

# Machine Learning Models for Discovering Biological Interactions

**Ivan Marcelo Carrera Izurieta**

*Advisor: Dra. Ines Dutra*

Dep. Ciencia de Computadores. UP.

*Co-advisor: Dr. Eduardo Tejera*

Grupo de Químio-Bioinformática. UDLA.

Doutoramento em Ciência de Computadores.  
Faculdade de Ciências. Universidade do Porto.

October 27, 2020



FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO

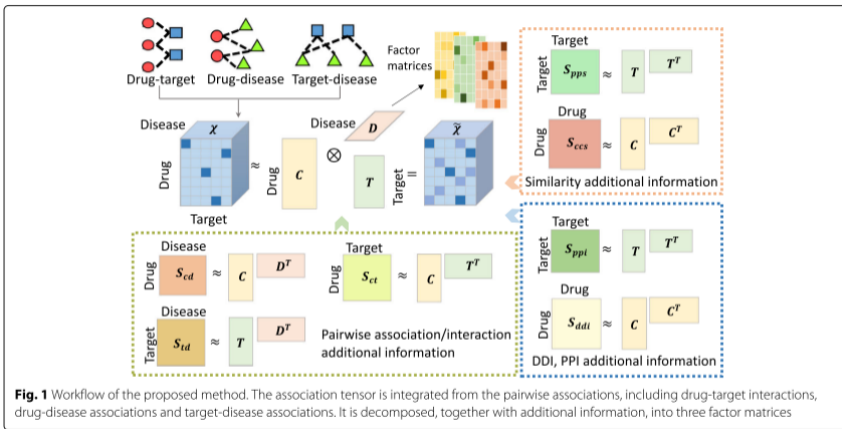


## Introduction

- ▶ The main goal of this work is to develop a set of computational models for the *in silico* discovery of new drug activities.
- ▶ The contribution of this work focuses on novel ways for describing cellular lines and diseases, as well as the discovery of novel drug applications.
- ▶ In the context of Bioinformatics, the motivation of this work is to

## Introduction: Drug Discovery process

Drug discovery process: discover new applications for known drugs.



**Fig. 1** Workflow of the proposed method. The association tensor is integrated from the pairwise associations, including drug-target interactions, drug-disease associations and target-disease associations. It is decomposed, together with additional information, into three factor matrices

[Zhang et al., 2018]

## Introduction: Development stages

This work is divided into 4 stages, namely:

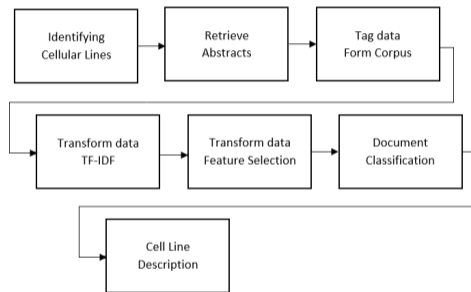
- ▶ (1) Cell Fishing,
- ▶ (2) Cell characterization,
- ▶ (3) Disease gene prioritization, and
- ▶ (4) Drug discovery.

## Development stages: Cell Fishing

- ▶ Cell Fishing is discovering drug-cell line interactions.
- ▶ Database on interactions between cell lines and chemical compounds
- ▶ Characterize cell lines with the set of chemical compounds they interact with.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.
- ▶ Published: [Tejera et al., 2019]

## Development stages: Cell Characterization

- ▶ In order to discover drug-cell line interactions, we need a way for better describing cell lines.
- ▶ Characterize cell lines using text mining.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.
- ▶ Published: [Carrera et al., 2020]
- ▶ Submitted to BIBM 2020.



