

Software Engineering Capstone

Safety Knowledge Graph

GradVek 2.0

Wednesday, May 10 2023

01-problem

02-solution

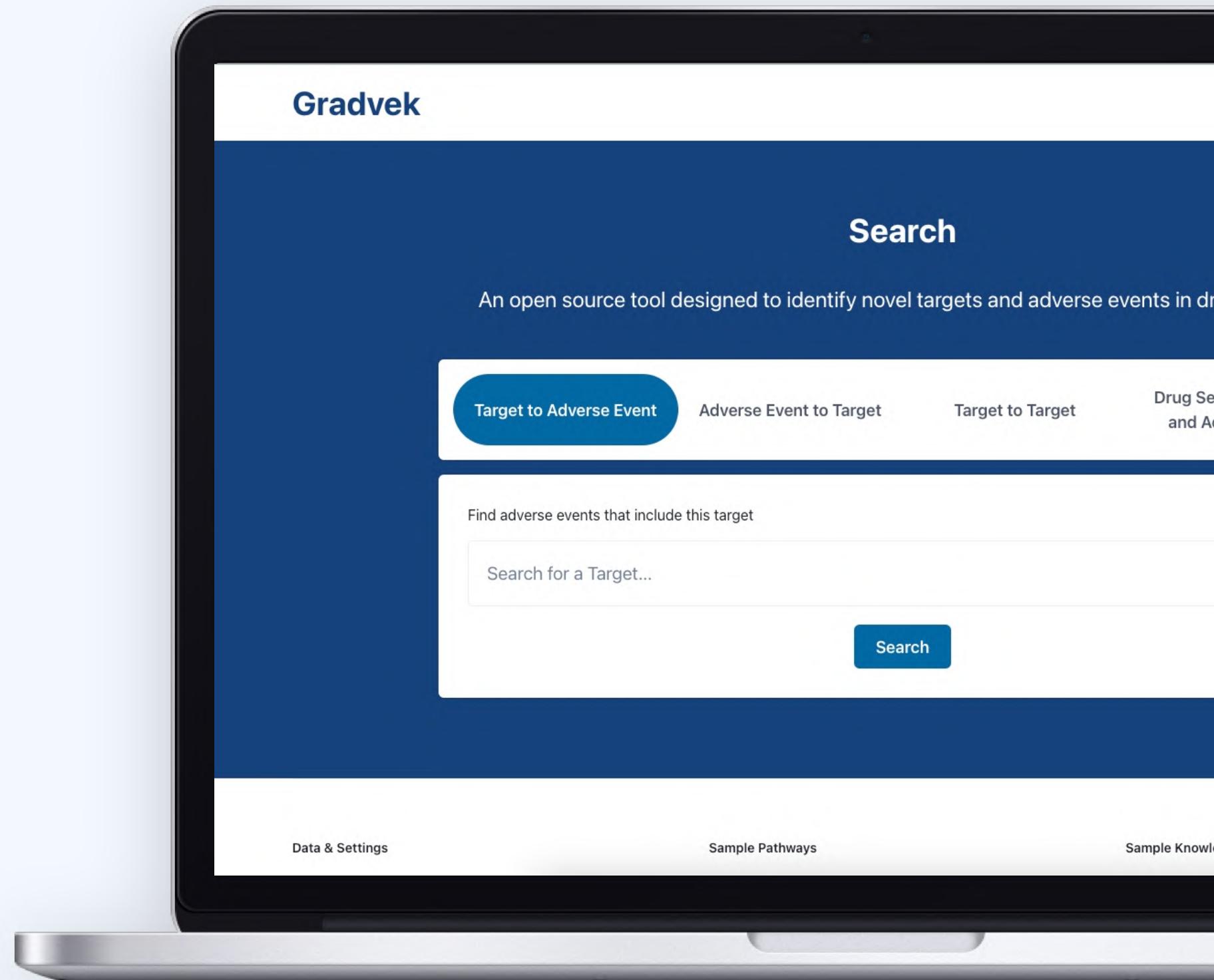
03-milestones

04-next steps

Gradvek 2.0

Gradvek is a tool designed to identify novel targets and adverse events in drug research.

It serves as a comprehensive resource connecting various entities such as drugs, adverse events, targets, and genes.



