

# Interações entre Droga e Doença por meio de Genes

Mateus Siqueira Batista   Nicolas Bissoli Nattis

MC536 - Instituto de Computação, UNICAMP

2020

# Proposta

Obter dados de interações entre genes e drogas, e entre genes e doenças.

Através disso, podemos relacionar a interação entre estas drogas e as doenças.

## Exemplo:

- Droga A ativa o gene X.
- Gene X tem relação de causa com as doenças  $\alpha$ ,  $\gamma$ .
- Portanto, a droga A tem relação de causa com as doenças  $\alpha$ ,  $\gamma$ .



- Dados sobre interações droga-gene e o genoma drogável.
- Extraído de mais de trinta fontes confiáveis.
- Dados extraídos via arquivo .tsv disponibilizados.



- Plataforma contendo uma das maiores coleções publicamente disponíveis de genes e variantes associados a doenças humanas.
- Dados extraídos via banco de dados SQLite disponibilizado.

A relação de droga-doença ocorre de modo intuitivo a partir das relações de doença e droga com o gene.

|             |          | Droga-Gene |          |
|-------------|----------|------------|----------|
|             |          | ativação   | inibição |
| Gene-Doença | ativação | ativação   | inibição |
|             | inibição | inibição   | ativação |

# Tipos de Interação

## Activator

Activator

Inducer

Agonist

Binder

Chaperone

Partial Agonist

Potentiator

Stimulator

Substrate

## Inhibitor

Antibody

Antagonist

Inhibitor

Antisense  
Oligonucleotide

Blocker

Inhibitory  
Allosteric  
Modulator

Inverse Agonist

Partial Antagonist

Suppressor

DgiDb

Susceptibility  
Mutation

Chromosomal  
Rearrangement

Somatic Causal  
Mutation

Fusion Gene

Germline Causal  
Mutation

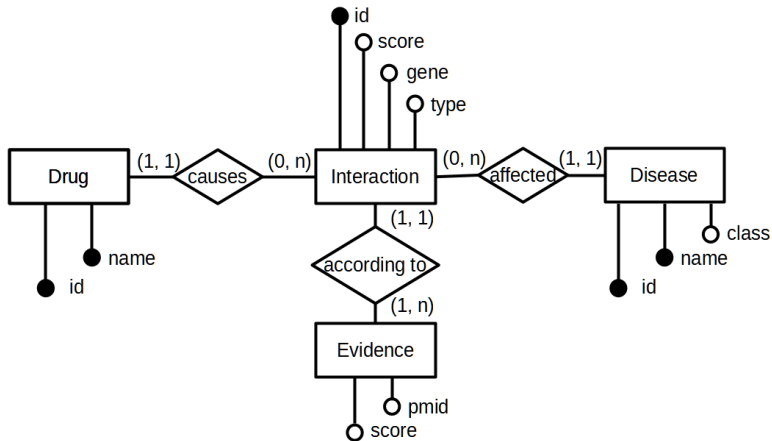
GeneticVariation

Biomarker

Therapeutic

DisgeNet

# Modelo Conceitual



# Modelo Lógico

Drug(DrugId, Name)

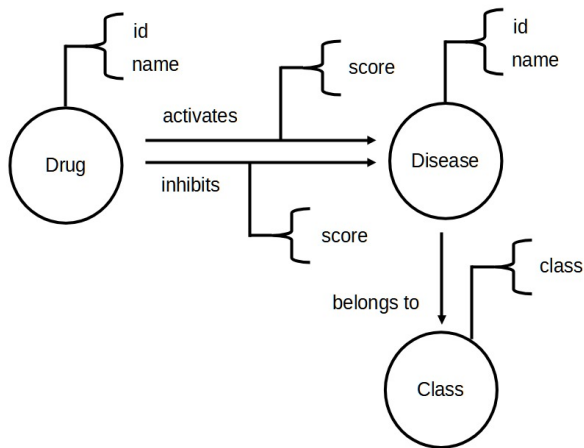
Disease(DiseaseId, Name, Class)



# Modelo Lógico

```
Interaction(InteractionId, DrugId, DiseaseId, Score,  
Gene, Type)  
Evidence(InteractionId, Pmid, Score)
```

# Modelo Lógico de Redes Complexas



# Evolução do projeto

- Removemos classes de drogas e aliases (fora do escopo).
- Dividimos as interações do Neo4j em duas.
- Usamos SQLite e TSV ao invés de requisições HTTP (muito lento).

