**Target super family classification**

In this project, we classified most targets into 13 protein superfamilies. Some of the remaining unclassified targets are placed under a 14th class, ‘Unknown’. The Names of these 14 classes are as follows:

1. Adhesion
2. Ion channel
3. Kinase
4. Enzyme
5. GPCR
6. Transporter
7. Membrane receptor
8. Secreted protein
9. Structural protein
10. Epigenetic regulator
11. Nuclear receptor
12. Transcription factor
13. Surface antigen
14. Unknown

A diagram of data flow

Description automatically generated

**Figure 1:** Schema for target protein tables (called TARGET INFORMATION) in ChEMBL 33 (cropped). From this PROTEIN\_CLASSIFICATION table contains target class information under column ‘PROTEIN\_CLASS\_DESC’. We mapped this column into 14 superfamilies (Supplementary File 1).

These 14 classes were made from target class information available in the database schema of ChEMBL33 (**Table**: PROTEIN\_CLASSIFICATION, column: PROTEIN\_CLASS\_DESC, **Figure 1**). The column: PROTEIN\_CLASS\_DESC, contains only the sub-class information but doesn’t contain a super family. So, we manually mapped the sub-families to super families. For instance, we mapped subfamily: ‘enzyme kinase

protein kinase agc sgk’ to superfamily: ‘Kinase’, ‘enzyme transferase’ to ‘Enzyme’, ‘enzyme hydrolase’ to ‘Enzyme’, ‘membrane receptor 7tm3 smallmol neurotransmitter metabotropic glutamate receptor’ to ‘GPCR’, ‘auxiliary transport protein ca beta’ to ‘Transporter’, ‘enzyme reductase’ to ’Enzyme’ and so on. Supplementary File 1 shows the mapping of ‘ChEMBL\_PROTEIN\_CLASS\_DESC’ into our protein\_super\_class for 2978 target proteins in our database.

**Compound classifications**

1. **Molecular classification based on drug type**

This classification was obtained from DrugBank. Of 6259 molecules, 432 are biological molecules (antibodies, gene therapies, vaccines, recombinant therapeutic protein or other biological molecules), and 5827 are small chemical molecules.

1. **Compound classification based clinical status**

We also classified all the 6259 molecules into six classes based on their clinical status. These six classes are as follows:

* Approved
* Vet approved
* Nutraceutical
* Experimental
* Investigational
* Illicit

Approved drugs contain those molecules that have been officially accepted for commercialization for at least one of the indications. From the list of 6243 molecules, 2277 are approved drugs in our database. Vet-approved drugs are approved for at least one of the indications in animal treatments. There are 36 vet-approved drugs in our database. Nutraceuticals are drugs regulated and processed at a pharmaceutical grade and have a demonstrable nutritional effect. There are 30 nutraceuticals in our database. Experimental compounds have not yet been investigated or approved in clinical but have been experimentally shown to bind specific proteins in mammals, bacteria, viruses, fungi, or parasites. Our database contains 2975 experimental compounds. Investigational compounds have not been approved yet but are currently in clinical trials (Phase I, Phase II or III). There are 906 investigational compounds. Finally, illicit compounds include those that are legally banned or selectively banned in most developed nations (such as cocaine and heroin). There are 35 illicit compounds.

There were also some scenarios where drugs were placed under multiple clinical statuses, e.g. Approved & investigational. This means that a particular drug is approved for an indication and is currently in clinical trials for another indication. We needed to have only one clinical status for each drug to implement the drug-target interaction network visualisation module. In such cases, we assigned the drug to an ‘approved’ class (e.g., Denileukin diftitox). The same was the rule applied for ‘Vet approved’. For cases where drugs were assigned both ‘illicit’ and ‘experimental’ clinical statuses, we placed those drugs under the ‘illicit’ category (e.g. drug, Methadyl acetate).

**Interactive Visualization for Drug-Target Interactions**

The drug-target interaction is a highly interactive D3-based module (a JavaScript library) that displays drug-target interactions for the selected drug(s). We used the directed graph available in the D3 library. Furthermore, we used HTML and CSS to improve the visualizations further.

The visualization module contains two nodes: a drug node (pill-shaped) and a target protein node (round-shaped). The target protein nodes are colour-coded according to 14 superfamilies as defined in the section above. Upon clicking a target protein, a popup menu displays protein 3D structures and other protein information such as sequence (fasta format), protein names and family class. The 3D structures of the proteins were obtained by querying proteins in PDB and selecting the top hit. In the case of proteins whose structures are not available in PDB, we obtained predicted structures from Alphafold. We wrote scripts to scrap structures for these 2978 proteins.

The drug node (represented by pill shape) has two colours; the left colour represents molecule type, i.e.

Small chemical molecules or biological molecules. The right colour on the pill represents the clinical status of the molecule, e.g. approved, vet-approved, illicit, investigational, experimental and nutraceuticals. The web module allows end users to change the colours for both drug and protein nodes. Upon clicking a drug node, a pop-up displays more information on the drug, such as drug 2D structures, toxicities, physiochemical properties, etc.

Upon clicking on the link between the drug and the target node, a popup displays unique mutations related to that interaction.

Finally, the target protein node's size depends on the number of connected drugs to that particular target. Target nodes associated with a bigger number of drugs have bigger size.

**Interactive Features**

* **Filtration:** End users can filter out nodes based on target super family, drug clinical status or drug type.
* **Threshold (slider):** End users can play with slider (threshold) to bring more nodes (if any) to the screen.
* **Changing the location of nodes:** End users can drag target or drug nodes to change its position on the screen.
* **Zoom:** End users can zoom in/out to increase the size of nodes and links on the screen.
* **Redraw:** Upon clicking the redraw button, the visualization module re-adjusts the network to optimal positions.
* **Export option:** After doing filtering or adjusting the network as per the needs of the end users, he/she can export the high-quality figure. He/she can freely use the exported figure for his/her manuscript or any other purpose.