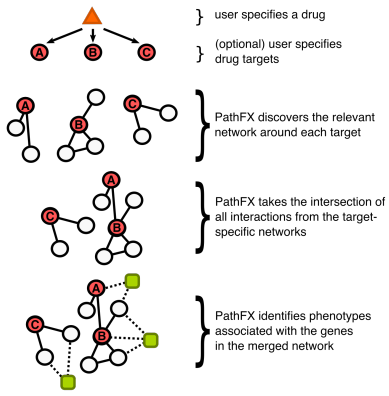
Supplementary Figures and Tables

Summary of PathFX algorithm and data sources

The PathFX algorithm was designed to look at the pathways-level effects (pathway "FX") of a drug intervention. Under the hood, PathFX is an interaction-network tool that searches for the most relevant protein-protein interactions around a drug's target(s), and then determines for which phenotypes the network is enriched relative to the entire interaction network (Supplementary Figure 1). We merged data from iRefWeb v4.1 (Turinsky *et al.*, 2014), PharmGKB(Whirl-Carrillo *et al.*, 2012), and a novel set of drug-protein binding information (published in (Wilson *et al.*, 2018) to create an interactome containing protein-protein, gene-protein, gene-gene, and drug-gene interactions. We scored interactions based on the amount and quality of evidence supporting the interaction. We wrote a custom depth-first network search algorithm with fast-tracking to identify the relevant interactions around a drug target(s) of interest and empirically derived a threshold for stopping the search (Wilson *et al.*, 2018). This threshold was derived to prevent over-representation from high degree network genes/proteins that result from study bias. To discover phenotypes associated with networks around a drug target(s), we merged gene-phenotype data from ClinVar(Landrum *et al.*, 2014), OMIM(Amberger *et al.*, 2009), PheGenI(Ramos *et al.*, 2014), DisGeNet(Piñero *et al.*, 2015; 2017), and eQTL data from the GWAS catalogue(Welter *et al.*, 2013). We assessed the association of a phenotype to a set of network genes/proteins using a Fisher's exact test. To control for annotation bias in the number of genes associated with phenotypes, we removed associations based on an empirically-derived p-value threshold. This threshold is derived by determining the expected association significance using networks created with random targets.

The algorithm and the web application provide tabular results of associated phenotypes ranked by significance, and a summary figure and table of phenotype clusters based on semantic similarity between phenotypes. As previously reported in (Wilson *et al.*, 2018), PathFX discovered associations between a drug's target(s) and the intend-to-treat disease for 558 of 1364 (40.9%) drug-disease pairs; this is our best estimate of the algorithm's sensitivity. The computational complexity of the PathFX algorithm (without phenotype clustering) is 0(n) and the expected run time is less than one minute.



Supplementary Figure 1. An abbreviated schematic of PathFX. The user inputs a drug and an optional list of gene names of drug-binding targets. PathFX identifies the most relevant interacting partners with the drug target(s) and then takes the intersection of this set of interactions to create a merged pathway for the drug. PathFX identifies the phenotypes associated with network proteins/genes.

Adaptations of PathFX for PathFXweb

PathFXweb uses the PathFX algorithm as published in (Wilson *et al.*, 2018). Because downloading and installing the UMLS Metathesaurus can be cumbersome, we have included a version of phenotype clustering with PathFXweb. This feature takes the top 50 phenotypes ranked from PathFX and clusters them using semantic similarly. Our wrapper code extends the capabilities of umls-interface.pl and umls-similarity.pl (McInnes *et al.*, 2009). The computational complexity of this feature is $O(n^2)$ and takes several (~12-24) hours to run with a set of 50 phenotypes on our server. Our preferred browser is Google Chrome, however, the application has been tested on Safari and Firefox as well.

Data Update and Future Releases

PathFXweb will be updated annually. With each data update, we will update the interactome and the gene-phenotype associations. We will empirically derive the interaction and p-value thresholds (described above) with each data update. The version number is documented in the "pathfx log" file and users will have access to previous versions if they wish to recreate previous analyses.

