



Article

# DRUGPATH: The Drug Gene Pathway Meta-Database

Rajeev Jaundoo <sup>1,†</sup> and Travis J. A. Craddock <sup>2,3,\*</sup>

<sup>1</sup> Department of Biomedical Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada; jaundoo@ualberta.ca

<sup>2</sup> Institute for Neuro-Immune Medicine, Nova Southeastern University, Fort Lauderdale, FL 33313, USA

<sup>3</sup> Departments of Psychology and Neuroscience, Computer Science, and Clinical Immunology, Nova Southeastern University, Fort Lauderdale, FL 33313, USA

\* Correspondence: tcraddock@nova.edu

† These authors contributed equally to this work.

Received: 29 February 2020; Accepted: 21 April 2020; Published: 30 April 2020



**Abstract:** The complexity of modern-day diseases often requires drug treatment therapies consisting of multiple pharmaceutical interventions, which can lead to adverse drug reactions for patients. *A priori* prediction of these reactions would not only improve the quality of life for patients but also save both time and money in regards to pharmaceutical research. Consequently, the drug-gene-pathway (DRUGPATH) meta-database was developed to map known interactions between drugs, genes, and pathways among other information in order to easily identify potential adverse drug events. DRUGPATH utilizes expert-curated sources such as PharmGKB, DrugBank, and the FDA's NDC database to identify known as well as previously unknown/overlooked relationships, and currently contains 12,940 unique drugs, 3933 unique pathways, 5185 unique targets, and 3662 unique genes. Moreover, there are 59,561 unique drug-gene interactions, 77,808 unique gene-pathway interactions, and over 1 million unique drug-pathway interactions.

**Keywords:** polypharmacology; database; gene; drug; target; combination therapy

## 1. Introduction

The concept of “one disease, one drug” is no longer relevant when applied to complex chronic diseases where drug combination treatments can lead to adverse drug-drug interactions [1,2]. In HIV for example, protease inhibitors such as ritonavir are used but have been known to affect the metabolism of other drugs [3]. To ameliorate adverse interactions such as these, a technique known as targeted drug delivery attempts to reduce the possibility of side effects and other drawbacks of treatment by delivering drugs to specific targets in order to localize their pharmacological activity [2], however, pharmaceuticals tend to be non-specific and on average bind to at least six molecular targets [4], which can lead to off-target interactions that increase the complexity of treatment. That being said, drug repurposing can also be performed where the off-target interactions of a pharmaceutical are utilized to treat diseases not originally targeted. In any case, the ability to predict all interactions of a given drug combination therapy would allow for more effective treatment regimens with fewer side effects, leading to better outcomes towards patient health.

One example of a combination therapy is the use of enbrel and mifepristone as treatment course proposed by our group for Gulf War Illness (GWI) [5,6]. GWI is a chronic multi-symptom illness characterized by gastrointestinal, muscle, and cognitive dysfunction stemming from exposure to battlefield neurotoxins during the 1990–91 Persian Gulf War [5–7]. Prior analysis of gene expression in peripheral blood mononuclear cells [8] as well as the dynamic regulatory modeling

of the neuro-endocrine-immune response to an exercise challenge [7,9] predicted that an initial inhibition of Th1 cytokines, followed by glucocorticoid receptor (GCR) inhibition, would bring patients from the chronic homeostatic regulation induced by neurotoxin exposure back to a more healthy, stable homeostatic regulatory state [7,9,10]. While subsequent analysis [5] suggested that the best treatment course may be accomplished using enbrel, a tumor necrosis factor alpha (TNF $\alpha$ ) inhibitor [11], in combination with the antiglucocorticoid mifepristone [12], great precaution must be used due to the novelty of this treatment regimen as well as the chemical sensitivity found in patients with GWI.

In order to address issues such as this, the drug-gene-pathway (DRUGPATH) meta-database was developed in order to provide a comprehensive mapping of possible interactions between drugs, genes, targets, and biological pathways to inform clinicians of potential adverse effects. The service is accessible at <http://drugpath.app>. To ensure validity of the resulting interaction database, and to test its functionality, we assessed two drug combinations. The first is the interaction between citalopram and phenelzine. This interaction was chosen as a validation exercise as combining citalopram, a selective serotonin reuptake inhibitor [13], and phenelzine, a monoamine oxidase inhibitor [14], is known to increase the risk of a rare but serious condition called serotonin syndrome, which may include symptoms such as confusion, hallucination, seizure, extreme changes in blood pressure, increased heart rate, fever, excessive sweating, shivering or shaking, blurred vision, muscle spasm or stiffness, tremor, incoordination, stomach cramp, nausea, vomiting, and diarrhea. As such, we expect to find overlap in pathways related to serotonin function as well as pathways related to the resulting symptoms of serotonin syndrome. The second interaction chosen as an exploratory drug combination of enbrel and mifepristone, as it is novel, and relevant to current therapeutic interventions in GWI.

## 2. Results

DRUGPATH amalgamates various expert-curated sources including the Pharmacogenomics Knowledgebase (PharmGKB), DrugBank, the HUGO Gene Nomenclature Committee (HGNC), Guide to Pharmacology (GTP), and the U.S. Food and Drug Administration's National Drug Code Database (FDA NDC) among others into a single resource available for querying using the graphical interface; see Table 1 for a complete list of all sources used. DRUGPATH contains interactions between drugs, genes, and pathways among other information (e.g., half-life, FDA approval status, indication) in order to easily identify potential adverse drug events (see Table 2 for statistics). Overall, DRUGPATH was developed in order to map the various interactions of a potential drug combination treatment, allowing researchers and clinicians alike to predict as well as mitigate side effects.

**Table 1.** The sources and their roles within the DRUGPATH database.

Source	Role
ConsensusPathDB (CPDB) [15,16]	pathways, pathway identifiers, gene symbols
DrugBank [17,18]	gene symbols, drug-target interactions, drug names, identifiers, half-lives, FDA approval status, and indications
FDA NDC [19]	drug names, FDA approval status
Guide to Pharmacology (GTP) [20]	drug names and identifiers, gene symbols, and targets
HGNC [21]	gene entrez, ensembl, and symbol identifiers
PharmGKB [22]	drug names, drug-gene interactions
repoDB [23]	drug names and indications
Toxin and Toxin-Target Database (T3DB) [24,25]	gene symbols, drug names, identifiers, half-lives, FDA approval status, and indications, drug-target interactions

**Table 2.** DRUGPATH's statistics.

Category	Available
Drugs	12,940
Genes	3662
Targets	5185
Pathways	3933
Drug-Gene Interactions	59,561
Drug-Pathway Interactions	1,053,633
Gene-Pathway Interactions	77,808

The interactions between two sets of drugs, citalopram and phenelzine as well as enbrel and mifepristone, are shown in the following sections using screenshots from the DRUGPATH graphical user interface, which allows users to easily query the DRUGPATH database in order to obtain the drug-gene and drug-pathway interactions (see Figure 1). This interface also allows users to search via gene ensembl, entrez, and symbol identifiers, target names, drug identifiers, indications, and half lives, FDA approval status, and finally pathway names. Any combination of the mentioned options is allowed, so searching for interactions between the TNF gene, "TNF", and the "signal transduction" pathway for example would be a valid query. Note that the link between a drug and gene is defined as a drug-gene interaction, and likewise, the link between a drug and target or a drug and pathway are drug-target and drug-pathway interactions respectively.

