PHARMACOLOGIC SUBSTANCES & CORONAVIRUS

Data analysis on relationships between **pharmacologic substances or targets** & **disease**

TABLE OF CONTENTS

DESCRIPTION OF THE ASSIGNMENT

- We have two datasets that contain valuable information about the relationship of several chemicals & drugs with coronavirus.

  1. In the 1st dataset these relations are directly reported.
  2. In the 2nd dataset the pharmacologic substances show their effectiveness on coronavirus through genes, proteins or receptors.

- Also, we have available data from ATC, a system that classifies active ingredients of drugs according to the organ or system on which they act.

DESCRIPTION OF THE ASSIGNMENT

The tasks of the assignment are divided in 2 categories regarding to which dataset will be used:

1. Based on the 1st dataset:
   - We need to distinguish drugs from chemicals based on ATC's list of drugs.
   - We have to report the Top 10 Key Opinion Leaders for every drug identified in our data.

2. Based on the 2nd dataset:
   - We need to identify the drugs as before and for each drug to report the corresponding collection of proteins and coronavirus concepts.
   - We have to show which substances have similar targets.

DRUG IDENTIFICATION

In order to identify which substances of our list are included also in ATC ontology, we created a mechanism that checks for all of our substances :

- If the ATC list has a drug with the same name or part of it (e.g. zinc ~ zinc products).

- Or if there is a drug with similar name (e.g. cyclosporine ~ ciclosporin). This was implemented by calculating pairwise string similarity scores.

IDENTIFIED DRUGS IN 1ST CASE

Substances in Case 1	Substances in ATC
cyclosporine	ciclosporin
dexamethasone	dexamethasone and antibiotics
tretinoin	tretinoin
omacetaxine	omacetaxine mepesuccinate
oritavancin	oritavancin
mercaptopurine	mercaptopurine
thioguanine	tioguanine
⋮	⋮

* THE FULL LIST OF IDENTIFIED DRUGS ARE REPORTED IN "IDENTIFIED_DRUGS_CASE_1.XLSX" FILE.

IDENTIFIED DRUGS IN 1ST CASE

" The **word cloud** consists of all the substances that were identified in ATC and the size corresponds to the <u>frequency</u> of this substance in our dataset "
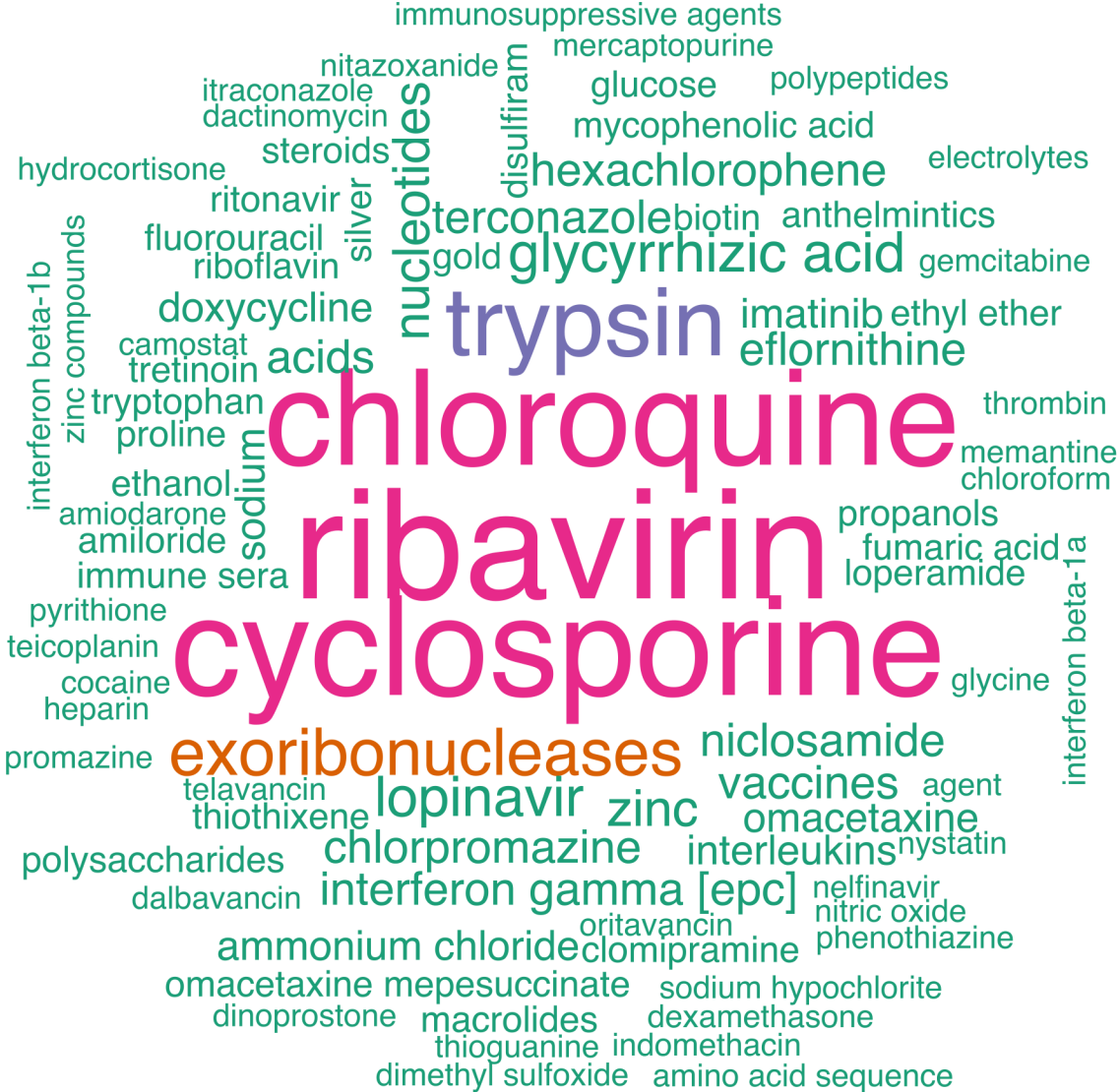
TOP 10 KEY OPINION LEADERS PER DRUG