Supplementary Figure 2. User interface for registering and running PathFXweb. (**A**) The home page prompts the user to login (indicated with red arrows). (**B**) After registering, the user can access the analysis page. Analysis parameters are entered into a form (highlighted with red rectangle) and can track job status using the "PathFX Jobs" tab (red arrow, upper right), or underneath the USER_NAME (red arrow, middle right). Example analyses are included as blue radio buttons below the form. (**C**) After running an analysis, the "PathFX Jobs" page lists a table of results and will indicate whether a job is running or ready for download or viewing. Clicking on the "download" (red box, far right) will access a copy of the zipped results that were also emailed to the user. Clicking on the "visualize" (red box, far left) will open the network visualization where the user sees and modifies the network and exports results. Note: results will be deleted after 7 days to save server storage.

Supplementary Table 1. List of results files and descriptions

File Name/Extension	Description
One or more files with the ending 'neighborhood.txt'	These are the protein neighborhoods for the individual drug targets.
One or more files with the ending 'specific_neighborhood.txt'	These are the protein neighborhoods after controlling for study bias in the interaction network. For further information, please see the website and (Wilson *et al.*, 2018).
The file ending with 'merged_neighborhood.txt'	This is the full drug network after controlling for study bias and considering interactions from all drug targets. The edge scores between proteins and gene variants reflect the amount and quality of evidence supporting the interaction. This edge score is explained in (Wilson *et al.*, 2018).
The file ending with 'merged_neighborhood__assoc_table.txt'	This is a table of phenotypes associated with the network. The top 50 phenotypes from this table are used for phenotype clustering if the user enables this feature.
The file ending with 'merged_neighborhood__assoc_database_sources.txt'	This file lists the database sources (e.g. ClinVar, OMIM) of gene-to-phenotype associations used in PathFX.
The *.pkl files, including: lin_pandas_matrix.pkl, disease_clusters_lin_1.7.pkl, merged_neighborhood__cui_list_.pkl	These are intermediate files from the phenotype clustering phase of the algorithm.
The cluster_membership_*.txt files	These are tables of phenotype clusters and the phenotypes assigned to each cluster. This file is only generated if phenotype clustering is enabled.
Two .png figures showing the results of the phenotype clustering, one file ending with 'labeledClusters_dendogram_full_1.7.png', and one file ending with 'unlabeled_dendogram_full_1.7.png'	In the former, the dendrogram labels show the top words associated with a particular cluster and in the later, the dendrogram labels show the number of phenotypes collapsed into a cluster or in the case of single-phenotypes clusters, the label shows the individual concept unique identifier (CUI). The "top words" labeling was chosen as a short-hand to represent each cluster without cluttering the image. We recommend that users look at the full disease list in the cluster_membership_*.txt file to assess which phenotypes are associated with the cluster.
A file ending with 'merged_neighborhood__withDrugTargsAndPhens.txt'	This is the complete protein network with phenotype interactions that is used by the server for creating the network visualization and can also be used on the desktop version of Cytoscape. The edge scores in this file are set to 1.0 for all drug-to-protein interactions and gene-to-phenotype interactions. The edge scores between proteins and gene variants reflect the amount and quality of evidence supporting the interaction. This edge score is explained in (Wilson and Altman, 2018).
A file ending with 'network_nodeType.txt'	This file specifies the entity type in the network file. It is also used by the server for color-coding the network visualization and can also be used on the desktop version of Cytoscape.
A .json file.	This stores the configuration information for Cytoscape network visualization. This file is included when the user chooses to visualize the network after their job has completed.
A README file	This contains the user name, analysis parameters, analysis date, and PathFX version for data provenance.

Example Analysis with Metformin

Here we prototyped an analysis with the antidiabetic drug, metformin. After the user navigates to the "Run PathFX" page (Supplementary Figure 2B), they can enter the following parameters or select the "Metformin" radio button (Supplementary Figure 2B) to automate the example query:

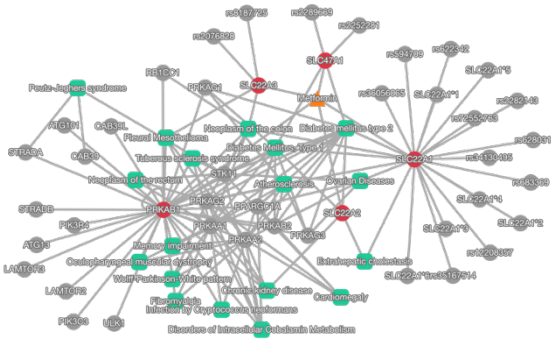
Name: PathFX Example: Metformin

Analysis Name: Exploring phenotypes associated with Metformin

Drug Name: Metformin

Drug Targets: [blank]