● Team

Frontend

David Aviles

Backend

Alyssa Bédard
Kevin Patel
Nathan Sheely

Research

Dixon Gross
Ali PiyarAli

NEXT

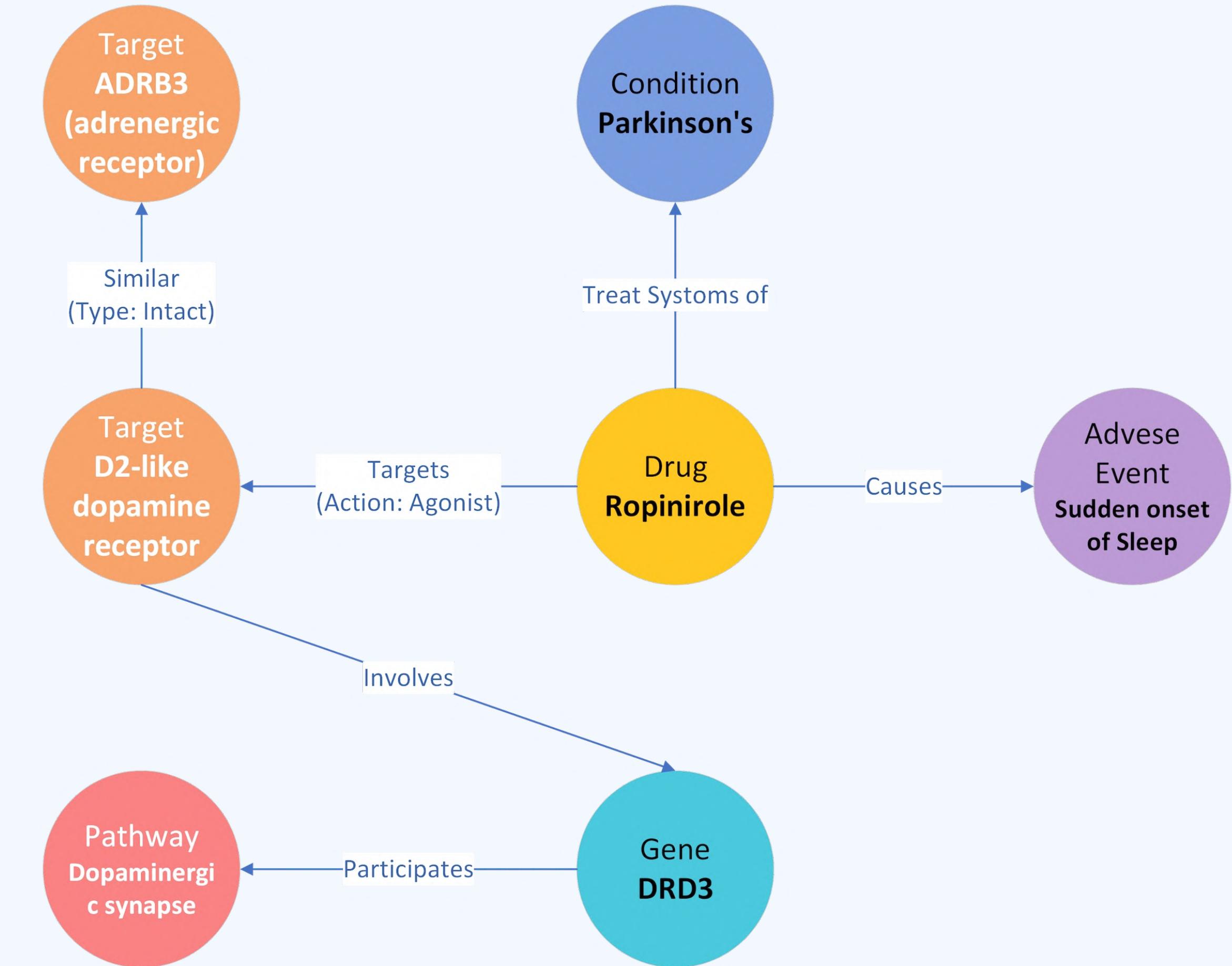
django 


 Open Targets

DRD3: Dopamine Receptor D3

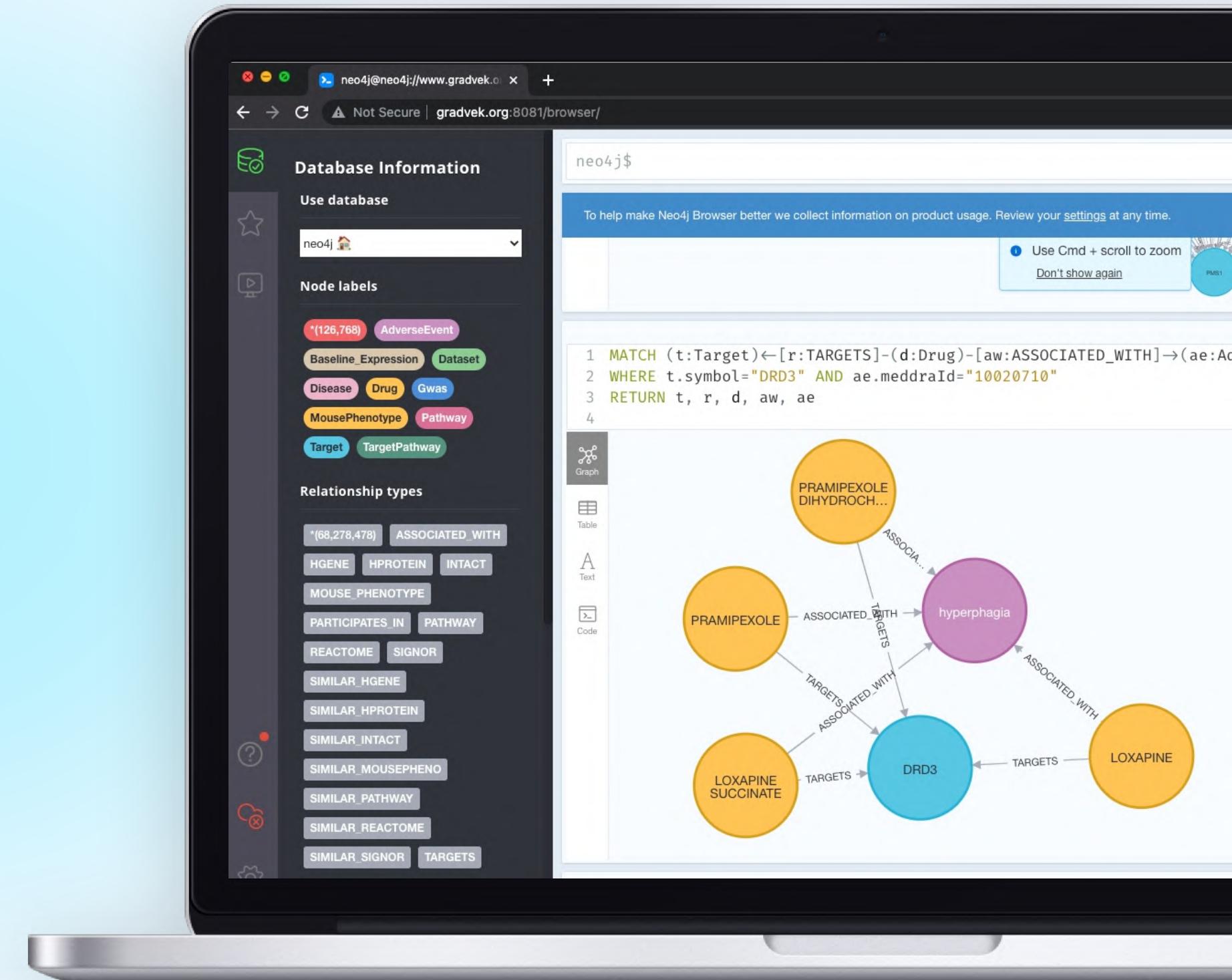
Target proteins are the proteins that drugs aim to bind to and modify their function for therapeutic purposes

Novel targets refer to previously unidentified or undiscovered proteins that have the potential to be modulated by drugs for therapeutic purposes



Knowledge Graph

Demo



neo4j@neo4j://gradvek.org:8047

Not Secure | gradvek.org:8081/browser/

Database Information

Use database

neo4j

Node labels

- *(126,768) AdverseEvent
- Baseline_Expression
- Dataset
- Disease
- Drug
- Gwas
- MousePhenotype
- Pathway
- Target
- TargetPathway

Relationship types

- *(68,278,478) ASSOCIATED_WITH
- HGENE
- HPROTEIN
- INTACT
- MOUSE_PHENOTYPE
- PARTICIPATES_IN
- PATHWAY
- REACTOME
- SIGNOR
- SIMILAR_HGENE
- SIMILAR_HPROTEIN
- SIMILAR_INTACT
- SIMILAR_MOUSEPHENO
- SIMILAR_PATHWAY
- SIMILAR_REACTOME
- SIMILAR_SIGNOR
- TARGETS

Property keys

- actionType
- adverseEventId
- chemblId
- critval
- dataset

neo4j\$

To help make Neo4j Browser better we collect information on product usage. Review your [settings](#) at any time.

```
1 MATCH (t:Target)-[r:TARGETS]-(d:Drug)-[aw:ASSOCIATED_WITH]→(ae:AdverseEvent)
2 WHERE t.symbol="DRD3" AND ae.meddraId="10020710"
3 RETURN t, r, d, aw, ae
```

Graph

Table

Text

Code

Overview

Node labels

- * (6) Target (1) AdverseEvent (1) Drug (4)