## Development stages: Cell Characterization

- ▶ In order to discover drug-cell line interactions, we need a way for better describing cell lines.
- ▶ Characterize cell lines using text mining.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.
- ▶ Published: [Carrera et al., 2020]
- ▶ Submitted to BIBM 2020.

	title	index	document	cellId
0	Survey of ATCC stocks of human cell lines for HeLa cont...	PMD:566722	Survey of ATCC stocks of human cell lines for HeLa cont...	CVCL_2260
1	Comparison of gene expression profiles of conjunctival c...	PMD:19630270	Comparison of gene expression profiles of conjunctival c...	CVCL_2260
2	Check your cultures! A list of cross contaminated or misi...	PMD:20143388	Check your cultures! A list of cross contaminated or misi...	CVCL_2260
3	Identification by C banding of two human prostate tumo...	PMD:2744886	Identification by C banding of two human prostate tumo...	CVCL_4806
4	Feasibility of drug screening with panels of human tumo...	PMD:3325022	Feasibility of drug screening with panels of human tumo...	CVCL_4806
5	Human tumor lines for cancer research	PMD:3518977	Human tumor lines for cancer research DOI: 10 3109 073...	CVCL_4806
6	Human urologic cancer cell lines	PMD:6244232	Human urologic cancer cell lines 1 Invest Urol 1980 Mar...	CVCL_4806
7	Molecular characterization of human prostate carcinoma...	PMD:14518029	Molecular characterization of human prostate carcinoma...	CVCL_4806
8	Synthesis of chondrocytic keratan sulphate containing pr...	PMD:2045389	Synthesis of chondrocytic keratan sulphate containing pr...	CVCL_M605
9	Marshall Urist Award Interstitial collagenase gene expres...	PMD:10906887	Marshall Urist Award Interstitial collagenase gene expres...	CVCL_M605
10	Phenotypic characterization of chondrosarcoma derived ...	PMD:16030953	Phenotypic characterization of chondrosarcoma derived ...	CVCL_M605
11	A comprehensive transcriptional portrait of human canc...	PMD:25485619	A comprehensive transcriptional portrait of human canc...	CVCL_M605
12	Structural electrophysiological biochemical and pharm...	PMD:16829	Structural electrophysiological biochemical and pharm...	CVCL_0464
13	Synapse formation between two clonal cell lines	PMD:193191	Synapse formation between two clonal cell lines Clonal n...	CVCL_0464
14	Culture and characteristics of hormone responsive neuro...	PMD:2985920	Culture and characteristics of hormone responsive neuro...	CVCL_0464

## Development stages: Cell Characterization

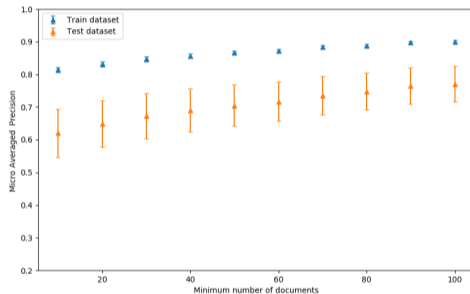
- ▶ In order to discover drug-cell line interactions, we need a way for better describing cell lines.
- ▶ Characterize cell lines using text mining.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.
- ▶ Published: [Carrera et al., 2020]
- ▶ Submitted to BIBM 2020.

	0	1	2	3	4	5	6	7	
0	0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00284	0.00000	-0
1	0.00001	-0.00000	0.00000	-0.00245	0.00002	-0.00000	-0.00001	0.00000	-0
2	0.00000	0.17482	-0.00004	0.00000	-0.00000	-0.04145	0.19039	-0.00000	-0
3	-0.00000	-0.00183	-0.03235	0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.0
4	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0
5	-0.00002	0.00000	-0.00000	-0.00000	-0.00000	-0.00001	-0.00001	0.00000	-0
6	-0.00005	0.00002	-0.00001	0.00014	-0.00000	-0.00004	-0.00018	-0.00004	-0
7	0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0
8	-0.00001	0.00000	0.00000	-0.00000	-0.00000	-0.00129	-0.00001	-0.00000	-0
9	-0.00000	-0.00000	-0.00704	0.00000	-0.01160	0.00000	0.00004	-0.00000	0.0
10	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00001	-0.00426	-0.00000	-0
11	-0.06807	0.00002	0.00002	-0.00001	-0.00001	-0.00006	-0.00012	-0.00000	-0
12	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0
13	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	-0
14	-0.00676	-0.01006	0.00003	-0.00001	-0.00000	-0.00000	-0.01150	0.00000	-0



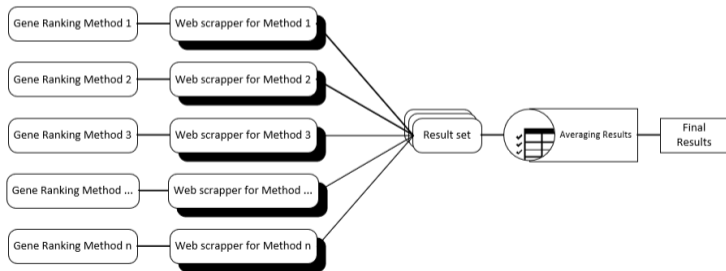
## Development stages: Cell Characterization

- ▶ In order to discover drug-cell line interactions, we need a way for better describing cell lines.
- ▶ Characterize cell lines using text mining.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.
- ▶ Published: [Carrera et al., 2020]
- ▶ Submitted to BIBM 2020.