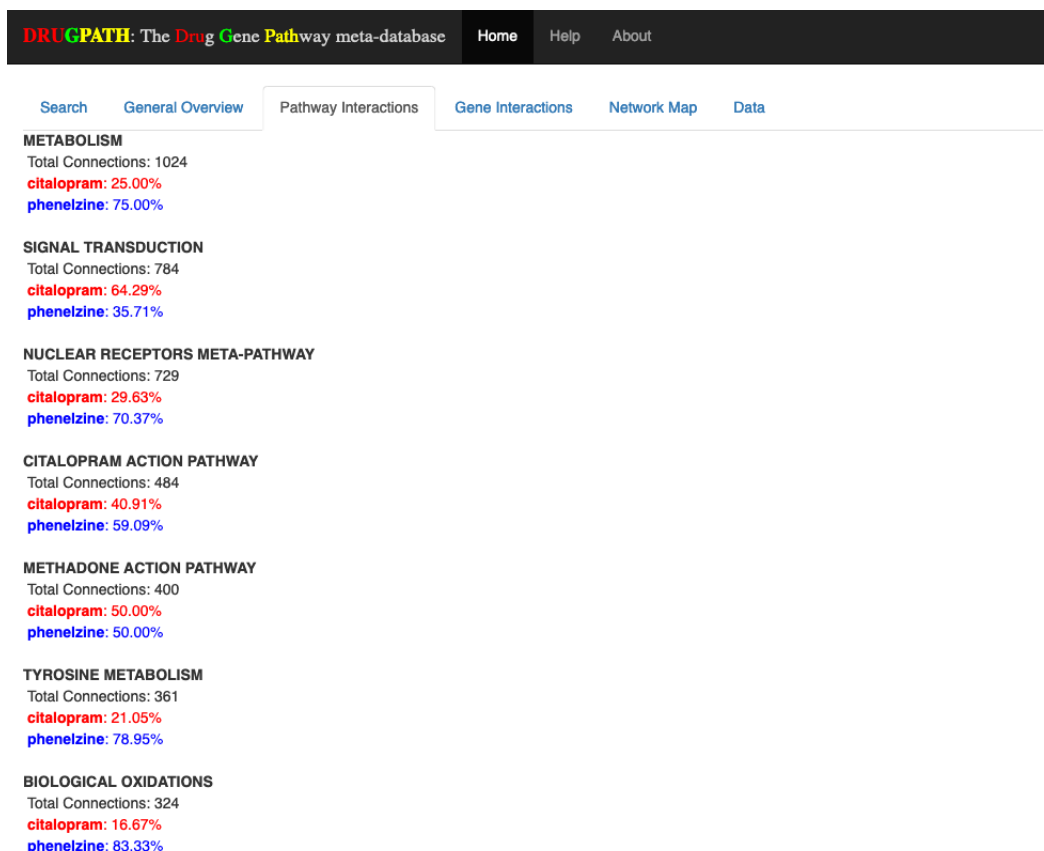
**Figure 1.** DRUGPATH Graphical User Interface.

## 2.1. Citalopram and Phenelzine

In this section the results from searching for all interactions between the drugs citalopram and phenelzine are shown. The output contained in each tab, as well as how to interpret it, is explained in each subsection below. Note that searching DRUGPATH via the graphical interface is shown in Materials and Methods.

### 2.1.1. Pathway Interactions

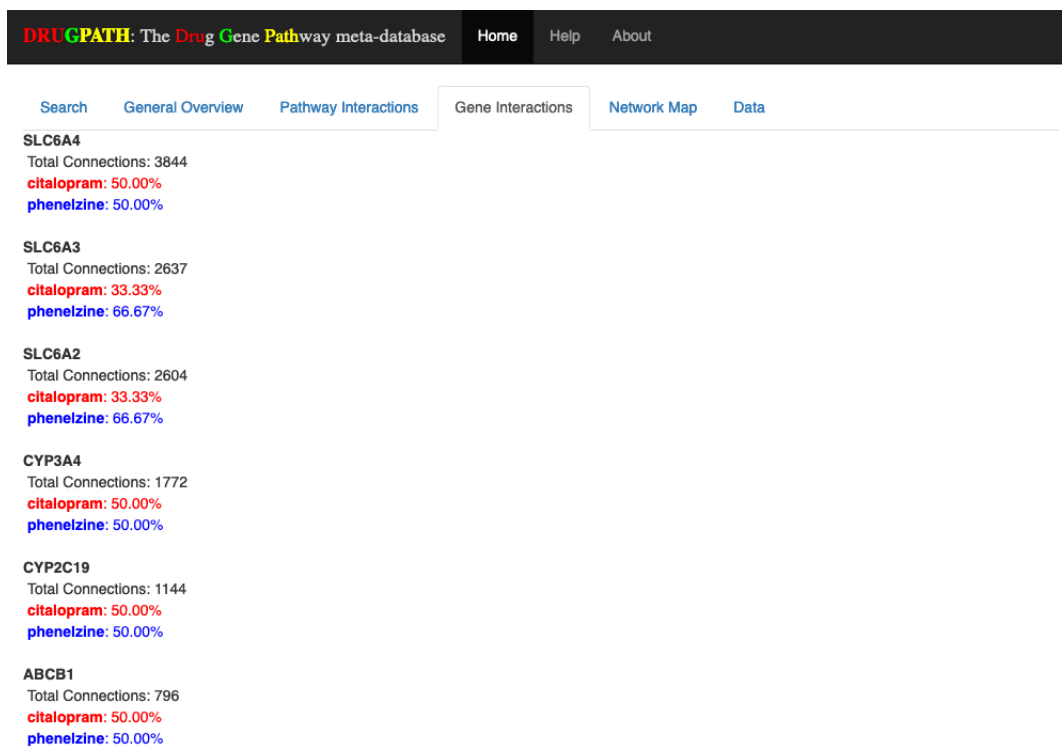
The pathways shared by both citalopram and phenelzine are shown in the Pathway Interactions tab in Figure 2, where the percentages next to each drug indicates how many of their total genes are associated to the pathway. For example, the "metabolism" pathway has a total of 1024 associated genes, the majority of which (75%) are associated with phenelzine. This type of information is valuable when developing drug combination treatments because potential drug-drug interactions can be identified and mitigated if possible, as in the case of the "signal transduction" pathway that is far more likely to be affected by citalopram than phenelzine. DRUGPATH found a total of 403 shared pathways between citalopram and phenelzine (see Supplementary Information S1).



**Figure 2.** Pathway interactions between citalopram and phenelzine. Note that only the first few results are shown.

### 2.1.2. Gene Interactions

Similar to the Pathway Interactions tab, the Gene Interactions tab shown in Figure 3 outputs the genes shared by all drugs. Here, the percentages next to each drug name correspond to how many connections the drug has to the particular gene, or in other words, the strength of association between the drug and gene. The number of connections refers to how many unique drug-gene-target entries there are in DRUGPATH. To illustrate, the cytochrome P450 3A4 (CYP3A4) gene has a total of 1772 connections stemming from citalopram and phenelzine, where each drug is just as likely as the other to affect this particular gene since their percentages are both 50%. On the other hand SLC6A3 is more likely to be affected by phenelzine than citalopram, so if this gene should not be affected during treatment it would be best to find a substitute. DRUGPATH found a total of 59 shared genes between citalopram and phenelzine (see Supplementary Information S2).



**Figure 3.** Genes interactions between citalopram and phenelzine. Note that only the first few results are shown.

### 2.1.3. Network Map

An interactive network map consisting of drug to gene to pathway interactions is shown in the Network Map tab in Figure 4. This was created using visNetwork [26] and igraph [27] with the Fruchterman and Reingold (FR) layout [28], which is a force directed layout where nodes are treated similar to atomic particles; connected nodes will ‘attract’ to one another and be pulled towards a particular direction depending on the number of connections [28,29]. The legend is displayed on the left of the map, where drug nodes are red, genes are blue, and pathways are green. Users can zoom in using either the scrollwheel on their mouse or by clicking the  $\ominus$  and  $\oplus$  buttons on the bottom right side which zooms outwards and inwards respectively. The arrow keys on the bottom left allow the user to move about the map, otherwise the mouse can be used to click and drag as well. When network maps have many drugs, genes, and pathways they can become crowded and difficult to view, so its helpful that clicking on a given node highlights only other connected nodes. For instance, clicking on the “TNF” node will only highlight drugs, genes, pathways directly connected to it; see Figure 5. Alternatively, the drop-down menu located in the top left allows for easy selection of any particular node without having to search and click.