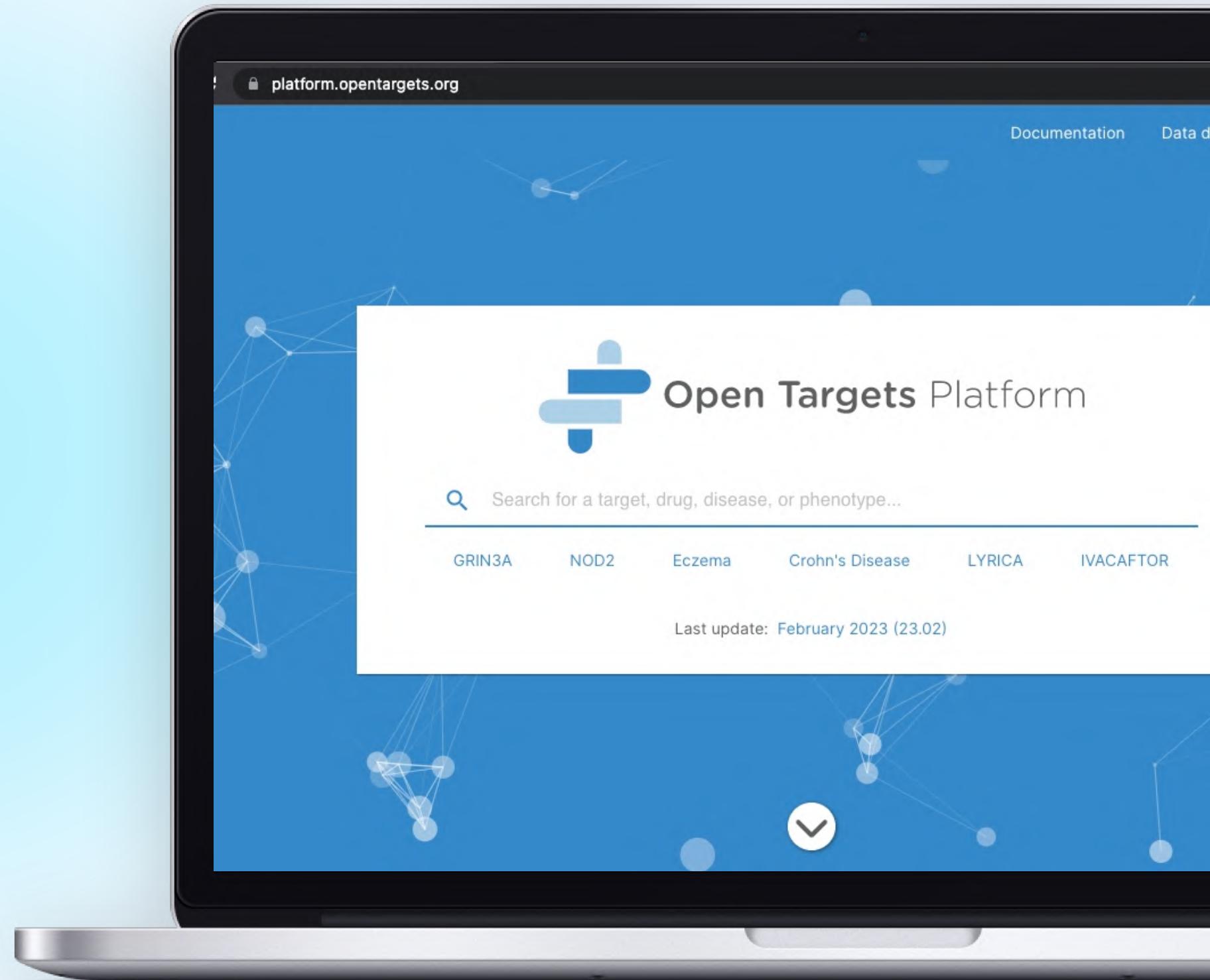
Relationship types

- * (8) TARGETS (4) ASSOCIATED_WITH (4)

Displaying 6 nodes, 8 relationships.

Open Targets

Demo



DRD3 profile page | Open Targets

platform.opentargets.org/target/ENSG00000151577

DRD3 dopamine receptor D3

Ensembl: ENSG00000151577 | UniProt: P35462 | GeneCards: DRD3 | HGNC: DRD3

View DRD3 in Open Targets Genetics

Associated diseases **Profile**

KD Description Dopamine receptor whose activity is mediated by G proteins which inhibit adenylyl cyclase. Promotes cell proliferation.

S Synonyms D(3) dopamine receptor, DRD3, Dopamine D3 receptor, D3DR, ETM1, FET1, essential tremor 1

TR Tractability Assessment available

CP Safety no data

BE Chemical Probes 4 chemical probes

GO Baseline Expression RNA

GC ProtVista Positional, Structural and Functional Information

PV Genetic Constraint medium constraint

MI Molecular Interactions 1029 physical or functional interactors

PV Cancer Hallmarks no data

MI Mouse Phenotypes 38 distinct phenotypes

PW Comparative Genomics 9 orthologues and 25 paralogues

CH Subcellular Location 1 subcellular locations

B Bibliography 2,281 publications

KD Known Drugs 38 drugs with 127 indications

CG Known Drugs Clinical precedence for drugs with investigational or approved indications targeting **DRD3** according to their curated mechanism of action. Source: ChEMBL.