## Development stages: Disease Characterization

- ▶ In order to discover drug-disease interactions, we need a way for describing diseases: gene prioritization.
- ▶ There are a lot of gene prioritization algorithms, the best way is a consensus [Tejera et al., 2017].
- ▶ Characterize disease using a prioritized set of genes.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.



## Development stages: Disease Characterization

- ▶ In order to discover drug-disease interactions, we need a way for describing diseases: gene prioritization.
- ▶ There are a lot of gene prioritization algorithms, the best way is a consensus [Tejera et al., 2017].
- ▶ Characterize disease using a prioritized set of genes.
- ▶ Given a new chemical compound, Cell Fishing obtains a set of candidate cell lines that could be hit by the chemical compound.

Diabetes Mellitus, Type 1

Phenotype:  Phenotype as found in MedGen:

CUID:  MIM Number:

**Diabetes Mellitus, Type 1**

Show  entries

Gene Symbol	Value
PTPN22	0.84017639576723
INS	0.83098011582457
CTLA4	0.83580563461748
HLA-DQA1	0.82979667039628
HLA-DQB1	0.78367354210524
HNF1A	0.73943810253664
HLA-DRB1	0.72996229407285
TNF	0.70302432196825
GAD2	0.68930730289409
MICA	0.68490775961572

Showing 1 to 25 of 351 entries

Previous

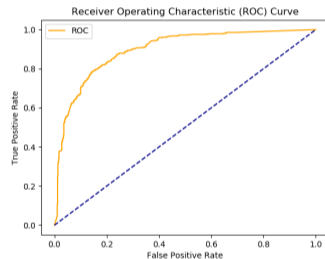
These results were processed in:

CIPHER  Genedistiller  Genie  Snp

Disgenet  Guildify  Glad4u  Bee

## Development stages: Drug Discovery





- ▶ Hypothesis: discover drug-disease interactions, modeling the problem as a Recommender System.
- ▶ Proof of concept: model cell lines as a set of drugs, model drugs as a set of cell lines.
- ▶ Compute new drug-cell line interactions as a Recommender System.
- ▶ Remaining work: introduce the cell line characterizations, change target to disease



## Publications and Submissions

- ▶ [Tejera et al., 2019] PLoS ONE (October 2019)
- ▶ [Carrera et al., 2020] ACM BCB 2020 (September 2020)
- ▶ BIBM 2020 (result in November 2020)

## References I

-  Carrera, I., Dutra, I., and Tejera, E. (2020).  
Representing Cellular Lines with SVM and Text Processing.
-  Tejera, E., Carrera, I., Jimenes-Vargas, K., Armijos-Jaramillo, V., Sánchez-Rodríguez, A., Cruz-Montegudo, M., and Perez-Castillo, Y. (2019).  
Cell fishing: A similarity based approach and machine learning strategy for multiple cell lines-compound sensitivity prediction.  
*PLoS ONE*, 14(10):e0223276.
-  Tejera, E., Cruz-Montegudo, M., Burgos, G., Sánchez, M.-E., Sánchez-Rodríguez, A., Pérez-Castillo, Y., Borges, F., Cordeiro, M. N. D. S., Paz-y Miño, C., and Rebelo, I. (2017).  
Consensus strategy in genes prioritization and combined bioinformatics analysis for preeclampsia pathogenesis.  
*BMC medical genomics*, 10(1):50.
-  Zhang, L., Chen, X., Guan, N.-N., Liu, H., and Li, J.-Q. (2018).  
A Hybrid Interpolation Weighted Collaborative Filtering Method for Anti-cancer Drug Response Prediction.  
*Frontiers in Pharmacology*, 9(SEP).