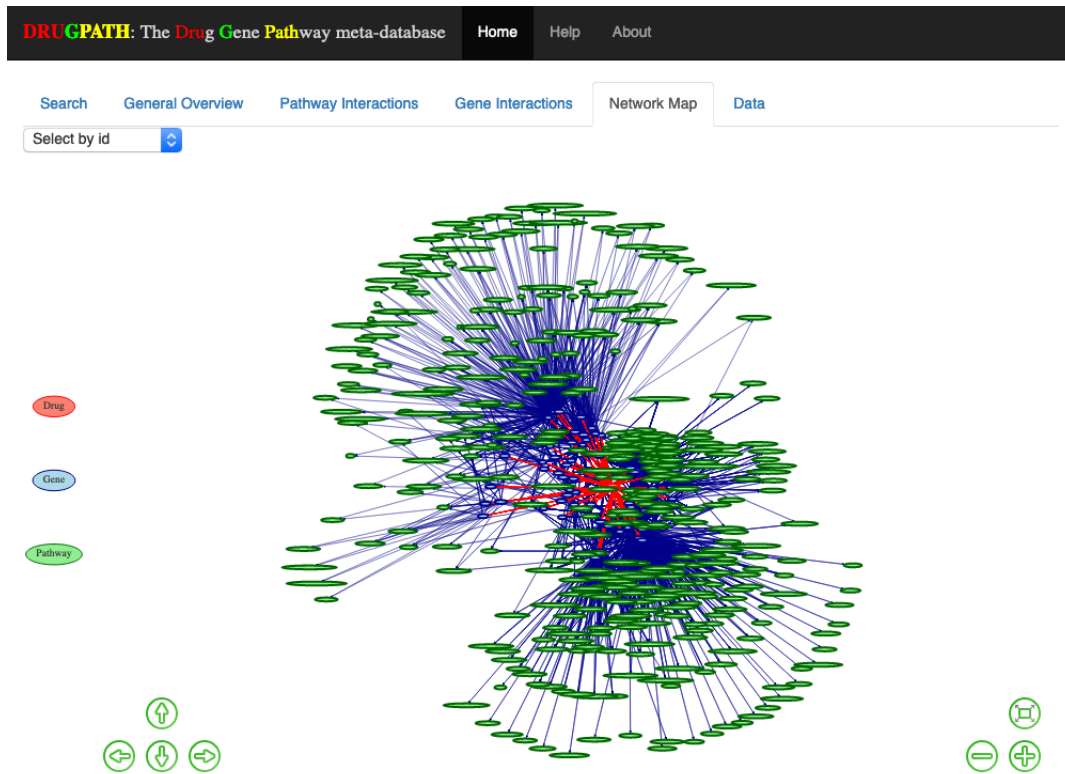


Figure 4. Overview of all the genes and pathways associated with citalopram and phenelzine.

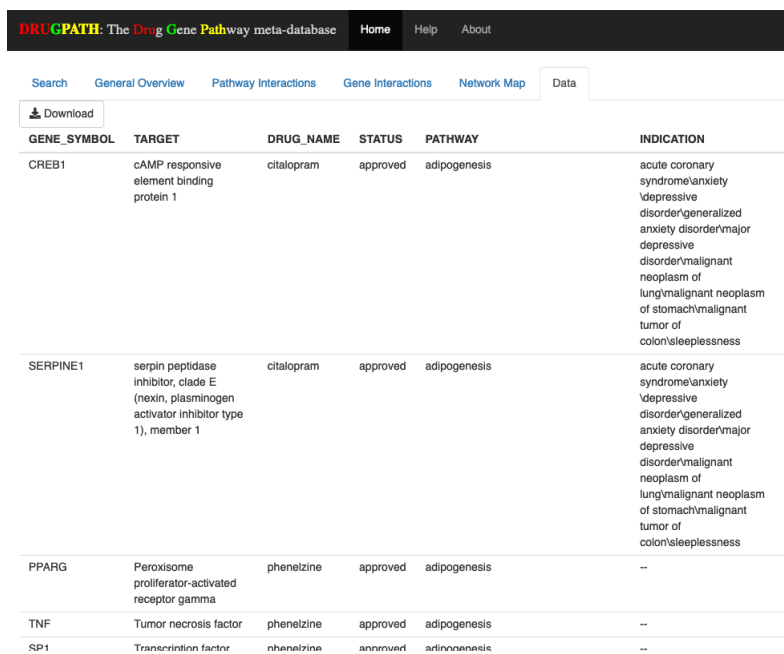


Figure 5. Closer examination of all drugs and pathways associated with the gene TNF. Any drug, gene, or pathway not associated with this gene are greyed out and made transparent.

### 2.1.4. Data

The Data tab shown in Figure 6 allows users to download the raw data corresponding to their search query in comma separated value (CSV) format. This data is pulled directly from the DRUGPATH database and is used to produce the information shown in all other tabs. In this tab users are able to scroll through all of the data, and to download they can simply click the “Download” button located in the top left.





GENE_SYMBOL	TARGET	DRUG_NAME	STATUS	PATHWAY	INDICATION
CREB1	cAMP responsive element binding protein 1	citalopram	approved	adipogenesis	acute coronary syndrome/anxiety disorder/generalized anxiety disorder/major depressive disorder/malignant neoplasm of lung/malignant neoplasm of stomach/malignant tumor of colon/sleeplessness
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	citalopram	approved	adipogenesis	acute coronary syndrome/anxiety disorder/generalized anxiety disorder/major depressive disorder/malignant neoplasm of lung/malignant neoplasm of stomach/malignant tumor of colon/sleeplessness
PPARG	Peroxisome proliferator-activated receptor gamma	phenelzine	approved	adipogenesis	--
TNF	Tumor necrosis factor	phenelzine	approved	adipogenesis	--
SP1	Transcription factor	phenelzine	approved	adipogenesis	--

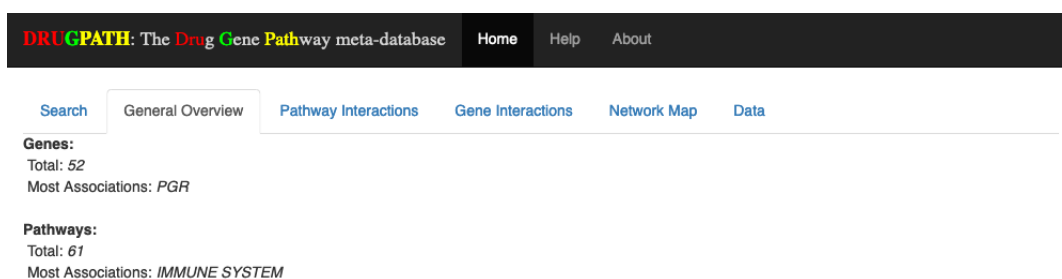
Figure 6. Data available for download.

## 2.2. Enbrel and Mifepristone

Similar to above with citalopram and phenelzine, this section displays the pathway and gene interactions between the drugs enbrel and mifepristone.

### 2.2.1. General Overview

Overall, there were a total of 52 genes and 61 pathways between enbrel and mifepristone, with the progesterone receptor (PGR) gene and the immune system pathway having the most associations. See Figure 7. These results are consistent with the known function of these drugs.



Category	Total	Most Associations
Genes	52	PGR
Pathways	61	IMMUNE SYSTEM

Figure 7. General overview of the interacting genes and pathways between enbrel and mifepristone.