Search

Download table as **JSON** **TSV** **API query**

Drug information Disease information Clinical trials information

https://platform.opentargets.org/target/ENSG00000151577

01–problem

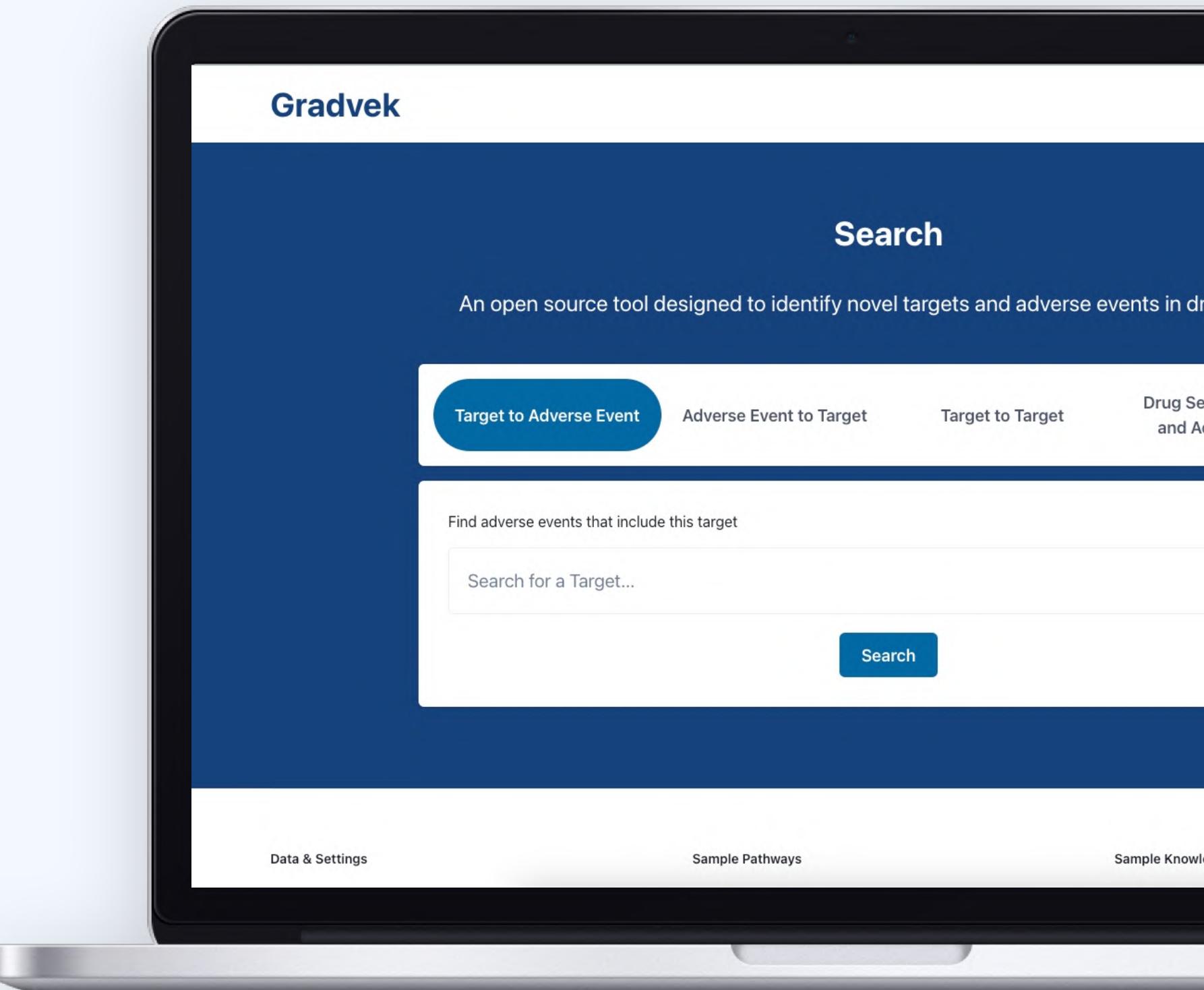
02–solution

03–milestones

04–next steps

Demo

<http://www.gradvek.org/>



01-problem

02-solution

03-milestones

04-next steps

Starting to build

- Rough front end build
- Decision to rebuild backend
- Downloading and parsing data into Neo4J



M4 ?

Milestone 1

Milestone 2

Milestone 3

Next steps

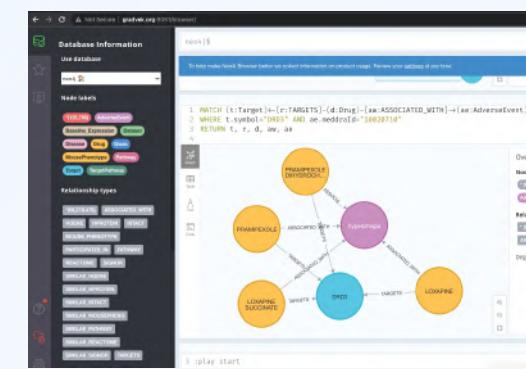
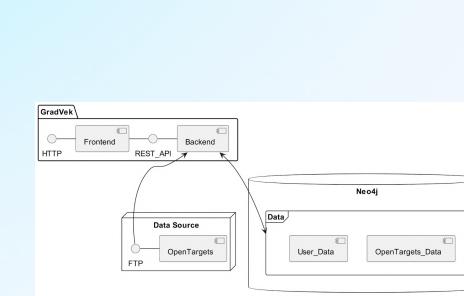
Learning and Planning

- Learning about underlying science
- Researching tools such as Neo4J
- Reading and understanding GradVek 1.0 code



Building and integrating

- Makefile, ./check_env, and Docker
- Similarity endpoints
- UI results page implementation
- **Challenge:** Dealing with large sets of data



Literature Similarity

The Open Targets literature similarity score is a Word2Vec model, not a data source

Clarified Data Discrepancies

Reactome molecular interactions seem to be missing where Reactome Pathways data shows an existing relationship

Further UI enhancements

Site copy tweaks, and further user experience refinement still possible

● Safety Knowledge Graph : GradVek 2.0 – Highlights

Target to Target similarity search

Pairwise similarity calculated across 8 data sources to help users find promising drug targets

Nodes and Relationships

Nodes: 124,519

Relationships: 68,278,478

Redesigned user interface

Prioritizing insights over information overload

Refactored and optimized code base

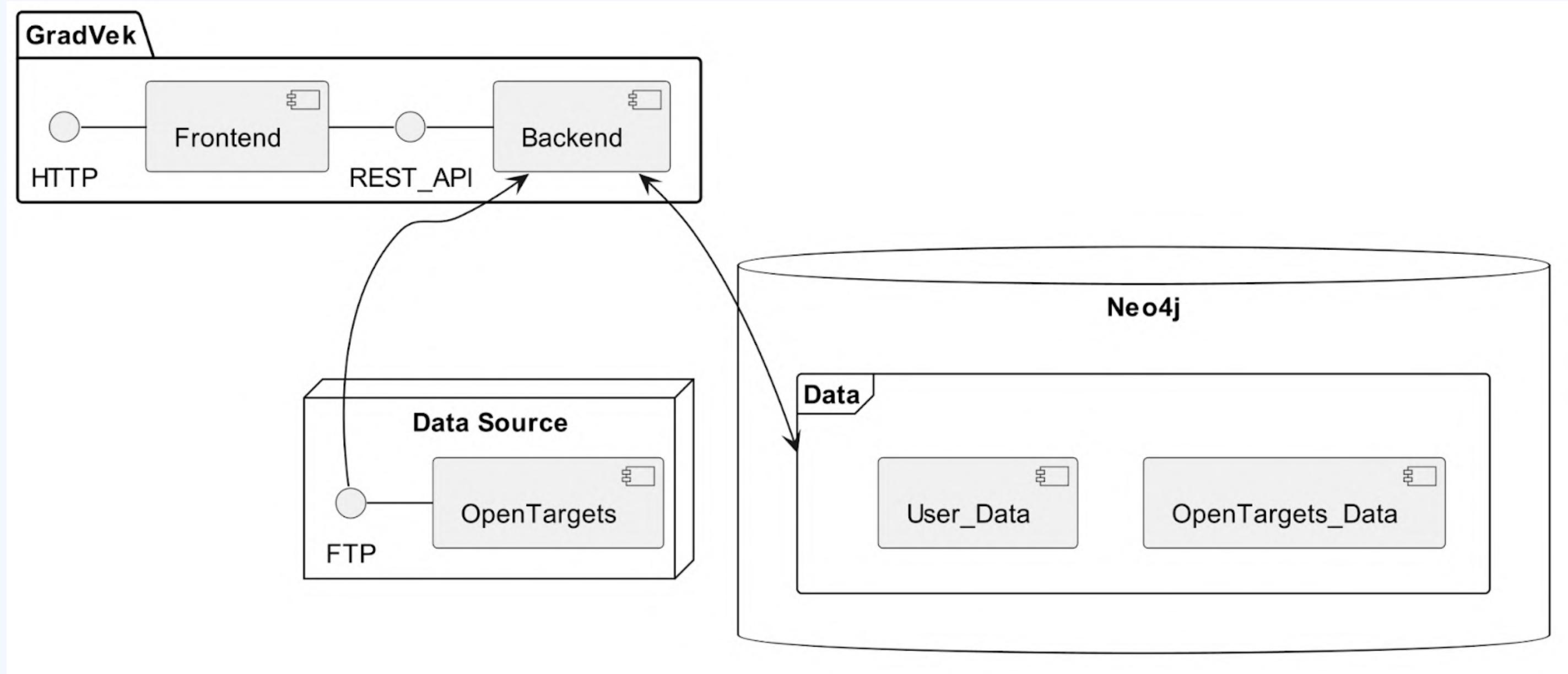
Data load time improved from over an hour to a few minutes for original sources

