Bio-Join Phase 3 Presentation

BiS332 course by Professor Doheon Lee

Surayouth Phuksawattanachai (20180813)
Alexander Semenov (20226159)
Lana Abu Hassan (20170842)

Motivation: Drug - Disease relation

Question 1

What is the most suitable drug that I should use to treat this disease?

Question 2

Which diseases can I use this drug for?

Question 3

If I use this drug, what genes will it affect?

New data

Source: [Comparative Toxicogenomics Database]

Data: extended OMIM with Disease vocabulary (MEDIC)

disease_name	id	alt_id	gene_symb
10p Deletion Syndrome (Partial)	MESH:C538288		None
13q deletion syndrome	MESH:C535484		None
16p11.2 Deletion Syndrome	MESH:C579850		None
17,20-Lyase Deficiency, Isolated	MESH:C567076		None
17-Hydroxysteroid Dehydrogenase Deficiency	MESH:C537805	OMIM:264300	EDH17B3
17-Hydroxysteroid Dehydrogenase Deficiency	MESH:C537805	OMIM:264300	HSD17B3
18-Hydroxylase deficiency	MESH:C537806	OMIM:203400|OMIM:610600	CYP11B2
22q11 Deletion Syndrome	MESH:D058165		None
2,4-Dienoyl-CoA Reductase Deficiency	MESH:C565624	OMIM:616034	C5orf33
2,4-Dienoyl-CoA Reductase Deficiency	MESH:C565624	OMIM:616034	DECRD

(10 rows)

New data

Source: Comparative Toxicogenomics Database

Data: Chemical vocabulary and Chemical–disease associations

drug_name	drug_id	disease_name	disease_id
06-Paris-LA-66 protocol	C046983	Precursor Cell Lymphoblastic Leukemia-Lymphoma	MESH:D054198
10074-G5	C534883	Adenocarcinoma	MESH:D000230
10074-G5	C534883	Adenocarcinoma of Lung	MESH:D000077192
10074-G5	C534883	Alopecia	MESH:D000505
10074-G5	C534883	Androgen-Insensitivity Syndrome	MESH:D013734
10074-G5	C534883	Astrocytoma	MESH:D001254
10074-G5	C534883	Autistic Disorder	MESH:D001321
10074-G5	C534883	Breast Neoplasms	MESH:D001943
10074-G5	C534883	Breast Neoplasms, Male	MESH:D018567
10074-G5	C534883	Bulbospinal neuronopathy, X-linked recessive	MESH:C537017

(10 rows)

Drug-Diseases

Given a disease, find drug to treat it

```
Please provide a disease name: Liver Diseases
Use these drugs:
ToxVal | Drug Name
-----------------
99.98  | 7-methyl-5-(1-((3-(trifluoromethyl)ph
99.98  | argemone oil
99.98  | calcium propionate
99.98  | Thiamine
99.93  | 4-amino-1,8-naphthalimide
-----------------
```

Given drug, find diseases that can be treated with it

```
Please provide a drug name: Ibuprofen
These disease can be treated:
Prevalence | Diseases
----------------
999360.0   | Graft Occlusion, Vascular
998891.0   | Carcinoma, Ductal, Breast
998741.0   | Anterior Cruciate Ligament Injuries
996922.0   | Squamous Cell Carcinoma of Head and Neck
996453.0   | Hyperalgesia
-----------------
```

Drug-Genes

Given a drug, find genes affected

```
Please provide a drug name: Ibuprofen
These genes are affected:
Popularity | Gene Symbols
----------------
100.0      | PRKCH
99.89      | ADH1B
99.75      | MST1R
99.75      | MDD1
99.74      | ICOS
99.51      | NOP56
99.45      | UVSSA
99.37      | HGF
99.36      | NUP214
99.19      | ISG15
-----------------
```

Given drug, find diseases that can be treated with it

```
Please provide a drug name: Ibuprofen
These chromosomes are most commonly affected:
Chromosome 1: 74 affected genes
Chromosome 2: 52 affected genes
Chromosome 6: 52 affected genes
```

Disease- Gene

Given chromosome number , find associated diseases

```
Please provide a chromosome number: 5
These disease are assosiated with chromosome '5':
-------------------------------------------------
Heparin Cofactor II Deficiency
2-Hydroxyglutaricaciduria
Graft Occlusion, Vascular
Lymphomatoid Granulomatosis
Cardiomyopathy, Familial Hypertrophic, 13
-----------------
```

Ranking Output

Since the output might be too large, ranking of drugs, genes and diseases is needed

Our criteria was: toxicity/bioavailability, infection/mortality rate, popularity/importance

- Found no public data
- We made mock data for:

  Chemicals: Toxicity
  Genes: Popularity
  Diseases: Prevalence

Future perspective: replace the current data with real data e.g. https://go.drugbank.com/

Ranking Output

disease_id	prevalence
MESH:C564334	286725
MESH:C563409	154111
MESH:D034161	934590
MESH:D012868	479340
MESH:C565185	554613
OMIM:613158	407127
MESH:C537252	875811
MESH:D015210	258325
MESH:C567261	220644
MESH:D007640	773602

drug_id	tox
C065545	8.16
C552122	67.69
C097880	95.09
C581203	57.77
C452139	44.15
C528580	43.83
C013082	30.78
C034410	86.9
C052610	44.92
C519197	13.68

gene_symb	popularity
A1BG	88.29
A2M	26.72
A2MP1	81.22
NAT1	67.65
NAT2	73.86
NATP	52.38
SERPINA3	60.01
AADAC	63.94
AAMP	27.98
AANAT	55.27

Bonus: Statistics

Question 4

What is the disease distribution on chromosomes?

Question 5

What are the most universal drugs?

The interface:

```
Choose a statistic?
```

```
1. Statistics: Count number of diseases for each chromosome
2. Statistics: Drugs that can treat the most diseases
```

Bonus: Statistics

Disease distribution on chromosomes

```
Chromosome-Diseases stats:
Chromosome 1: 524
Chromosome 2: 385
Chromosome 3: 336
Chromosome 4: 220
Chromosome 5: 249
Chromosome 6: 294
Chromosome 7: 240
Chromosome 8: 178
Chromosome 9: 188
Chromosome 10: 228
Chromosome 11: 357
Chromosome 12: 313
Chromosome 13: 111
Chromosome 14: 156
Chromosome 15: 157
Chromosome 16: 204
Chromosome 17: 322
Chromosome 18: 90
Chromosome 19: 250
Chromosome 20: 148
Chromosome 21: 47
Chromosome 22: 115
Chromosome X: 325
```

Most universal drugs

```
The most universal drugs:
Num of Diseases it can treat | Drug name
-------------------------------------
5789 | bisphenol A
5542 | Benzo(a)pyrene
5461 | Tetrachlorodibenzodioxin
5364 | Valproic Acid
5015 | Acetaminophen
4533 | Aflatoxin B1
4452 | Tobacco Smoke Pollution
4399 | sodium arsenite
4268 | Ethinyl Estradiol
4254 | Estradiol
```

Thank you for you attention

Any questions?

Check out the source code on:
https://github.com/Tsatsch/Biojoin