### 2.2.2. Pathway Interactions

Of the 61 pathways shared between enbrel and mifepristone, the first few pathway interactions are shown in Figure 8. The full results (see Supplementary Information S3) show many immune-related pathways such as Th17 cell differentiation, interleukin (IL) 2 (IL-2), IL-4, and IL-13 pathways among many others. As enbrel targets the immune pathways this is a consistent result. Beyond this there are also pathway interactions involved with leptin signaling, gastrin, and ghrelin, which suggests a potential effect on appetite and digestion. Additionally, there are interactions in adipogenesis, adipocytokine and visceral fat deposit pathways suggesting a potential effect on weight gain or loss.

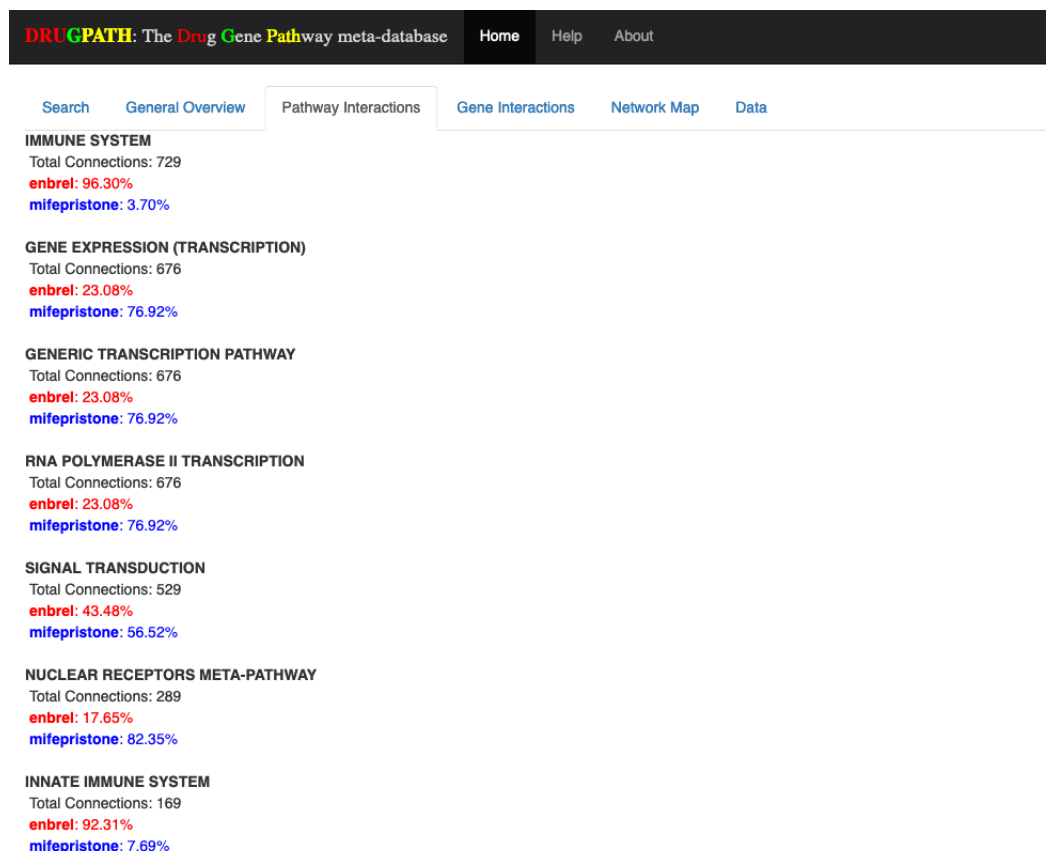


Figure 8. First few pathway interactions between enbrel and mifepristone.

### 2.2.3. Gene Interactions

While all 61 pathways associated with enbrel and mifepristone were affected by both drugs, none of the 52 total genes are shared, or interact, between these drugs. In this case, a message of ‘None available’ appears in the Gene Interactions tab (see Figure 9).

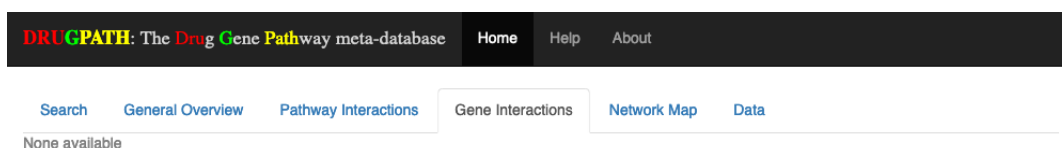


Figure 9. Message that appears when no gene interactions are found.

## 3. Discussion

DRUGPATH is a meta-database that consolidates expert-curated sources in order to map the interactions between drugs, genes, targets, and pathways, of which there are currently  $\approx 60,000$  drug-gene interactions,  $\approx 94,000$  gene-pathway interactions, and over 1.2 million drug-pathway interactions. In addition to all of the sources used in DRUGPATH, there are numerous other freely available biological, pharmacological, genetic, and even interaction databases, however, to our knowledge none of these other databases currently provide drug-gene, drug-pathway, and gene-pathway interactions directly. While DrugBank, PhID, and others provide drug-drug interaction checkers on their websites, none goes so far as to provide both the pathways and genes predicted to be involved with a given interaction.

For instance, Figures 2–6 show the predicted results of the drugs citalopram, a selective serotonin reuptake inhibitor [13], and phenelzine, a monoamine oxidase inhibitor [14]. These two drugs in combination can induce serotonin syndrome [30], but services such as DrugBank and RXList



only state at most whether an interaction occurs and its effects, but not the genes and pathways involved. DRUGPATH on the other hand finds 481 shared pathways and 59 shared genes between citalopram and phenelzine, which includes serotonin related pathways such as “selective serotonin reuptake inhibitor pathway” and “serotonergic synapse” pathway, but also dopaminergic-involved pathways (e.g., “dopaminergic synapse pathway” and “naltrexone action pathway” [31]) as well as “androgen and estrogen biosynthesis and metabolism pathway” and “codeine and morphine metabolism pathway”, which have been verified in literature to be involved with both citalopram and phenelzine’s pharmacological actions [32–36]. These findings lend confidence in DRUGPATH’s ability to identify potential reactive drug combinations.

The combination of enbrel and mifepristone in Figures 7–9 were performed to demonstrate the predictive capabilities of DRUGPATH since there is currently no literature available documenting the interactions between these two drugs. Here, the major hypothesis of GWI pathophysiology involves the combination of stress and exposure to battlefield neurotoxins triggering a neuroinflammatory cascade, leading to altered homeostatic regulation [5–7]. Our computational models have predicted the inhibition of Th1 inflammatory cytokines (e.g., TNF $\alpha$ ) followed by inhibition of the GCR would return the patient back to stable homeostatic regulation [7,10]. Such models indicated the use of enbrel and mifepristone. Enbrel is a TNF $\alpha$  inhibitor FDA approved to treat arthritis and plaque psoriasis, with adverse interactions including headache, infection, rhinitis, and sepsis [11,37–39]. Mifepristone is a high affinity progesterone and glucocorticoid receptor antagonist used for pregnancy abortion, with adverse effects such as diarrhea, fever, cramping, nausea, and vomiting [12,40,41]. DRUGPATH predicted that these two drugs would interact via immune pathways, such as the aptly named “immune pathway” affected by almost all genes connected to enbrel and the ELANE gene from mifepristone. This interaction suggests that these drugs in combination would likely affect immune functioning, which aligns to the proposed treatment strategy from our group of enbrel and mifepristone for GWI [5,7,10]. This is consistent with the expected effect of these drugs, and would normally be monitored when undertaking these drug treatment courses. The identification of overlap in effects on leptin signaling, gastrin, and ghrelin pathways suggests a potential effect on appetite and digestion, which suggests additional monitoring is warranted of digestion in patients undergoing a combined enbrel-mifepristone treatment course. Likewise interactions in adipogenesis, adipocytokine and visceral fat deposit pathways suggests increased monitoring of a patients weight gain or loss during this drug combination.

