising diseases, ieither because thery are consider a less tractive market or because of the lack of knowledge on the disease. **Drug repositioning** emerges as a quick and effective way to obtain treatments that have been already approved for clinical use. Merging the **knowledge** about mechanisms of **disease** and **drug action** together with the arising machine learning methodologies, here we present **DRExM3L**, a rational data-driven approach to drug repositioning that utilizes explainable machine learning and **mechanistic models of signal transduction** to predict potentially causal relationships between proteins of interest, in this case targets of known drugs, **(KDTS)** and the disease mechanisms identified in the "**Disease Map**". These data-driven approaches are a great promise in accelerating the generation of knowledge, particularly in fields where little is known, such as rare or emerging diseases.

MATERIALS & METHODS

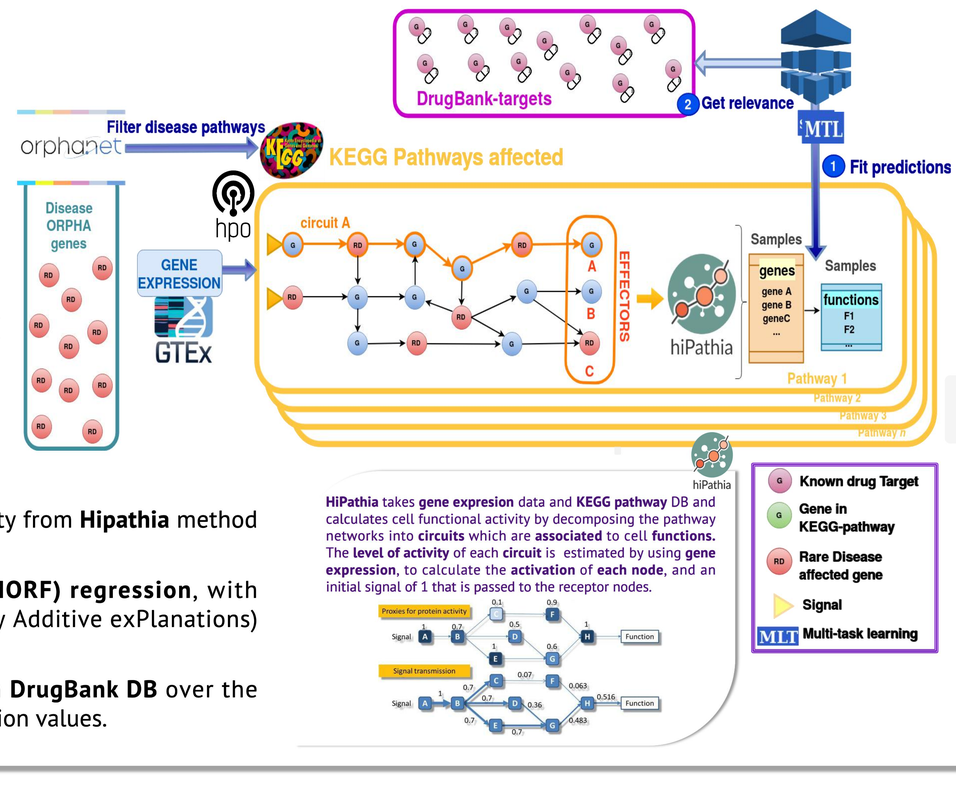
1. Public Data integration



2. Mechanistic & ML modelling

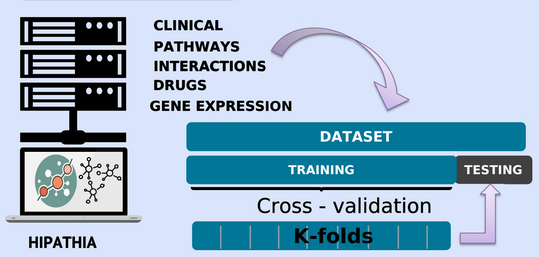
STEPS:

- **KEGG** database (DB) for the information of each disease map.
- Construction of the disease map by extracting all the circuits* containing the genes responsible for each rare neoplastic disease according to **ORPHANET DB**.
- Information about the approved drug targets from **DrugBank.**
- Using **GTEx** expression data, we obtained circuits activity from **Hipathia** method and associated **Gene Ontology** and **Uniprot** functions.
- Implementation of **Multi-Output Random Forest (MORF) regression**, with repeated **cross-validation** and unbiased **SHAP** (SHapley Additive exPlanations) values.
- Infering the effect of the approved drug targets from **DrugBank DB** over the activity of the argets over the Disease Map circuit activation values.



RESULTS

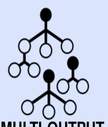
ML Trainning



1. **Optimization of hyperparameters**: Search for the best model.
2. **Obtain model performance**: Check the model works fine
3. **Compute relevance distributions** of all trained models.

