Web Scraper + API
Take-Home Challenge

Background: Drugbank is an online curated knowledgebase for approved, withdrawn and investigational drugs. Along with information of drug structures and clinical trials, DrugBank provides information on the gene targets of each drug, which is crucial in better understanding a drug's mechanism of action. While DrugBank does provide an XML download of its contents, it is frequently updated online in between releases, making web scraping necessary. We would like you to build a web scraper that can capture the gene name of all targets for a list of given drugs.

Assignment: We're asking that you 1) build a web scraper that collects the gene name of all targets for a predefined set of DrugBank IDs, and 2) build an API endpoint that accepts a DrugBank ID and returns the gene targets.

Task 1 Web Scraper Functional Requirements: Your program should do the following:

- Make network requests to fetch a list of gene names of all targets (when available) for a predefined set of DrugBank IDs
- Persist the DrugBank IDs along with their gene targets
- Be able to be run by a single terminal command

Task 2 API Functional Requirements: Your program should do the following:

- Accept a DrugBank ID and return gene target in a modern data exchange format
- Instructions showing how to stand up your API on a linux or macos machine
- Example GET url request or bash script that makes a curl request to your API

Non-Functional Requirements:

- Python, please!

Library suggestions & Ideas:
You have complete freedom to use the environment/packages you are most comfortable/familiar with!

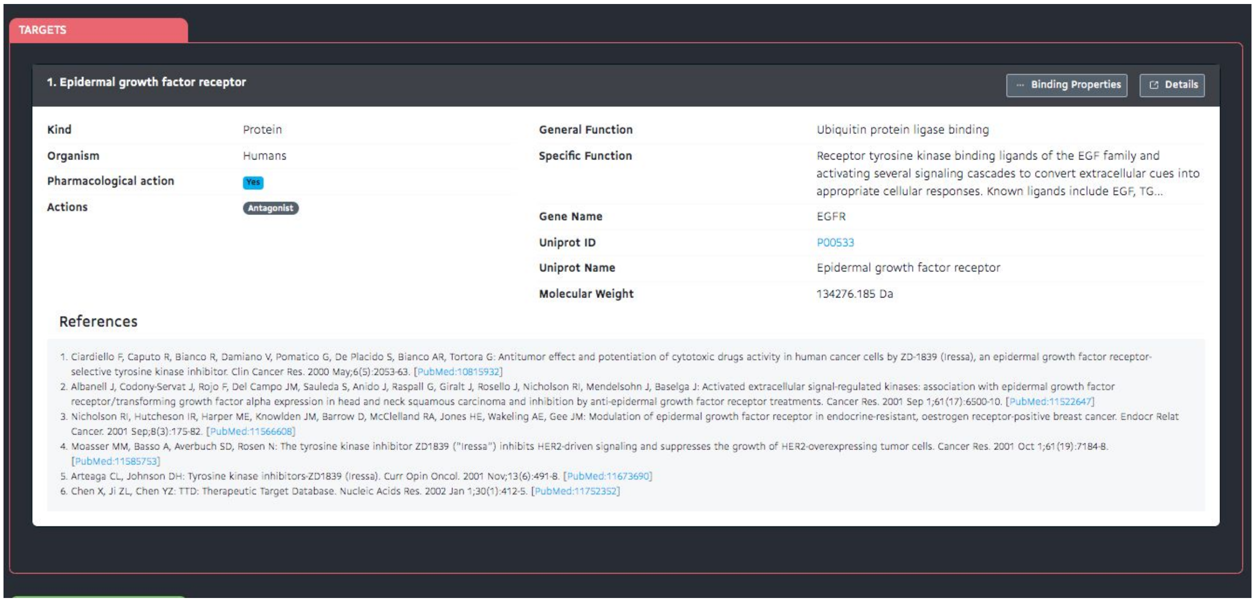
Deliverable: You should zip your code and e-mail the submission to coryandar@onethree.bio. A github repo would be even better.

Evaluation: You will be evaluated on the correctness of your implementation and the quality of your code. Proper functionality is the most important. Ideally, the code should be clean, easy to read, and well commented. Your code should be modular, with proper separation of concerns.

Task 1 Details (Predefined set of DrugBank IDs):
Scrape and store the approximate synonyms of the following IDs of interest*:
-DB00619
-DB01048
-DB14093
-DB00173
-DB00734
-DB00218
-DB05196
-DB09095
-DB01053
-DB00274

***Note:** extra credit if you can programmatically accomplish this for all drugs within DrugBank

An example of the data we are interested in (for DrugBank ID DB00317) can be seen under the "Targets" section in the following link: https://www.drugbank.ca/drugs/DB00317. We have also attached a screenshot below.



1. Epidermal growth factor receptor

Kind	Protein
Organism	Humans
Pharmacological action	Yes
Actions	Antagonist

General Function	Ubiquitin protein ligase binding
Specific Function	Receptor tyrosine kinase binding ligands of the EGF family and activating several signaling cascades to convert extracellular cues into appropriate cellular responses. Known ligands include EGF, TG...
Gene Name	EGFR
Uniprot ID	P00533
Uniprot Name	Epidermal growth factor receptor
Molecular Weight	134276.185 Da

References

1. Ciardiello F, Caputo R, Bianco R, Damiano V, Pomatico G, De Placido S, Bianco AR, Tortora G: Antitumor effect and potentiation of cytotoxic drugs activity in human cancer cells by ZD-1839 (Iressa), an epidermal growth factor receptor-selective tyrosine kinase inhibitor. Clin Cancer Res. 2000 May;6(5):2053-63. [PubMed:10815932]
2. Albanell J, Codony-Servat J, Rojo F, Del Campo JM, Sauleda S, Anido J, Raspall G, Giralt J, Rosello J, Nicholson RI, Mendelsohn J, Baselga J: Activated extracellular signal-regulated kinases: association with epidermal growth factor receptor/transforming growth factor alpha expression in head and neck squamous carcinoma and inhibition by anti-epidermal growth factor receptor treatments. Cancer Res. 2001 Sep 1;61(17):6500-10. [PubMed:11522647]
3. Nicholson RI, Hutcheson IR, Harper ME, Knowlden JM, Barrow D, McClelland RA, Jones HE, Wakeling AE, Gee JM: Modulation of epidermal growth factor receptor in endocrine-resistant, oestrogen receptor-positive breast cancer. Endocr Relat Cancer. 2001 Sep;8(3):175-82. [PubMed:11566608]
4. Moasser MM, Basso A, Averbuch SD, Rosen N: The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER2-overexpressing tumor cells. Cancer Res. 2001 Oct 1;61(19):7184-8. [PubMed:11585753]
5. Arteaga CL, Johnson DH: Tyrosine kinase inhibitors-ZD1839 (Iressa). Curr Opin Oncol. 2001 Nov;13(6):491-8. [PubMed:11673690]
6. Chen X, Ji ZL, Chen YZ: TTD: Therapeutic Target Database. Nucleic Acids Res. 2002 Jan 1;30(1):412-5. [PubMed:11752352]

As an example, our final datastore would contain all of the gene targets for DB00317 (Gefitinib), which includes only one target, EGFR.

Task 2 Details (API Endpoint):

Feel free to use any technologies you are most comfortable with. Again, the API should accept a DrugBank ID and return the approximate synonyms. To verify functionality, you should expect to run the API locally on your machine and provide a curl command in bash that sends the request and displays the response to us.