# Transformação

- Transformar DISGENET SQLite em TSV.
- Ler TSV das fontes.
- Produzir TSV final.
- Transformar TSV final em SQLite.
- Transformar TSV final em CSV para Neo4j.

# Tratamento de Dados

## ■ Confibialidade

- Os dados coletados possuem níveis de confiabilidade variáveis, que foram levados em conta e tranformados em um Score.
- Há um Score atrelado ao **Dgldb** e outro ao **DisGeNET**. Assim, obtivemos um Score global da interação multiplicando um pelo outro.

# Destaques

```
# Importação das iterações

print('Generating interactions')
with open('../data/external/disgenet/interactions.tsv', 'w') as f:
    f.write('DiseaseId\tGene\tScore\tPMID\tType\n')
    for (number, gene) in enumerate(genes):
        print(f'\r{(number + 1) * 100 // len(genes)} % ({number + 1} / {len(genes)}), end='')
        for disease, type, pmid, score in conn.execute(
            'SELECT d.diseaseNID, d.associationType, d.pmid, d.score'
            + ' FROM geneDiseaseNetwork AS d, geneAttributes as g'
            + ' WHERE g.geneName = ? AND g.geneNID = d.geneNID', [gene]):
            f.write(f'{diseases_id[disease]}\t{gene}\t{score}\t{pmid}\t{type}\n')
```

# Destaques

```
# Parse de dados do DgIDB

def parse() -> DgIdb:
    r = DgIdb()

    dgidb = pd.read_csv(
        '../data/external/dgidb/interactions.tsv',
        sep='\t',
        encoding='utf-8')
    dgidb = dgidb.dropna()
    scaler = MinMaxScaler()
    dgidb[['interaction_group_score']] = scaler.fit_transform(
        dgidb[['interaction_group_score']])

    for _, row in dgidb.iterrows():
        drug = row['drug_concept_id']
        gene = row['gene_name']
        drug_name = row['drug_name']
        pmids = str(row['PMIDs']).split(',')
        score = row['interaction_group_score']

        types = str(row['interaction_types']).split(',')
        ty = None
        for x in types:
            try:
                y = dgidb_interaction_types[x]
                if ty == None or ty == y:
                    ty = y
            else:
                ty = None
                break
        except KeyError:
            pass
        if ty == None:
            continue
        r.drug_names[drug] = drug_name
        r.interactions[gene].append(
            DrugGeneInteraction(drug, ty, score, pmids))

    return r
```

# Destaques

```
# Geração de nós

print('Generating nodes.csv')
with open('../data/processed/neo4j/nodes.csv', 'w', newline='') as fout:
    out = csv.writer(fout)
    out.writerow(['id:ID', 'name', ':LABEL'])

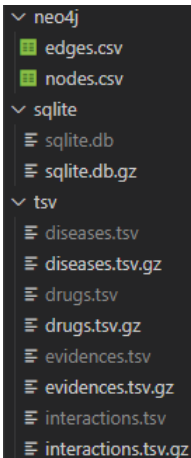
    with gzip.open('../data/processed/tsv/drugs.tsv.gz', 'rt') as f:
        dict = csv.DictReader(f, delimiter='\t')
        out.writerows([[row['Id'], row['Name'], 'Drug'] for row in dict])

    with gzip.open('../data/processed/tsv/diseases.tsv.gz', 'rt') as f:
        dict = csv.DictReader(f, delimiter='\t')
        for row in dict:
            out.writerow([row['Id'], row['Name'], 'Disease'])
            for class_ in row['Class'].split(';'):
                disease_classes[class_].append(row['Id'])

    out.writerows([[x, x, 'Class'] for x in disease_classes])
```



# Dados publicados



# Perguntas de Análise

Selecione o nome das drogas e doenças das N interações com maior pontuação.

```
SELECT Dr.Name, Di.Name, I.Type, I.Score  
FROM Interaction as I, Drug as Dr, Disease as Di  
WHERE Dr.Id = I.DrugId AND Di.Id = I.DiseaseId  
ORDER BY I.Score DESC LIMIT 10;
```

# Perguntas de Análise

Selecione o nome das drogas e doenças das N interações com maior pontuação.