1. Training the model from expression ...

.. to predict pathways activities

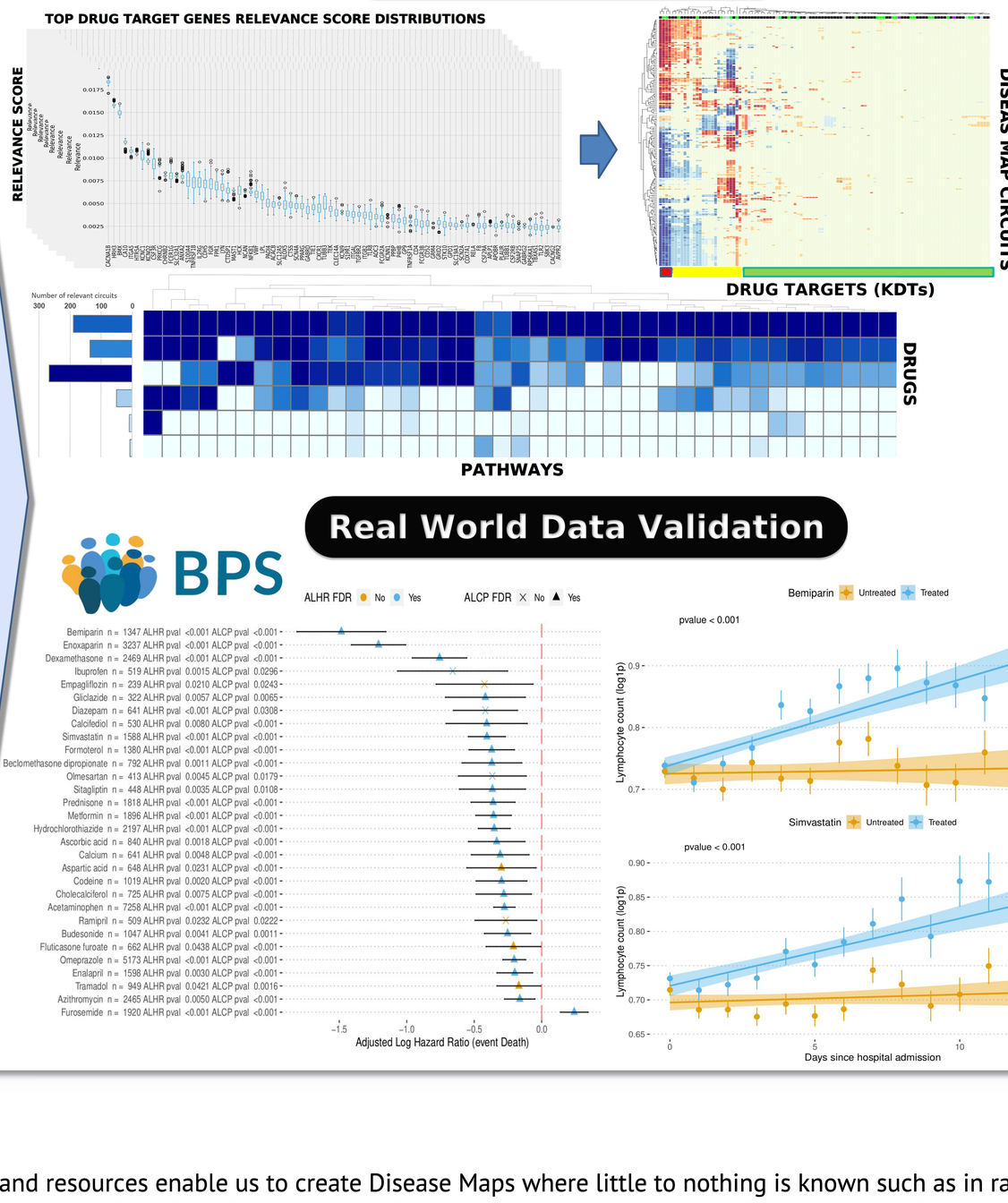


MULTI-OUTPUT RANDOM FOREST

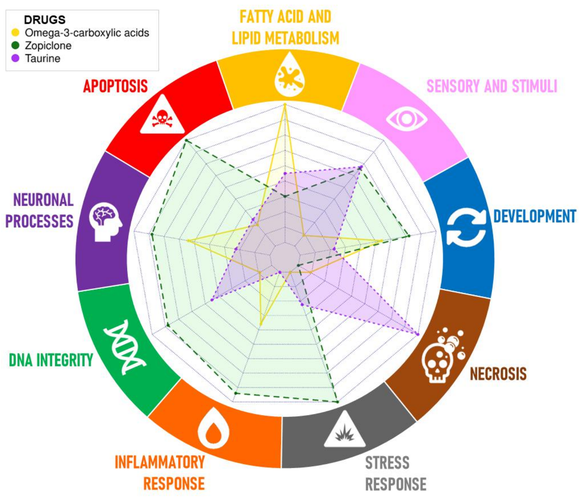
2. Predicting the interactions Drug target - Disease Map

3. Making the model explainable: drug targets score extraction with direction

KDT scores Analysis

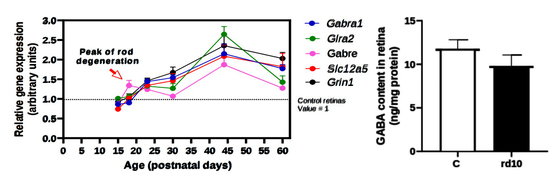


Functional Analysis



Real World Data Validation

Experimental Validation



Data driven

Drug repositioning

HIGHLIGHTS

The **integration** of information from different **databases** and resources enable us to create Disease Maps where little to nothing is known such as in rare or arising diseases.
The automatization of the expansion process used in the construction of the **Disease Maps** of action, enables the extension of its use to any disease with associated genes affected.
Relevance **evaluation** of **drug targets** from approved drugs **over** the **Disease Map** revealed top gene-targets for the prediction of the disease actionable pathways as well as functional modules.
The integration of **RWD for the *in-silico* validation** of promising drugs accurately predicted the protective effect, accelerating the selection and experimental validation of **promising candidates.**

CONTACT US !

marina.estebanm@gmail.com

FEDER Fondo Europeo de Desarrollo Regional
UNIÓN EUROPEA "Una manera de hacer Europa"
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
Instituto de Salud Carlos III
Junta de Andalucía
Consejería de Salud y Consumo
Fundación Progreso y Salud
ciberer
IBIS INSTITUTODEBIOMEDICINADESEVILLA
iMPaCT DATA
Fundación BBVA
AGENCIA ESTATAL DE INVESTIGACIÓN