GradVek 2.0

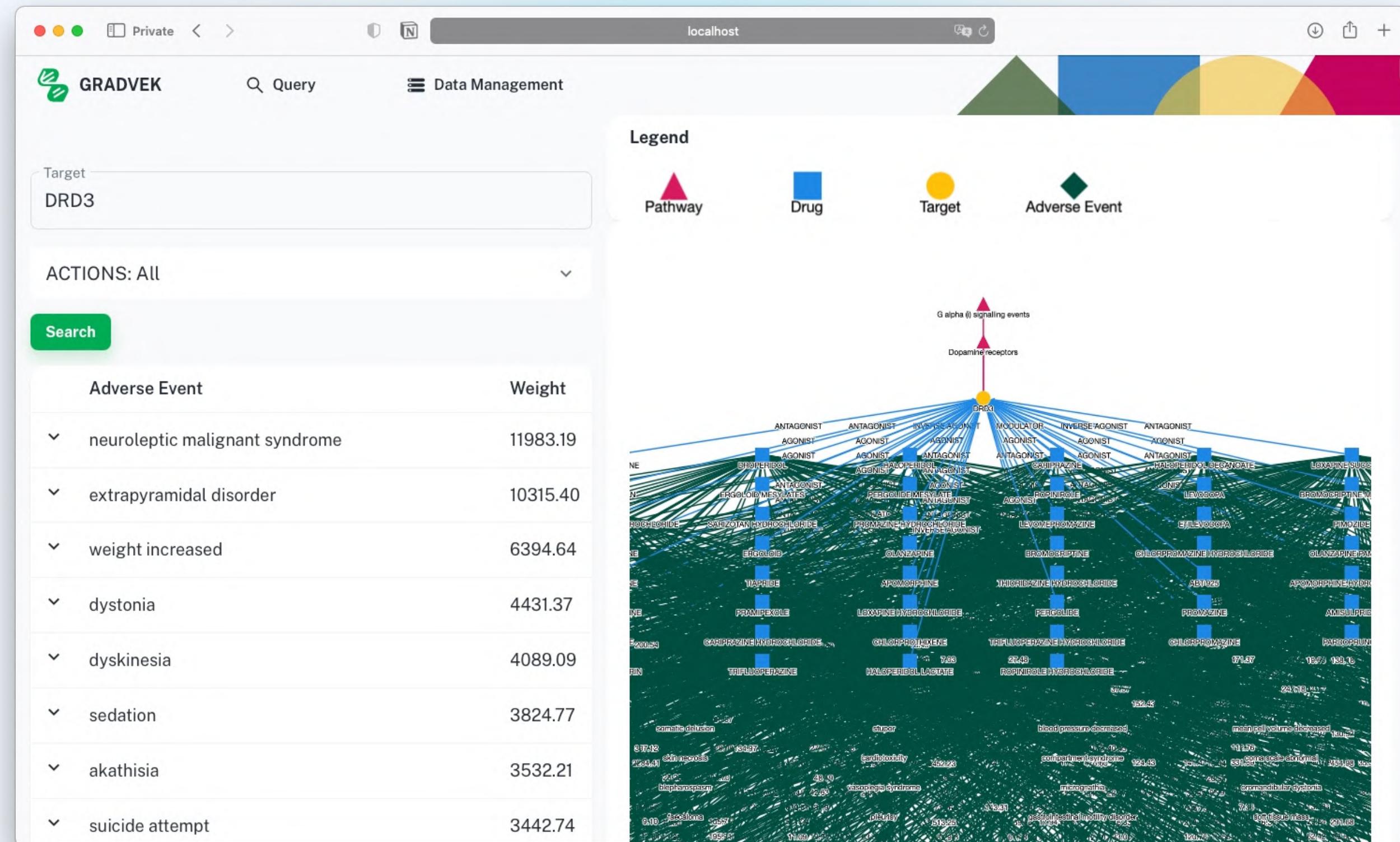
Appendix

Wednesday, May 10 2023

Architecture Diagram - simplified



GradVek 1.0 UI



Application Requirement

UI/UX

Redesign of 1.0 to easily show insights

Data Updates

Quarterly / On Demand Updates of Data

Computation

Similarity Calculations for Descriptors & Targets

Knowledge Graph

Rebuild Knowledge Graph with new data

API

Public API of Data

Weights/ Threshold

Ability to set min threshold and recompute data

Performance Requirements

1 UI must retrieve **results quickly** (pre-calculate)

2 Focus on **useful data** (based on minimum threshold)

3 Generate Knowledge Graph with **filters** in a **timely manner**

4 Provide **feedback** to **user** while all data loads

Initial Timeline

1 pt = 5 hours

- 15 % - UI/UX - 19 pt (~ 95 Hours)
- 25 % - UI Implementation - 35 pt (~ 175 Hours)
- 5% - Data Ingestion - 7 pt (~ 35 Hours)
- 15% - Computation & Processing - 19 pt (~ 95 Hours)

M2

-
- 10 % - Knowledge Graph - 11 pt (~ 55 Hours)
 - 10 % - API - 16 pt (~ 80 Hours)
 - 20 % - Testing - 25 pt (~ 125 Hours)

M3



Risks

- Computing **similarity scores** and developing appropriate **thresholds** will involve frequent customer input
- We are assuming a **simple similarity metric**, a **Jaccard Index**, will be able to provide some useful information
- Several pieces of planned work depend on having those **similarity metrics**
- This tool is intended to actually be used to assist with **drug research**

Open Target Data Types (Descriptors)

| | |
|--|--------------------------------------|
| Signor | Signaling interactions database |
| Reactome | Pathways and reactions |
| Intact | Molecular interactions |
| Mouse Phenotypes | Mouse phenotypic data |
| Tissues affected by Gene Expression | Tissues where genes are expressed. |
| Tissues affected by Protein Expression | Tissues where proteins are expressed |
| GWAS Traits | Genetic association with traits |
| Pathways | Functional pathways |