| Dr .Name                 | Di .Name  | I .Type | I .Score          |
|--------------------------|---|---------|-------------------|
| NITISINONE               | Tyrosinemia, Type III   | 0       | 0.861673561599624 |
| BUROSUMAB                | Autosomal dominant hypophosphatemic rickets                               | 0       | 0.833348003403456 |
| NITISINONE               | Hawkinsinuria   | 0       | 0.669172021242262 |
| THIAMINE                 | THIAMINE METABOLISM DYSFUNCTION SYNDROME 5 (EPISODIC ENCEPHALOPATHY TYPE) | 1       | 0.591677082416454 |
| BUROSUMAB                | TUMORAL CALCINOSIS, HYPERPHOSPHATEMIC, FAMILIAL, 1                        | 0       | 0.500008802042074 |
| BUROSUMAB                | TUMORAL CALCINOSIS, HYPERPHOSPHATEMIC, FAMILIAL, 2                        | 0       | 0.500008802042074 |
| BUROSUMAB                | Hypophosphatemic Rickets  | 0       | 0.416674001701728 |
| BUROSUMAB                | TUMORAL CALCINOSIS, HYPERPHOSPHATEMIC, FAMILIAL                           | 0       | 0.416674001701728 |
| NITISINONE               | Tyrosinemias  | 0       | 0.40333656074876  |
| GLEMBATUMUMAB<br>VEDOTIN | AMYLOIDOSIS, PRIMARY LOCALIZED CUTANEOUS, 3                               | 1       | 0.4               |

# Perguntas de Análise

Quais drogas tem relação com a doença (Acute lymphocytic leukemia) C0023449?

```
SELECT Dr.Id, Dr.Name, I.Type, I.Score  
FROM INTERACTION as I, DRUG as Dr, Disease as Di  
WHERE Di.id ='C0023449' AND I.DiseaseId = Di.id  
ORDER BY I.Score DESC LIMIT 10;
```

# Perguntas de Análise

Quais drogas tem relação com a doença (Acute lymphocytic leukemia) C0023449?

| Dr.Id                | Dr.Name           | I.Type | I.Score            |
|----------------------|-------------------|--------|--------------------|
| chembl:CHEMBL398707  | HYDROMORPHONE     | 0      | 0.0388903558959012 |
| chembl:CHEMBL3545253 | FLORTAUCIPIR F 18 | 0      | 0.0388903558959012 |
| chembl:CHEMBL2       | PRAZOSIN          | 0      | 0.0388903558959012 |
| chembl:CHEMBL1621597 | IPRATROPIUM       | 0      | 0.0388903558959012 |
| chembl:CHEMBL157101  | KETOCONAZOLE      | 0      | 0.0388903558959012 |
| chembl:CHEMBL1670    | MITOTANE          | 0      | 0.0388903558959012 |
| chembl:CHEMBL723     | CARVEDIOL         | 0      | 0.0388903558959012 |
| chembl:CHEMBL157138  | LISURIDE          | 0      | 0.0388903558959012 |
| chembl:CHEMBL2103830 | FOSTAMATINIB      | 0      | 0.0388903558959012 |
| chembl:CHEMBL1201250 | BENZQUINAMIDE     | 0      | 0.0388903558959012 |

# Perguntas de Análise

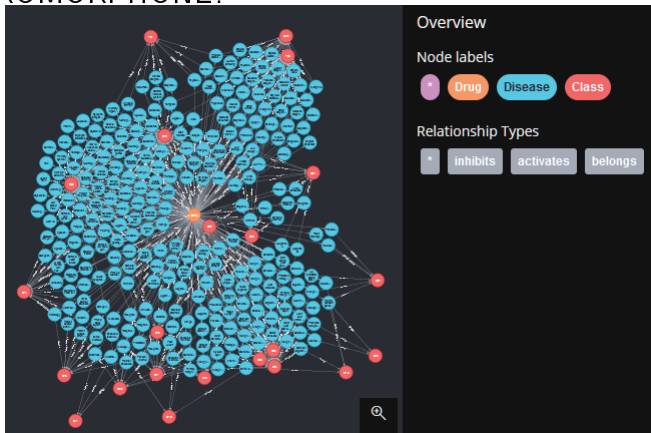
Quais classes de doenças são mais ativadas pela HYDROMORPHONE?

- Uma análise aprofundada é necessária, mas podemos partir da seguinte.

```
MATCH (dr:Drug)-[activates]->(di:Disease)
-[belongs]->(c:Class)
WHERE dr.name='HYDROMORPHONE'
RETURN dr, activates, di, belongs, c
```

# Perguntas de Análise

Quais classes de doenças são mais ativadas pela HYDROMORPHONE?



# Perguntas de Análise

- Podemos deduzir novas interações a partir do grafo?
  - Predição de links.
- Existem tipos de drogas e doenças que interagem mais entre si?
  - Comunidade.



Obrigado pela atenção!