The user clicks the "PathFX Jobs" page to download a zipped file of results or visualize the network (Supplementary Figure 2C). The network visualization defaults to a color scheme that highlights drug, drug-binding proteins, intermediate proteins, and phenotypes associated with network genes (Supplementary Figure 3). Users drag-and-drop nodes as they see fit; there is a "dandelion drag" feature where single-edge connections to a central hub node (resembling a dandelion seed head) move together with the hub node, simplifying reconfiguration. After configuring the network, the user exports the image in png format.



Supplementary Figure 3. Network visualization of metformin. The user queries the network image associated with metformin. The network includes the drug, metformin (orange triangle), metformin's protein binding partners (red circles), intermediate pathway proteins (grey circles), and phenotypes (green squares). The user toggles network entities to change their position before exporting the image to png format.

Example analysis with Metformin and Atorvastatin

Here we prototyped an example of a metformin and atorvastatin drug combination. After the user navigates to the "Run PathFX" page (Supplementary Figure 2B), they can enter the following parameters or select the "Drug Combo" radio button (Supplementary Figure 2B) to automate the example query:

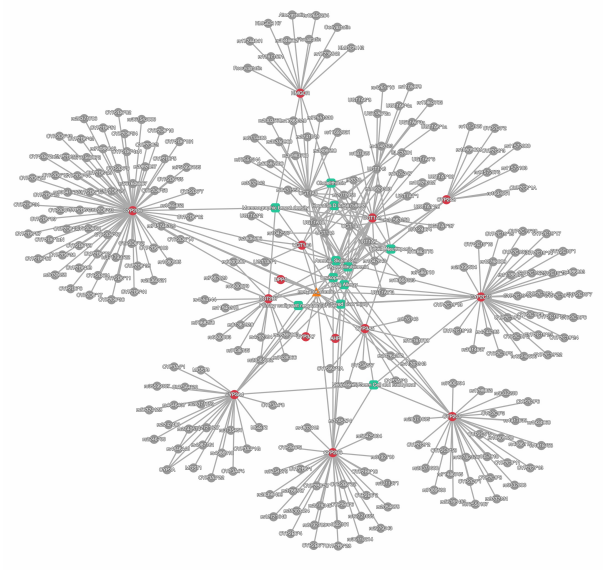
Name: PathFX Example: Metformin & Atorvastatin

Analysis Name: Exploring phenotypes associated with Metformin and Atorvastatin

Drug Name: met_ator_combo

Drug Targets: HMGCR,DPP4,AHR,CYP3A4,CYP3A5,CYP3A7,CYP2C8,CYP2D6,CYP2C9,CYP2C19,CYP2B6,UGT1A1,UGT1A3,UGT2B7

In this example, the user has queried the set of phenotypes associated with the union of drug targets from metformin, an anti-diabetic drug, and atorvastatin, a cholesterol-lowering drug. This query creates a network where the drug combination is represented as a single agent (orange triangle, Supplementary Figure 4), that is connected to the 14 drug targets (red circles, Supplementary Figure 4) associated with this combination. PathFXweb recovers 11 phenotypes associated with this combination (Supplementary Table 2).



Supplementary Figure 4. Network visualization of metformin and atorvastatin combination. The network includes the drug combination, "met_ator_combo" (orange triangle), the drug's protein binding partners (red circles), intermediate pathway proteins (grey circles), and phenotypes (green squares). The user toggles network entities to change their position before exporting the image to png format.

Supplementary Table 2. Association table result file from metformin and atorvastatin drug combination. PathFXweb generates an association table where phenotypes are ranked by their multiple-hypothesis-corrected p-value ("Benjamini-Hochberg"). The table details how many genes in the drug neighborhood are associated with the phenotype ("assoc in neigh") and how many genes are associated with the phenotype in the entire interactome ("assoc in intom"). The last column lists which genes from the drug neighborhood are associated with the phenotype ("genes").