cyclosporine	chloroquine	zinc
Sasaki Takashi	Benjannet Suzanne	Acharya Badri Narayan
Sato Yuka	Bergeron Eric	Bhagyawant Sameer S.
Tanaka Yoshikazu	Erickson Bobbie R	Dash Paban Kumar
Inoue Mai	Ksiazek Thomas G	Hu Xueying
Osawa Shuichi	Nichol Stuart T	Jiang Tengfei
Tanaka Satoka	Rollin Pierre E	Jiang Yunbo
Snijder Eric J	Seidah Nabil G	Kesari Pooja
Thiel Volker	Vincent Martin J	Li Jin-ping
Zevenhoven-Dobbe Jessika C	Afridi Shabbir Khan	Lin Ping
de Wilde Adriaan H	Aliyari Roghiyh	Liu Bang

* THE FULL LIST IS REPORTED IN "TOP10_KOL.XLSX" FILE.

CASE 2- IDENTIFIED DRUGS

Substances in Case 2	Substances in ATC
dexamethasone	dexamethasone and antibiotics
angiotensin ii	angiotensin ii
fructose	fructose
perindopril	perindopril and amlodipine
rosiglitazone	rosiglitazone
irbesartan	irbesartan and diuretics
valsartan	valsartan and amlodipine
⋮	⋮

* THE FULL LIST OF IDENTIFIED DRUGS ARE REPORTED IN "IDENTIFIED_DRUGS_CASE_2.XLSX" FILE.

PROTEINS & CORONAVIRUS CONCEPTS PER DRUG

Identified Drug	Collection of proteins	Coronavirus Concepts
dexamethasone	ACE2 gene | Elastases | Toll-like receptors | MAPK14 protein, human	SARS coronavirus
angiotensin ii	ACE2 gene	SARS coronavirus
fructose	ACE2 gene	SARS coronavirus
perindopril	ACE2 gene	SARS coronavirus
rosiglitazone	ACE2 gene | Toll-like receptors	SARS coronavirus
irbesartan	ACE2 gene | Ubiquitin	SARS coronavirus | Genus: Coronavirus
deoxycholic acid	BAG3 gene	SARS coronavirus
azacitidine	Peptidylprolyl Isomerase	Genus: Coronavirus
⋮	⋮	⋮

* THE FULL LIST IS REPORTED IN "PROTEINS & CONSEPTS_PER_SUBSTANCE.XLSX" FILE.