While analysis of these two drug combinations shows consistency with known drug effects, and shows the potential of DRUGPATH to identify potential adverse drug effects, DRUGPATH’s greatest strength is also its biggest limitation. Namely, it relies heavily on the information from the sources used in its curation. While curation performed by experts is often a favorable trait for biomedical databases, it can lead to errors when dealing with thousands to millions of entries. That being said, sources often make corrections with every new update of their database, which is why we established the requirement that a source must be actively maintained in order to be incorporated into DRUGPATH. Additionally, the architecture of DRUGPATH where each source is self-contained in their own sub-directory was a deliberate choice, since it allows users to not only enable/disable specific sources at will, but it makes applying and checking for updates more convenient and streamlined. Furthermore, DRUGPATH currently does not measure the strength of association between a gene and drug for instance, which is due to the fact that it can only provide data included in each database. While all sources contribute incredibly valuable information, experimental data such as binding energies are unfortunately not provided.

#### 4. Materials and Methods

DRUGPATH was built using MATLAB versions 2017b and 2018a, refs. [42,43] as well as Python version 3.5 [44], and is fully modular in nature. This means that each source is confined to their own sub-directories where the raw input file and all scripts required for pre-processing, data extraction,

and amalgamation to the DRUGPATH dataset are found. We refer to each source's sub-directories as 'plugins' because they can be independently modified without affecting any other source or editing the existing codebase. Not only that, but each source can be easily enabled or disabled using a plaintext configuration file. DRUGPATH was saved in the SQLite database file format, which is lightweight, cross-platform, and very fast for querying millions of entries compared to other formats such as CSV files or Microsoft Excel, making it accessible and easy to use [45].

Generating DRUGPATH consisted of first utilizing PharmGKB [22] to obtain drug-gene interactions, generic and proprietary drug names, and drug identifiers for services such as ChemSpider [46], Uniprot [47], ChEMBL [48–50], and so on. Next, drug-target interactions, additional drug identifiers, and half-life information were obtained from both DrugBank [17,18] as well as T3DB [24,25]. GTP [20] provided even more drug names and identifiers, and HGNC [21] added missing gene ensembl, gene entrez, and gene symbol identifiers for entries. Gene-pathway interactions were mapped using CPDB [15,16], and repoDB [23] provided drug indications. Finally, the database was formatted so that each entry contains only one drug to gene to target to pathway interaction.

One of the major issues when amalgamating many databases is redundancy. Here, targets such as IL-2 can have many different names such as "interleukin-2", "IL-2", "interleukin two", "IL2", "interleukin 2", and so on, while drugs can have a variety of generic as well as proprietary/trade names. To address this, the various drug identifiers used by each source including DrugBank, PubChem substance, PubChem compound, DrugBank, ChEBI, etc. identifiers were all saved in the DRUG\_ID column. This allowed DRUGPATH to map drugs by not only their names, but also their identifiers as well. For instance "mifepristone" (generic name), "RU-486" (generic name), "mifepristona" (Spain), "mifepriстон" (Germany), "mifépristone" (France), "mifegyne" (trade name), etc. were all associated with one another since they share the same DrugBank identifier. It's important to note that all of the unique drug names were saved in the DRUG\_NAME column so that users could easily search a given drug and DRUGPATH would return the appropriate results. In regards to combating redundancy in targets, the gene entrez, gene ensembl, and gene symbol identifiers, found in columns 1–3 in DRUGPATH respectively, were used in the same manner as drug identifiers to match targets that otherwise have non-matching names.

#### 4.1. Fetching and Updating

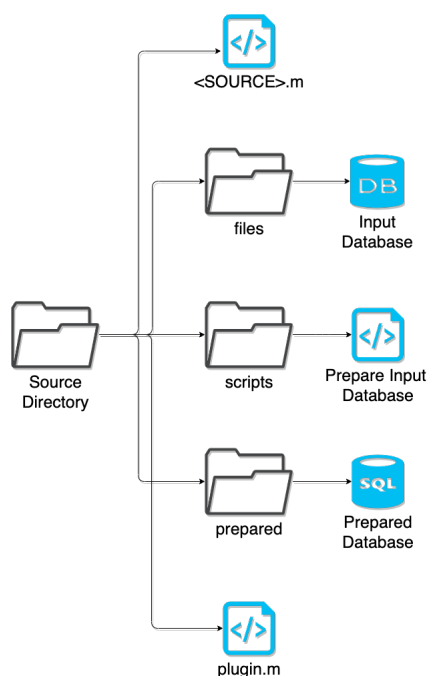
Before a new source can be added to DRUGPATH, it is first vetted against current literature as well as the other sources used in the dataset in an attempt to verify if the data is accurate. Furthermore, the source needs to be actively maintained where it's regularly receiving updates in order to be considered for inclusion, otherwise the likelihood of mapping false correlations from outdated information increases. That all being said, this vetting process is complicated due to a variety of factors, one of which includes misspellings of drug, gene, and/or target names within sources that are otherwise trustworthy. For example, the FDA NDC database contains information on all FDA approved drugs such as proprietary/generic brand names, but there were a few instances of misspelled drug names which made tasks such as mapping drugs to genes more arduous. Additionally, all of the sources that make up DRUGPATH contain thousands of entries, making detection of outdated and/or incorrect entries a very difficult task. Fortunately, these erroneous entries are often quickly corrected when reported, and they only make up a very small share of the entire database.

Updating is completed through a Python script that first downloads the source from its website into a temporary directory, performs a backup of the source's existing database within the plugin directory, and then moves the updated source file into the plugin directory.

#### 4.2. Plugins

Every source is contained in their own plugin directory, which includes the scripts for preparation and amalgamation. Here, the preparation script extracts only relevant information, such as data on humans rather than animals and obtaining universal drug identifiers from ChEMBL or Uniprot instead

of the database's internal identifiers, and normalizes naming conventions (e.g., TNF- $\alpha$  vs TNF $\alpha$  vs TNF $\alpha$ ) to be compatible with the existing DRUGPATH entries. This extracted data is saved to a SQLite database file format, and the amalgamation script queries this SQLite database in order to merge it with the existing DRUGPATH database. See Figure 10 for a graphical overview of how the plugin directories are structured.

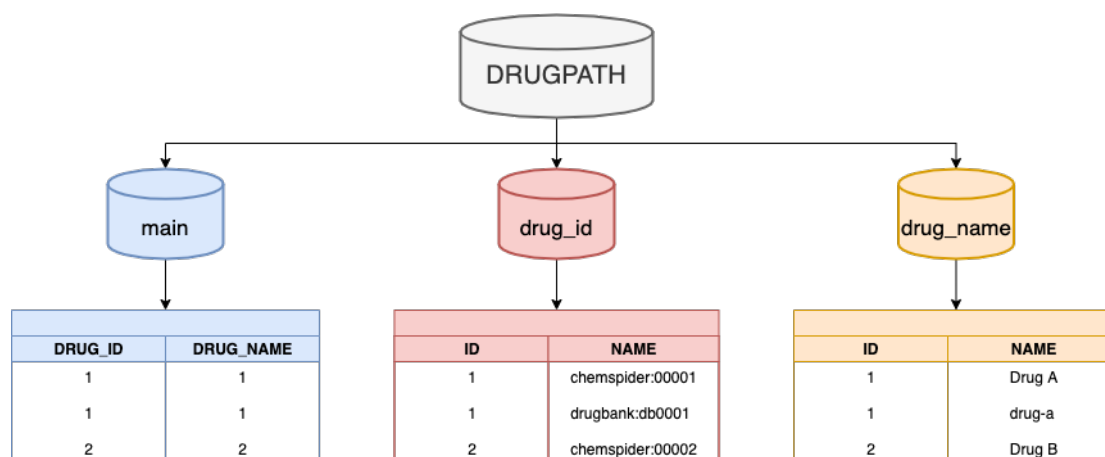


**Figure 10.** Structure of each source's sub-directory. `<SOURCE>.m` refers to the MATLAB script used to amalgamate the **Prepared Database** with the existing DRUGPATH database. `plugin.m` is called by the main MATLAB script that generates the database, and its job is to execute the script that prepares the raw input database as well as execute `<SOURCE>.m`.