Rank	phenotype	cui	assoc in neigh	assoc in intom	probability	Benjamini-Hochberg	genes
36	Drug Allergy	C0013182	11	79	9.19E-11	3.23E-05	CYP2C19,CYP2C8,CYP2C9,CYP2D6,CYP3A4,CYP3A5,UGT1A1,UGT1A7,UGT1A8,UGT1A9,UGT2B7
39	Hepatitis D Infection	C0011226	8	55	2.06E-09	3.50E-05	UGT1A,UGT1A1,UGT1A10,UGT1A3,UGT1A6,UGT1A7,UGT1A8,UGT1A9
42	Hyperbilirubinemia	C0020433	9	99	3.02E-08	3.77E-05	CYP2B6,UGT1A,UGT1A1,UGT1A10,UGT1A3,UGT1A6,UGT1A7,UGT1A8,UGT1A9
43	Febrile Neutropenia	C0746883	5	27	1.52E-07	3.86E-05	CYP3A5,UGT1A,UGT1A1,UGT1A6,UGT1A7
45	Anemia, Sickle Cell	C0002895	12	271	1.71E-06	4.04E-05	CYP2C19,CYP2C9,CYP2D6,UGT1A,UGT1A1,UGT1A10,UGT1A3,UGT1A6,UGT1A7,UGT1A8,UGT1A9,UGT2B7
48	Primary malignant neoplasm of liver	C0024620	4	25	3.04E-06	4.31E-05	CYP2C19,CYP2D6,CYP3A4,CYP3A5
50	Mammographic breast density	C1268717	6	70	4.00E-06	4.49E-05	CYP2D6,HMGCR,UGT1A,UGT1A1,UGT1A3,UGT2B7
51	Cholelithiasis	C0008350	9	171	4.87E-06	4.58E-05	HMGCR,UGT1A,UGT1A1,UGT1A10,UGT1A3,UGT1A6,UGT1A7,UGT1A8,UGT1A9
55	Drug-Induced Liver Injury	C0860207	9	198	1.76E-05	4.94E-05	AHR,CYP2B6,CYP2C19,CYP2C9,CYP2D6,CYP3A5,UGT1A1,UGT1A3,UGT1A9
56	Leukopenia	C0023530	11	301	2.75E-05	5.03E-05	CYP2B6,CYP2C8,CYP3A4,CYP3A5,UGT1A,UGT1A1,UGT1A6,UGT1A7,UGT1A8,UGT1A9,UGT2B7
57	Neoplasms, Germ Cell and Embryonal	C0027658	5	64	2.89E-05	5.12E-05	CYP2B6,CYP2C19,CYP2C9,CYP3A4,CYP3A5

Example analysis with experimental, anti-EGFR drug

Here we prototyped an example of an experimental drug that binds EGFR. After the user navigates to the "Run PathFX" page (Supplementary Figure 2B), they can enter the following parameters or select the "EGFR Drug" radio button (Supplementary Figure 2B) to automate the example query:

Name: PathFX Example: EGFR

Analysis Name: Exploring phenotypes associated with EGFR

Drug Name: EGFR_drug

Drug Targets: EGFR

In this example, the user has queried the set of phenotypes associated with the druggable target, EGFR. This query discovers a network of genes/proteins associated with EGFR (data not shown) and a set of 1331 phenotypes associated with this network (abbreviated results shown in Supplementary Table 3). For all queries, PathFX web also reports the databases from which the gene-to-phenotype associations originate (abbreviated results in Supplementary Table 4).

Supplementary Table 3. Abbreviated association table result file from EGFR_drug. PathFXweb generates an association table where phenotypes are ranked by their multiple-hypothesis-corrected p-value ("Benjamini-Hochberg"). The table details how many genes in the drug neighborhood are associated with the phenotype ("assoc in neigh") and how many genes are associated with the phenotype in the entire interactome ("assoc in intom"). The last column lists which genes from the drug neighborhood are associated with the phenotype ("genes"). A subset of the 1331 associated phenotypes are shown in this table.