CLUSTERING BASED ON PROTEIN TARGET
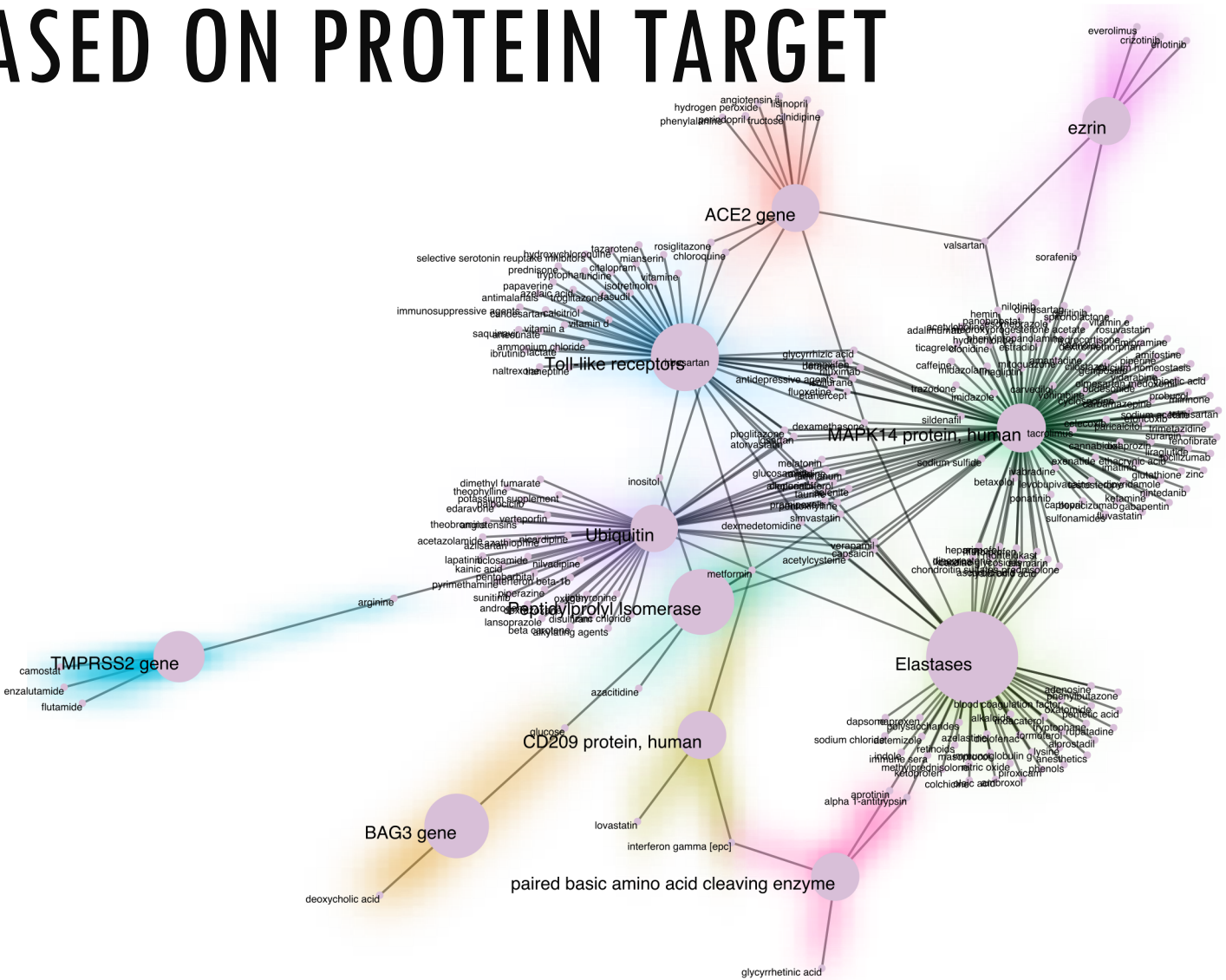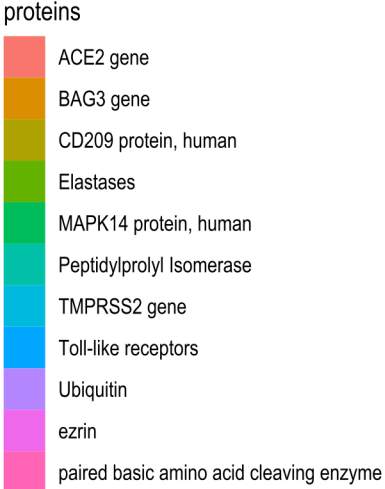
ADDITIONAL INFORMATION

- The identification algorithm in some cases found a drug with several names in ATC. For example, **"vaccines"** was found in **"Meningococcal vaccines", "Rubella vaccines"** and more.

- The **Key Opinion Leaders** (KOL) were found by extracting all the scientists/authors related to the specific drug. Then, by calculating how many times someone has been referred to this particular drug, we were able to report the Top 10 KOL.

- Some drugs have less than 10 KOL, because less than 10 scientists/authors have been referred to them. For instance **"riboflavin"** has only **Bowen Richard, Marschner Susanne** & **Keil Shawn D**.

- The whole assignment is implemented in **R** (see causaly_assignment_final.R file).

- All the Excel reports are extracted in folder "output" and the data input exists in folder "data".