DNA, Gene, Protein

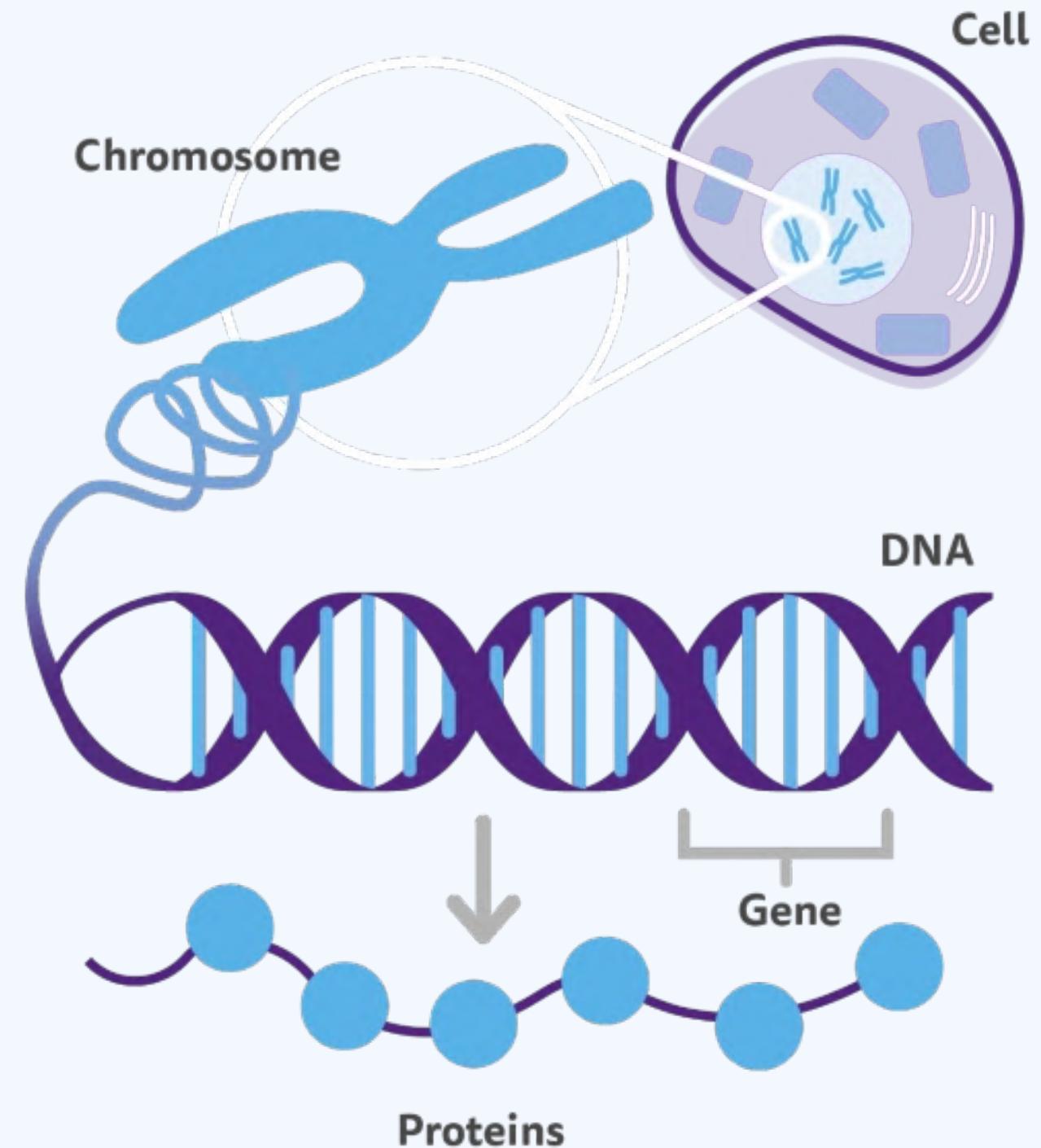
Cells: The smallest unit of life. Your body is made up of trillions of cells.

Chromosomes: Within each cell there are 23 pairs of chromosomes that you inherit from your birth parents.

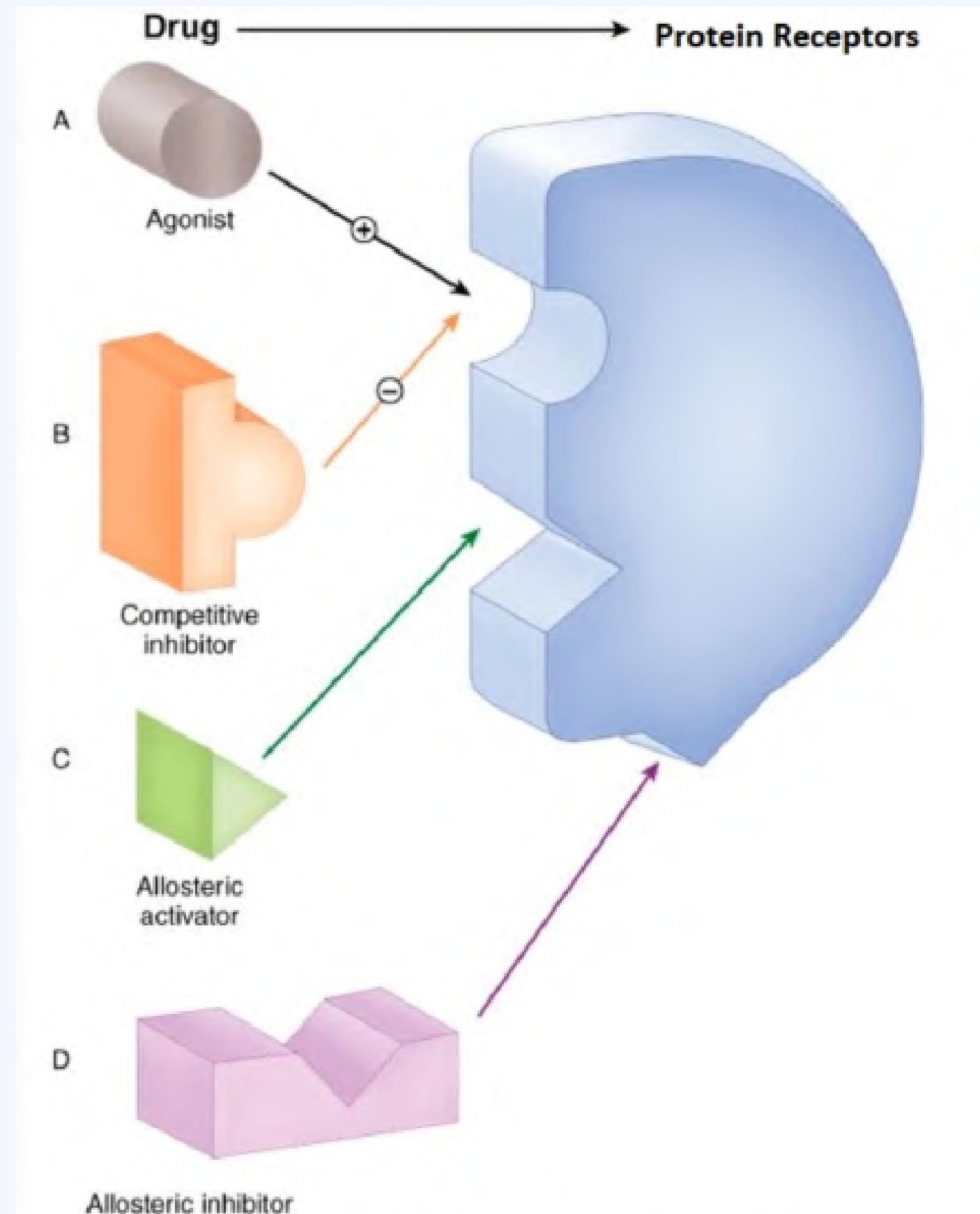
DNA: Each chromosome is made of strands of deoxyribonucleic acid, also known as DNA. DNA carries the genetic information needed to build and maintain everything in your body.