#### 4.3. Columns

There are a total of 10 columns within the DRUGPATH database, the details of which are explained below. As an overview, column 1 contains gene entrez identifiers, column 2 is comprised on gene ensembl identifiers, and gene symbols are located in column 3. There are various gene identification systems available, but these three were the most ubiquitous so they were all used. Column 4 contains the targets associated with the genes in previous columns (e.g., gene symbol "CYP3A4" corresponds to target "cytochrome P450 3A4"), and are used for drug-target interactions. Numerous identifiers for each drug are available in column 5, which directly contribute to all of the drug names found in column 6 by allowing different names to be mapped together across distinct sources. Column 7 includes the half-lives for as many drugs as possible, and column 8 contains the FDA status for all drugs found within the appropriate sources. Lastly column 9 is comprised of the pathways associated with each gene, and column 10 contains the FDA indications for the drugs.

The DRUGPATH SQLite database file contains 3 separate tables: main, drug\_id, and drug\_name. The 10 columns described above all exist in the main table, however, once DRUGPATH has been generated, one of the final post-processing steps is to replace both the drug identifiers in columns 5 and the drug names in column 6 with the same integer. This integer serves as an identifier within both the "drug\_id" and "drug\_name" tables, which contain the numerous identifiers as well as generic/trade names for the given drug respectively; see Figure 11. This procedure eliminates repeating the same long string of drug identifiers and names for multiple entries, which not only reduces the total size of the DRUGPATH SQLite file, but also makes querying the database faster and more efficient.



**Figure 11.** Structure of the DRUGPATH SQLite database, where **main**, **drug\_id**, and **drug\_name** are all separate tables within the file. Here, DRUG\_ID “1” corresponds to multiple identifiers within the **drug\_id** table for DRUG\_NAME “1”, which in turn represents the various names of Drug A within the **drug\_name** table. Note that within the **main** table, there are a total of 10 columns, but only the DRUG\_ID and DRUG\_NAME columns are shown in this figure.

#### 4.3.1. Gene Entrez, Ensembl, and Symbol Identifiers

Columns 1–3 of the DRUGPATH database contain gene identifiers for the genes that code for their corresponding targets in column 4. While there are various gene identifier formats that exist, DRUGPATH uses only three of the most prominent: the National Center for Biotechnology Information’s (NCBI) gene entrez, the European Bioinformatics Institute (EMBL-EBI) and the Wellcome Trust Sanger Institute’s gene ensembl, and finally HGNC’s gene symbols. All of the drug-gene relationships found in DRUGPATH are provided by PharmGKB, which builds its database using both manual curation as well as natural language processing techniques to extract this information from published pharmacogenetics and pharmacogenomics studies [22].

#### 4.3.2. Target

The targets in column 4 were provided for drug-target interactions, and are among the most volatile data in DRUGPATH due to the fact that so many targets have numerous names. To illustrate, the ‘Glutamate receptor ionotropic, NMDA 1’ and ‘Glutamate Ionotropic Receptor NMDA Type Subunit 1’ are two different names for the same target, and unlike drug names, there are not many sources available that map these target names together. For this reason target-pathway interactions were actually performed using gene symbols instead, where ‘GRIN1’ corresponds to both of the glutamate receptor names from the previous example.

#### 4.3.3. Drug Identifiers

The issue of various names corresponding to the same target occurs with drugs as well, where there are many generic as well as proprietary/trade names for a drug. This was ameliorated in part by the DrugBank, T3DB, and other sources that provided the names, but the bigger solution was to utilize unique drug identifiers from well established sources such as ChEMBL, ChemSpider, and Uniprot among others which represents all of a drug’s generic and trade names. However, unlike the gene identifiers where a selected few formats are most commonly used, a drug identifier may be available in one database but not in another, so once a new identifier was found within a source it was added to the existing drug identifiers located in column 5, creating an encompassing list of associated identifiers for a given drug. Figure 11 shows that this column is comprised of integers which correspond to the actual drug identifiers found in a separate “drug\_id” table. In essence, a single number can represent

numerous drug identifiers which saves space and speeds up user queries by reducing the amount of text to search through.

#### 4.3.4. Drug Name

Column 6 contains all of the drug names such as generic and trade names for each drug. Similar to the process used for drug identifiers, once a new name for a given drug was found it was automatically added to the list of drug names within this column, which improves the mapping of drug-gene and drug-pathway interactions as more and more databases are incorporated into DRUGPATH. As more associated names for a drug are linked to together, the accuracy and reliability of our network increases. Similar to column 5, the drug names in this column are replaced by a single integer, which serves as an identifier in the “drug\_name” table that corresponds to all of the possible generic/trade names for the given drug.

#### 4.3.5. Half Life

Column 7 contains the half-live information for as many drugs as possible. This data was sourced from DrugBank and T3DB, meaning that only drugs found in these two databases will have associated half-life information.

#### 4.3.6. Status

The FDA status for each drug is found in column 8 and can be one or more of: *approved*, *experimental*, *investigational*, *nutraceutical*, *illicit*, *vet approved* and/or *withdrawn*. Drugs are able to have one or more different statuses because they can be approved to treat one disease and in trial for approval to treat another (status: *approved* and *experimental*), may be approved for both humans and animals, such as Butorphanol (status: *approved* and *vet approved*), and so on. Similar to the half life column above, if the drug was not found in the source providing the information, in this case the FDA NDC database, then the drug will not have a status.

#### 4.3.7. Pathway

CPDB [15,16] provided the gene-pathway interactions found in column 9.

#### 4.3.8. Indication

Finally, column 10 contains each drug’s indication, or in other words, what disease or condition is being treated by this particular drug. DrugBank, T3DB, and repoDB were all utilized to provide this data.

### 4.4. Graphical Interface

Once DRUGPATH has been successfully generated, code written in the R programming language [51] was used to quickly and easily query the database to obtain interactions; see Figure 1. The graphical frontend was created using Shiny [52,53], which passes the user input to the backend R script in order to produce the gene and pathway interactions, as well as a network map that models the drug, gene, pathway interactions. On the homepage, the user can input drug names, gene and drug identifiers, half life information, status, and pathway names in order to search DRUGPATH. Note that any combination of drugs, identifiers, etc. can be queried at once. This graphical interface is located at the web address <http://drugpath.app>.

#### 4.4.1. Searching

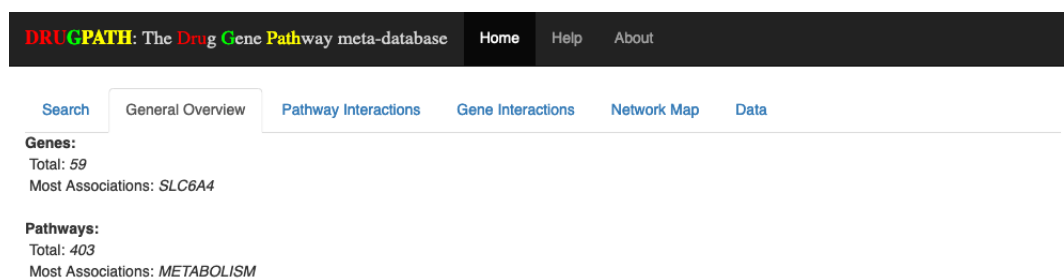
To query the database, input a comma-delimited list of search terms (e.g., “query1, query2, query3, etc.”) into the search bar, and check all categories that the search terms belong to. For example, to find interactions for the drugs “citalopram” and “phenelzine”, the search box would contain these two

terms, separated by a comma, and the checkbox for “Drug Name” would be checked. As search terms are entered, each individual term will be shown at the bottom of the page; see Figure 12. Additionally, the submit button is only shown once a search term has been entered and at least one column has been selected.