Rank	phenotype	cui	assoc in neigh	assoc in intom	probability	Benjamini-Hochberg	genes
1	Alzheimer Disease	C0002395	61	1699	0	1.82E-07	ADRB2,AKAP9,AP2A2,APOB,ARRB2,ARTN,BIN1,CAMK2A,CAMK2G,CHRM2,CSF2,DAB2,DLG4,DNM2,EGFR,ERBB4,FGF1,FGF2,FGF23,FGFR1,FGFR2,FGFR3,FRS3,FYN,GDNF,GFRA2,GRB2,GRIN1,GRIN2A,GRIN2B,HSP90AA1,IGF2R,IL2,IL2RB,IL3,IRS1,IRS2,ITSN1,JAK2,KL,LRP2,LRRK2,M6PR,NCAM1,NEFL,NRG1,PDGFRA,PDGFRB,PICALM,PIK3CA,PIK3CB,PIK3R1,PTPRA,RANBP9,SH3GL2,SLC18A3,SYNJ1,SYT1,TF,UBB,VAMP2
[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]
9	Inflammation	C0021368	30	647	0	1.64E-06	ADRB2,ANGPT1,APOB,BIN1,CAMK2A,CHRM2,CSF2,EGF,EGFR,FGF1,FGF2,FGF7,HGF,HSP90AA1,IL2,IL2RA,IL2RB,IL5,IL5RA,IRS1,KL,KRAS,NCAM1,NRG1,PDGFA,PDGFB,PDGFRB,RET,TF,TFRC
[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]
28	Neoplasm of the central nervous system	C0085136	14	145	0	5.11E-06	EGFR,ERBB2,FGFR1,GDNF,HGF,KRAS,NRAS,OCRL,PDGFB,PDGFRB,PIK3CA,PTPN11,RET,SOS1
30	Acute leukemia	C0085669	27	525	0	5.47E-06	ANGPT1,CALM3,CBL,CSF2,EPS15,FGFR1,HRAS,HSP90AA1,IL2,IL2RB,IL3,IL3RA,IL5,JAK1,JAK2,JAK3,KIT,KRAS,NCAM1,NRAS,PDGFRA,PICALM,PIK3CA,PIK3CB,PTPN11,PTPRA,TFRC
[…]	[…]	[…]	[…]	[…]	[…]	[…]	[…]

Supplementary Table 4. Abbreviated association source file for EGFR_drug. For all genes associated with a network phenotype, PathFX and PathFXweb report the database source for the association. This table includes an abbreviated view of the gene-to-phenotype relationships associated with the EGFR_drug network and highlights that some associations originate from single sources (e.g. ADRB2's association to "Alzheimer Disease" originates from DisGeNet) and some associations originate from multiple sources (e.g. DIAPHI's association to "Seizures, Febrile" originates from ClinVar, DisGeNet, and HumPhenOnt).

Gene	CUI	Phenotype	Source Databases
ADRB2	C0002395	Alzheimer Disease	DisGeNet
[…]	[…]	[…]	[…]
BIN1	C0002395	Alzheimer Disease	DisGeNet,PheGenI
[…]	[…]	[…]	[…]
PICALM	C0002395	Alzheimer Disease	DisGeNet,PheGenI
[…]	[…]	[…]	[…]
DIAPH1	C0036572	Seizures, Febrile	ClinVar,DisGeNet,HumPhenOnt
[…]	[…]	[…]	[…]

Supplementary References

Amberger,J. et al. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM(R)). Nucleic Acids Research, 37, D793–D796.
Landrum,M.J. et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Research, 42, D980–5.
McInnes,B.T. et al. (2009) UMLS-Interface and UMLS-Similarity : open source software for measuring paths and semantic similarity. AMIA Annu Symp Proc, 2009, 431–435.
Piñero,J. et al. (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Research, 45, D833–D839.
Piñero,J. et al. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford), 2015, bav028–bav028.
Ramos,E.M. et al. (2014) Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur. J. Hum. Genet., 22, 144–147.
Turinsky,A.L. et al. (2014) Navigating the global protein-protein interaction landscape using iRefWeb. Methods Mol. Biol., 1091, 315–331.
Welter,D. et al. (2013) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research, 42, D1001–D1006.
Whirl-Carrillo,M. et al. (2012) Pharmacogenomics Knowledge for Personalized Medicine. Clinical pharmacology and therapeutics, 92, 414–417.
Wilson,J.L. et al. (2018) PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development. PLoS Comput Biol, 14, e1006614–27.