Genes: Your DNA is made up of different sections called genes. Each gene is responsible for creating proteins that perform different tasks.

Proteins: Proteins are created by genes to carry out specific tasks in your body. Some of these tasks shape how you look or your body's ability to work smoothly.



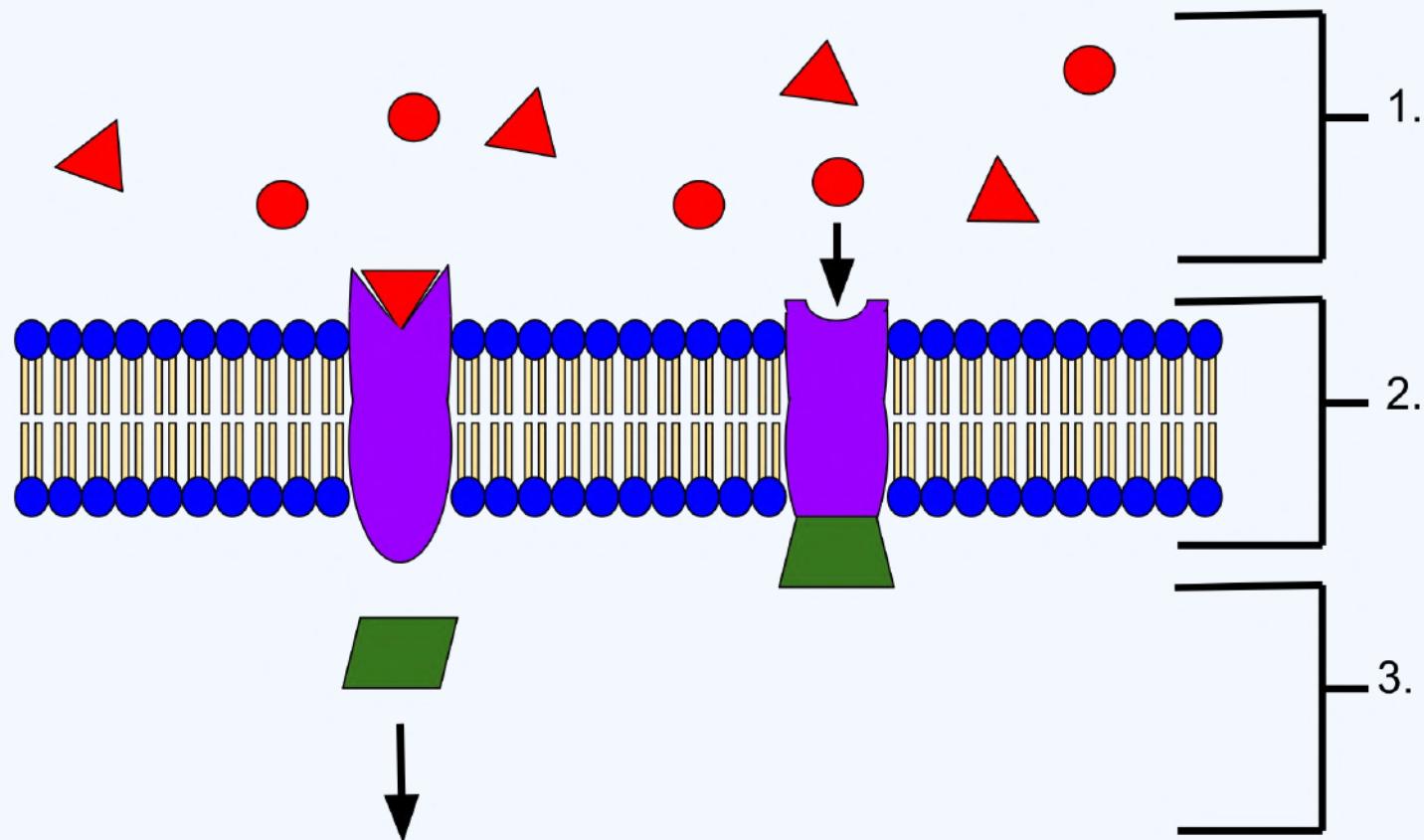
- **Agonist:** A drug that binds to a receptor and activates it, mimicking the action of an endogenous substance.
- **Antagonist:** A drug that binds to a receptor and blocks its activation by preventing the binding of other molecules, thereby inhibiting the receptor's function.
- **Inhibitor:** A drug that inhibits or decreases the activity of a specific enzyme, receptor, or other molecular target.
- **Activator:** A drug that enhances or increases the activity of a specific enzyme, receptor, or other molecular target.
- **Allosteric modulator:** A drug that binds to a specific site on a protein (allosteric site) and modulates its activity, either enhancing or inhibiting its function.
- **Antimetabolite:** A drug that interferes with the normal metabolic processes of cells by mimicking or blocking essential molecules.
- **Blocker:** A drug that blocks the activity of specific ion channels or transporters, thereby modulating the flow of ions or molecules across cell membranes.
- **Modulator:** A drug that modulates the function of a protein or receptor, often by altering its conformation or activity.
- **Reuptake inhibitor:** A drug that blocks the reuptake of certain neurotransmitters, thereby increasing their concentration in the synaptic cleft and prolonging their activity.
- **Stabilizer:** A drug that stabilizes certain molecular structures, such as membranes or proteins, to restore or maintain their normal function.
- **Antagonistic pair:** A combination of drugs where one drug acts as an agonist while the other acts as an antagonist, resulting in a combined effect that modulates the target's activity.
- **Prodrug:** A biologically inactive compound that is metabolized in the body into an active drug form.



Source: Katzung BG, Masters SB, Trevor AJ: *Basic & Clinical Pharmacology*, 11th Edition: <http://www.accessmedicine.com>
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