**Figure 12.** As the user inputs the search terms, they are shown at the bottom of the page.

#### 4.4.2. General Overview

Once the user has submitted his query, four new tabs are shown. The first of which is the General Overview tab (see Figure 13), which displays a quick synopsis of the genes and pathways associated with the search query. Continuing with the citalopram and phenelzine example, it is shown that there are a total of 59 genes and 471 pathways associated between these two drugs. The “Most Associations” metric refers to the gene or pathway with the most amount of connections, defined as the total number of unique drug-gene-target entries in DRUGPATH for the genes, and the total number of unique drug-gene-target-pathway entries for the pathways. Essentially, this is a measurement of how prominent a given gene or pathway is within an interaction network. In this case the “SLC6A4” gene and the “metabolism” pathway had the most connections between citalopram and phenelzine in this particular network.



**Figure 13.** Quick overview of the total number of genes and pathways found, as well as the gene and pathway that had the most connections.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/21/9/3171/s1>.

**Author Contributions:** R.J. wrote all code used to generate the database and authored the manuscript with T.J.A.C., who supervised the completion of the project, completely funded it, as well as reviewed and made edits and corrections to the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the United States Department of Defense Congressionally Directed Medical Research Program (CDMRP) awards (<http://cdmrp.army.mil/>), grant number [W81XWH-15-1-0582] (Craddock/Broderick-PIs), grant number [W81XWH-16-1-0632] (Craddock-PI), grant number [W81XWH-16-1-0552] (Craddock-PI) and grant number [W81XWH-13-2-0085] (Morris-PI, Broderick/Klimas-Co-PIs).

**Conflicts of Interest:** The authors declare no conflict of interest.



## Abbreviations

The following abbreviations are used in this manuscript:

CFS	Chronic fatigue syndrome
CPDB	ConsensusPathDB
CSBG	Clinical Systems Biology Group
CSV	Comma separated value
CYP3A4	Cytochrome P450 3A4
DRUGPATH	Drug-gene-pathway
EMBL-EBI	European Bioinformatics Institute
FDA NDC	U.S. Food and Drug Administration's National Drug Code
FR	Fruchterman and Reingold
GCR	Glucocorticoid receptor
GTP	Guide to Pharmacology
GWI	Gulf War Illness
HGNC	HUGO Gene Nomenclature Committee
HIV	Human immunodeficiency virus
IL	Interleukin
INIM	Institute for Neuro-Immune Medicine
NCBI	National Center for Biotechnology Information
PharmGKB	Pharmacogenomics Knowledgebase
PGR	Progesterone receptor
T3DB	Toxin and Toxin-Target Database
TNF $\alpha$	Tumor necrosis factor alpha

## References

1. Pretorius, E.; Klinker, H.; Rosenkranz, B. The role of therapeutic drug monitoring in the management of patients with human immunodeficiency virus infection. *Ther. Drug Monit.* **2011**, *33*, 265–274. [[CrossRef](#)] [[PubMed](#)]
2. Ramana, L.N.; Anand, A.R.; Sethuraman, S.; Krishnan, U.M. Targeting strategies for delivery of anti-HIV drugs. *J. Control. Release* **2014**, *192*, 271–283. [[CrossRef](#)] [[PubMed](#)]
3. Piscitelli, S.C.; Gallicano, K.D. Interactions among drugs for HIV and opportunistic infections. *N. Engl. J. Med.* **2001**, *344*, 984–996. [[CrossRef](#)] [[PubMed](#)]
4. Kell, D.B.; Dobson, P.D.; Bilsland, E.; Oliver, S.G. The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: What we (need to) know and how we can do so. *Drug Discov. Today* **2013**, *18*, 218–239. [[CrossRef](#)]
5. Jaundoo, R.; Bohmann, J.; Gutierrez, G.E.; Klimas, N.; Broderick, G.; Craddock, T.J. Using a consensus docking approach to predict adverse drug reactions in combination drug therapies for gulf war illness. *Int. J. Mol. Sci.* **2018**, *19*, 3355. [[CrossRef](#)]
6. Jeffrey, M.G.; Kregel, M.; Kibler, J.; Zundel, C.; Klimas, N.G.; Sullivan, K.; Craddock, T.J. Neuropsychological Findings in Gulf War Illness: A Review. *Front. Psychol.* **2019**, *10*, 2088. [[CrossRef](#)]
7. Craddock, T.J.; Del Rosario, R.R.; Rice, M.; Zysman, J.P.; Fletcher, M.A.; Klimas, N.G.; Broderick, G. Achieving Remission in Gulf War Illness: A Simulation-Based Approach to Treatment Design. *PLoS ONE* **2015**, *10*, e0132774. [[CrossRef](#)]
8. Craddock, T.J.; Harvey, J.M.; Nathanson, L.; Barnes, Z.M.; Klimas, N.G.; Fletcher, M.A.; Broderick, G. Using gene expression signatures to identify novel treatment strategies in gulf war illness. *BMC Med. Genom.* **2015**, *8*, 36. [[CrossRef](#)]
9. Craddock, T.J.; Fritsch, P.; Rice, M.A., Jr.; Del Rosario, R.M.; Miller, D.B.; Fletcher, M.A.; Klimas, N.G.; Broderick, G. A role for homeostatic drive in the perpetuation of complex chronic illness: Gulf War Illness and chronic fatigue syndrome. *PLoS ONE* **2014**, *9*, e100355.
10. Craddock, T.J.; Michalovicz, L.T.; Kelly, K.A.; Rice, M.A., Jr.; Miller, D.B.; Klimas, N.G.; Morris, M.; O'Callaghan, J.P.; Broderick, G. A Logic Model of Neuronal-Glial Interaction Suggests Altered Homeostatic Regulation in the Perpetuation of Neuroinflammation. *Front. Cell. Neurosci.* **2018**, *12*, 336. [[CrossRef](#)]

11. Culy, C.R.; Keating, G.M. Etanercept. *Drugs* **2002**, *62*, 2493–2537. [[CrossRef](#)] [[PubMed](#)]
12. Cadepond, F.; Ulmann, P.A.; Baulieu, P.E.E. RU486 (mifepristone): Mechanisms of action and clinical uses. *Annu. Rev. Med.* **1997**, *48*, 129–156. [[CrossRef](#)] [[PubMed](#)]
13. Hyttel, J. Citalopram? pharmacological profile of a specific serotonin uptake inhibitor with antidepressant activity. *Prog. Neuro-Psychopharm. Biol. Psychiatry* **1982**, *6*, 277–295. [[CrossRef](#)]
14. Robinson, D.; Nies, A.; Ravaris, C.; Ives, J.; Bartlett, D. Clinical pharmacology of phenelzine. *Arch. Gen. Psychiatry* **1978**, *35*, 629–635. [[CrossRef](#)] [[PubMed](#)]
15. Kamburov, A.; Wierling, C.; Lehrach, H.; Herwig, R. ConsensusPathDB—A database for integrating human functional interaction networks. *Nucleic Acids Res.* **2008**, *37*, D623–D628. [[CrossRef](#)]
16. Kamburov, A.; Pentchev, K.; Galicka, H.; Wierling, C.; Lehrach, H.; Herwig, R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* **2010**, *39*, D712–D717. [[CrossRef](#)]
17. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [[CrossRef](#)]
18. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2017**, *46*, D1074–D1082. [[CrossRef](#)]
19. United States Food and Drug Administration. National Drug Code Directory. Available online: <https://www.accessdata.fda.gov/cder/ndc/text.zip> (accessed on 23 May 2018).
20. Armstrong, J.F.; Faccenda, E.; Harding, S.D.; Pawson, A.J.; Southan, C.; Sharman, J.L.; Campo, B.; Cavanagh, D.R.; Alexander, S.P.; Davenport, A.P.; et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: Extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* **2020**, *48*, D1006–D1021. [[CrossRef](#)]
21. Gray, K.A.; Yates, B.; Seal, R.L.; Wright, M.W.; Bruford, E.A. Genenames. org: The HGNC resources in 2015. *Nucleic Acids Res.* **2014**, *43*, D1079–D1085. [[CrossRef](#)]
22. Whirl-Carrillo, M.; McDonagh, E.M.; Hebert, J.; Gong, L.; Sangkuhl, K.; Thorn, C.; Altman, R.B.; Klein, T.E. Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **2012**, *92*, 414–417. [[CrossRef](#)] [[PubMed](#)]
23. Brown, A.S.; Patel, C.J. A standard database for drug repositioning. *Sci. Data* **2017**, *4*, 1–7.
24. Lim, E.; Pon, A.; Djoumbou, Y.; Knox, C.; Shrivastava, S.; Guo, A.C.; Neveu, V.; Wishart, D.S. T3DB: A comprehensively annotated database of common toxins and their targets. *Nucleic Acids Res.* **2009**, *38*, D781–D786. [[CrossRef](#)] [[PubMed](#)]
25. Wishart, D.; Arndt, D.; Pon, A.; Sajed, T.; Guo, A.C.; Djoumbou, Y.; Knox, C.; Wilson, M.; Liang, Y.; Grant, J.; et al. T3DB: The toxic exposome database. *Nucleic Acids Res.* **2014**, *43*, D928–D934. [[CrossRef](#)]
26. Almende, B.V.; Thieurmel, B.; Robert, T. visNetwork: Network Visualization Using ‘vis.js’ Library. R Package Version 2.0.3. 2018. Available online: <https://cran.r-project.org/web/packages/visNetwork/index.html> (accessed on 1 January 2020).
27. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJ. Complex Syst.* **2006**, *1695*, 1–9.
28. Fruchterman, T.M.; Reingold, E.M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **1991**, *21*, 1129–1164. [[CrossRef](#)]
29. Gibson, H.; Faith, J.; Vickers, P. A survey of two-dimensional graph layout techniques for information visualisation. *Inf. Vis.* **2013**, *12*, 324–357. [[CrossRef](#)]
30. Schellander, R.; Donnerer, J. Antidepressants: Clinically relevant drug interactions to be considered. *Pharmacology* **2010**, *86*, 203–215. [[CrossRef](#)]
31. Bostwick, J.M.; Bucci, J.A. Internet sex addiction treated with naltrexone. In *Mayo Clinic Proceedings*; Elsevier: Amsterdam, The Netherlands, 2008; Volume 83, pp. 226–230.
32. Goel, N.; Plyler, K.S.; Daniels, D.; Bale, T.L. Androgenic influence on serotonergic activation of the HPA stress axis. *Endocrinology* **2010**, *152*, 2001–2010. [[CrossRef](#)]
33. Larsen, J.J.; Arnt, J. Spinal 5-HT or NA Uptake Inhibition Potentiates Supraspinal Morphine Antinociception in Rats. *Basic Clin. Pharmacol. Toxicol.* **1984**, *54*, 72–75. [[CrossRef](#)]
34. Singh, V.P.; Patil, C.S.; Jain, N.K.; Singh, A.; Kulkarni, S.K. Paradoxical effects of opioid antagonist naloxone on SSRI-induced analgesia and tolerance in mice. *Pharmacology* **2003**, *69*, 115–122. [[CrossRef](#)] [[PubMed](#)]

35. Smith, H.S. Opioid metabolism. In *Mayo Clinic Proceedings*; Elsevier: New York, NY, USA, 2009; Volume 84, pp. 613–624.
36. Sternbach, H. Danger of MAOI therapy after fluoxetine withdrawal. *Lancet* **1988**, *332*, 850–851.
37. United States Food and Drug Administration. ENBREL<sup>®</sup> (Etanercept). Available online: [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2008/enbrel\\_pi.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2008/enbrel_pi.pdf) (accessed on 1 January 2020).
38. Goffe, B.; Cather, J.C. Etanercept: An overview. *J. Am. Acad. Dermatol.* **2003**, *49*, 105–111. [[CrossRef](#)] [[PubMed](#)]
39. Shannon, E.; Daffy, J.; Jones, H.; Paulson, A.; Vicik, S.M. Etanercept (Enbrel<sup>®</sup>) alternative storage at ambient temperature. *Clin. Pharmacol. Adv. Appl.* **2017**, *9*, 87.
40. Honkanen, H.; Piaggio, G.; Von Hertzen, H.; Bártfai, G.; Erdenetungalag, R.; Gemzell-Danielsson, K.; Gopalan, S.; Horga, M.; Jerve, F.; Mittal, S.; et al. WHO multinational study of three misoprostol regimens after mifepristone for early medical abortion. II: Side effects and women's perceptions. *BJOG Int. J. Obstet. Gynaecol.* **2004**, *111*, 715–725.
41. Sarkar, N. Mifepristone: Bioavailability, pharmacokinetics and use-effectiveness. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2002**, *101*, 113–120. [[CrossRef](#)]
42. MATLAB. *Version 9.3.0.713579 (R2017b)*; The MathWorks, Inc.: Natick, MA, USA, 2017.
43. MATLAB. *Version 9.4.0.813654 (R2018a)*; The MathWorks, Inc.: Natick, MA, USA, 2018.
44. Van Rossum, G. *Python Tutorial*; Technical Report CS-R9526; Centrum voor Wiskunde en Informatica (CWI): Amsterdam, The Netherlands, 1995.
45. Junyan, L.; Shiguo, X.; Yijie, L. Application research of embedded database SQLite. In Proceedings of the 2009 International Forum on Information Technology and Applications, Chengdu, China, 15–17 May 2009; Volume 2, pp. 539–543.
46. Pence, H.E.; Williams, A. ChemSpider: An online chemical information resource. *J. Chem. Educ.* **2010**, *87*, 1123–1124. [[CrossRef](#)]
47. Consortium, U. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2016**, *45*, D158–D169.
48. Gaulton, A.; Kale, N.; van Westen, G.J.; Bellis, L.J.; Bento, A.P.; Davies, M.; Hersey, A.; Papadatos, G.; Forster, M.; Wege, P.; et al. The ChEMBL bioactivity database: An update. *Sci. Data* **2013**, *2*, 150032. [[CrossRef](#)]
49. Davies, M.; Nowotka, M.; Papadatos, G.; Atkinson, F.; van Westen, G.J.; Dedman, N.; Ochoa, R.; Overington, J.P. MyChEMBL: A virtual platform for distributing cheminformatics tools and Open Data. *Challenges* **2014**, *5*, 334–337. [[CrossRef](#)]
50. Jupp, S.; Malone, J.; Bolleman, J.; Brandizi, M.; Davies, M.; Garcia, L.; Gaulton, A.; Gehant, S.; Laibe, C.; Redaschi, N.; et al. The EBI RDF platform: Linked open data for the life sciences. *Bioinformatics* **2014**, *30*, 1338–1339. [[CrossRef](#)] [[PubMed](#)]
51. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
52. Chang, W.; Cheng, J.; Allaire, J.; Xie, Y.; McPherson, J. Shiny: Web Application Framework for R. R Package Version 1.1.0. 2018. Available online: <https://cran.r-project.org/web/packages/shiny/index.html> (accessed on 1 January 2020).
53. Sali, A. Shinycssloaders: Add CSS Loading Animations to 'Shiny' Outputs. R Package Version 0.2.0. 2017. Available online: <https://cran.r-project.org/web/packages/shinycssloaders/index.html> (accessed on 1 January 2020).