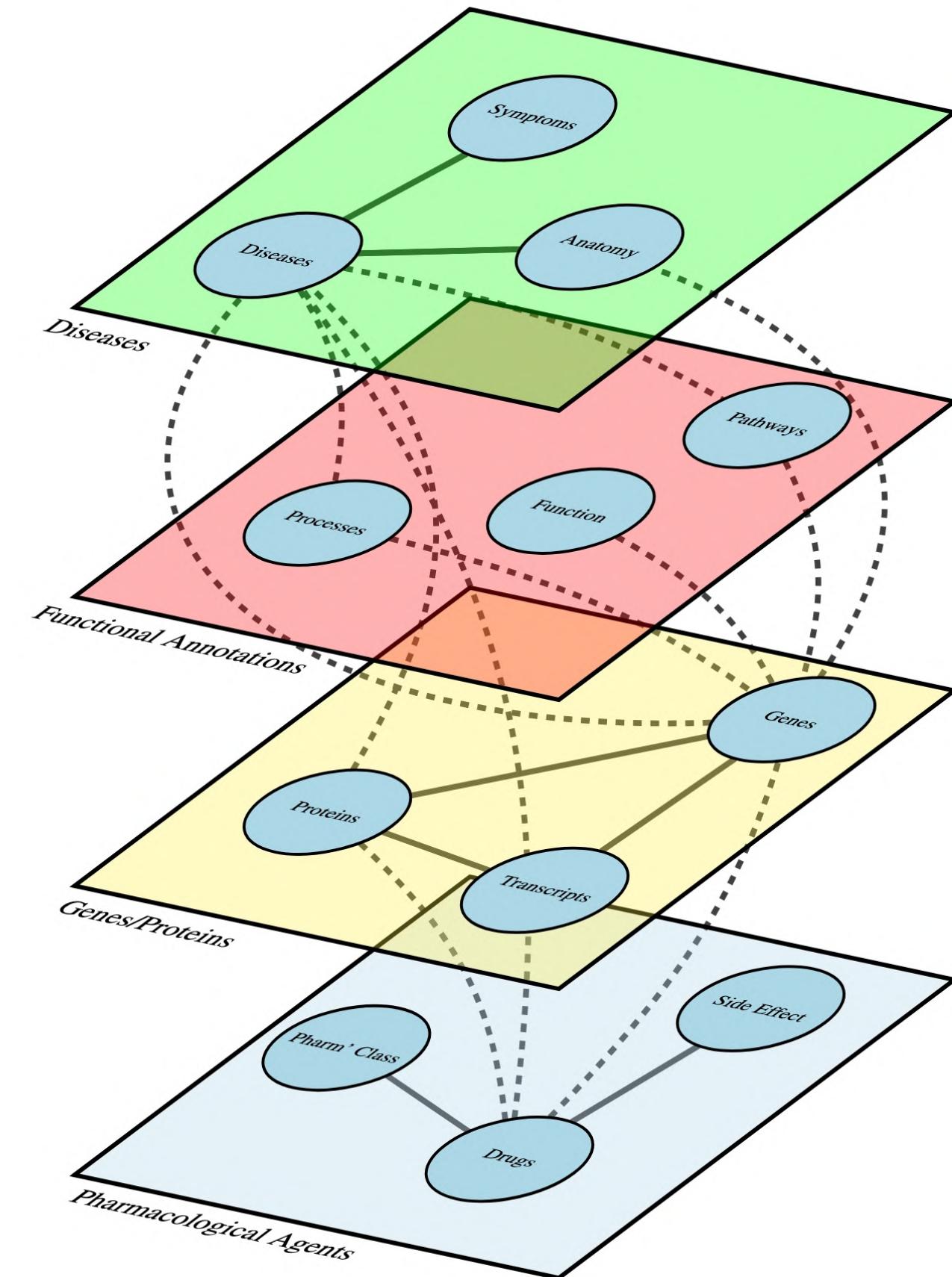
An example of membrane **receptors**.

- Ligands, located outside the cell
- Ligands connect to specific receptor proteins based on the shape of the active site of the protein.
- The receptor releases a messenger once the ligand has connected to the receptor.

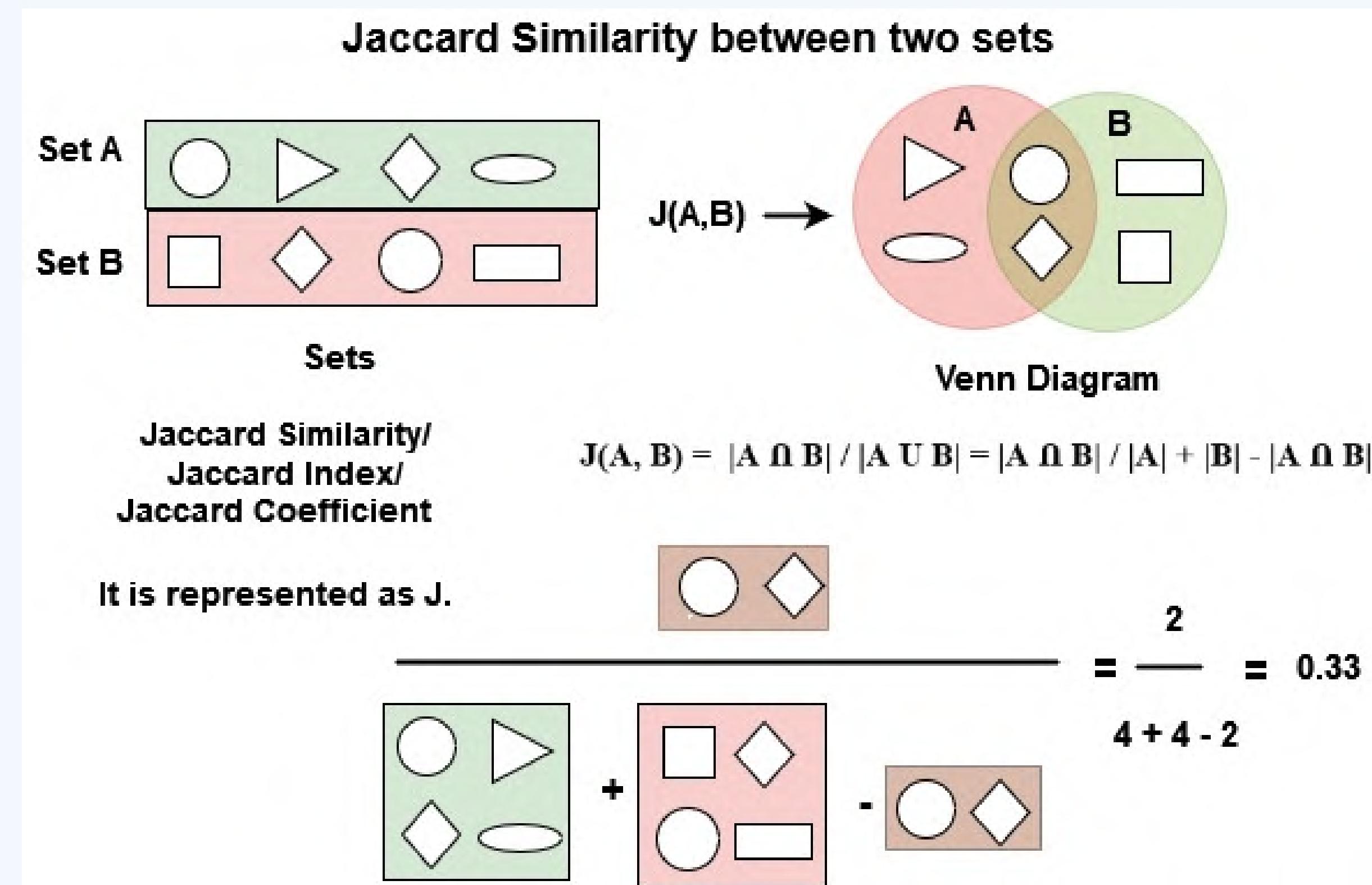


https://en.wikipedia.org/wiki/Receptor_%28biochemistry%29

Drug Discovery Knowledge Graph



Jaccard Similarity



Neo4j

