Standardizing Knowledge of Drug Effects: An Application of PheKnowLator for Drug Safety

Tiffany J. Callahan, Patrick B. Ryan, George Hripcsak

Background

Adverse events are a significant public health burden resulting in ~1.3 million emergency room visits and more than $3.5 billion dollars in annual medical costs.[1,2] The majority of adverse events are preventable and occur from known causes.[3] Ideally, adverse events would be identified during drug development or post-marketing surveillance, but even with strict regulations, extensive experimentation, and robust reporting, these methods are unable to account for every source of therapeutic and biological variance.[4,5] Despite decades of research, the majority of adverse events and side effects are not known or discovered until after they are observed.[6]

Systems pharmacology aims to describe the effects of a drug at the molecular level,[3,7–11] by integrating data from multiple temporal and spatial scales across all levels of biological organization.[7] These data are usually represented as a network or knowledge graph (KG), where nodes are biological entities (e.g., chemical compounds, proteins) and edges indicate relationships between these entities (e.g., interactions, drug-target affinity).[3,7,11,12] Systems pharmacology models have successfully predicted drug side effects,[6,13–15] developed precision therapeutics,[16,17] and facilitated drug repurposing.[18–20] While promising, the majority of these models focus on single entities providing an incomplete "biological milieu",[21] which may result in biased models and conclusions that are not biologically plausible.[3,7–9,11]

PheKnowLator (Phenotype Knowledge TransLator) Ecosystem[22,23] is an ecosystem for constructing ontologically-grounded KGs built on FAIR (findable, accessible, interoperable, reusable) data principles[24]. An overview of the PheKnowLator Ecosystem is shown in **Figure 1**. The usability of the PheKnowLator Ecosystem is facilitated through copious documentation, Jupyter Notebook demos, and interactive scripts that guide users through the construction process. Scalability within the PheKnowLator Ecosystem is achieved through intelligent build parallelization.

The goal of the proposed work is to demonstrate how a PheKnowLator KG can be traversed and used to construct features capable of discriminating different kinds of drug-outcome pairs.

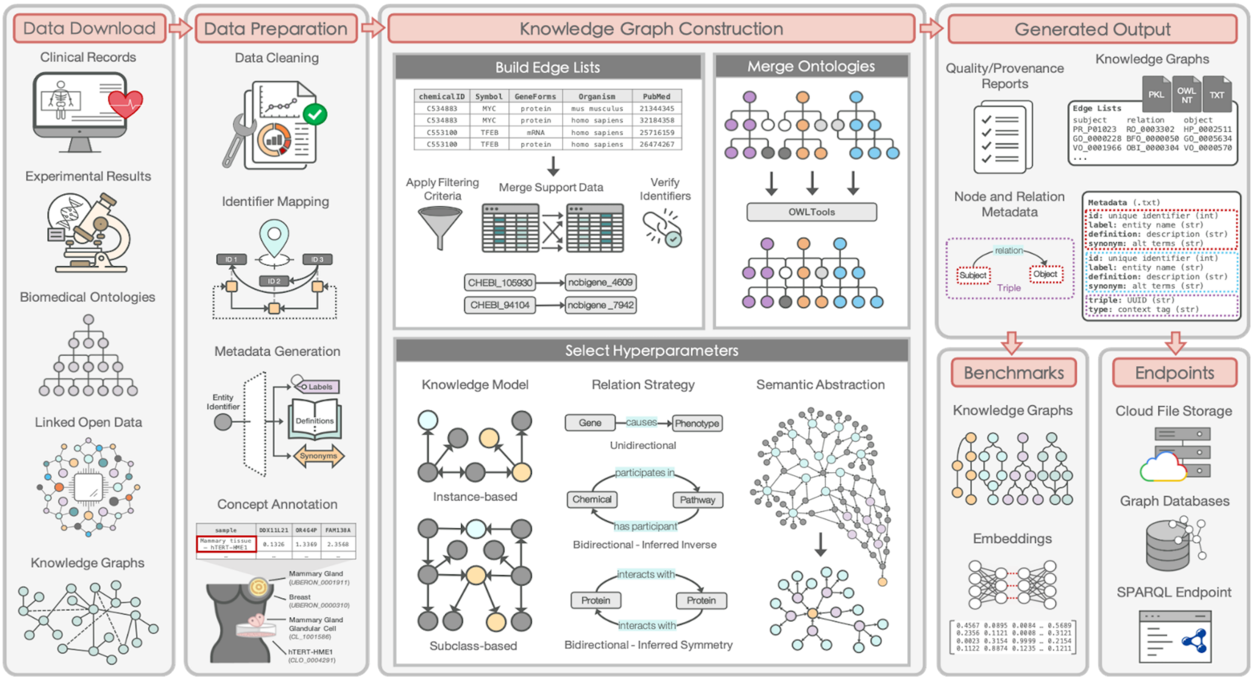
Methods

Knowledge Graph

The PheKnowLator Ecosystem provides monthly builds of open-source KGs designed to model the molecular mechanisms underlying human disease (PKT-KG). The knowledge model used to construct these builds was developed by a multidisciplinary team of domain experts and took over 3 years of collaborative discussion and empirical experimentation to complete (**Figure 2**). The PKT-KG was built with 12 Open Biomedical Ontology Foundry (OBO) ontologies and over 60 publicly available resources (the data sources are listed on the Wiki[25]). PKT-KG was visualized using the OpenOrd Force-Directed layout[26] provided by Gephi[27] (v0.9.2).

Network Descriptives

PKT-KG was explored using path-level statistics frequently used in the systems pharmacology domain,[28-31] which included: (1) **Efficiency.** A measure of how efficiently information is exchanged between nodes.



chemicalID	Symbol	GeneForms	Organism	PubMed
C534883	MYC	protein	mus musculus	21344345
C534883	MYC	protein	homo sapiens	32184358
C553100	TFEB	mRNA	homo sapiens	25716159
C553100	TFEB	protein	homo sapiens	26474267

sample	DDX11L21	OR4G4P	FAM138A
Mammary tissue – hTERT-HME1	0.1326	1.3369	2.3568
…	…	…	…

```
Edge Lists
subject    relation    object
PR_P01023  RO_0003302  HP_0002511
GO_0000228 BFO_0000050 GO_0005634
VO_0001966 OBI_0000304 VO_0000570
...
```

```
Metadata (.txt)
id: unique identifier (int)
label: entity name (str)
definition: description (str)
synonym: alt terms (str)
id: unique identifier (int)
label: entity name (str)
definition: description (str)
synonym: alt terms (str)
triple: UUID (str)
type: context tag (str)
```

```
0.4567 0.0895 0.0084 … 0.5689
0.2356 0.1121 0.0008 … 0.3121
0.0023 0.3154 0.9999 … 0.2154
0.1122 0.8874 0.1235 … 0.1211
```

Figure 1. The PheKnowLator Ecosystem[32] includes tools to download and prepare data, construct knowledge graphs, and generate a wide-range of outputs. These outputs support the production of benchmarks and are accessible through public endpoints. Acronyms - NT: N-Triples file format; OWL: Web Ontology Language; PKL: Python pickle file format; SPARQL: SPARQL Protocol and Resource Description Framework Query Language.

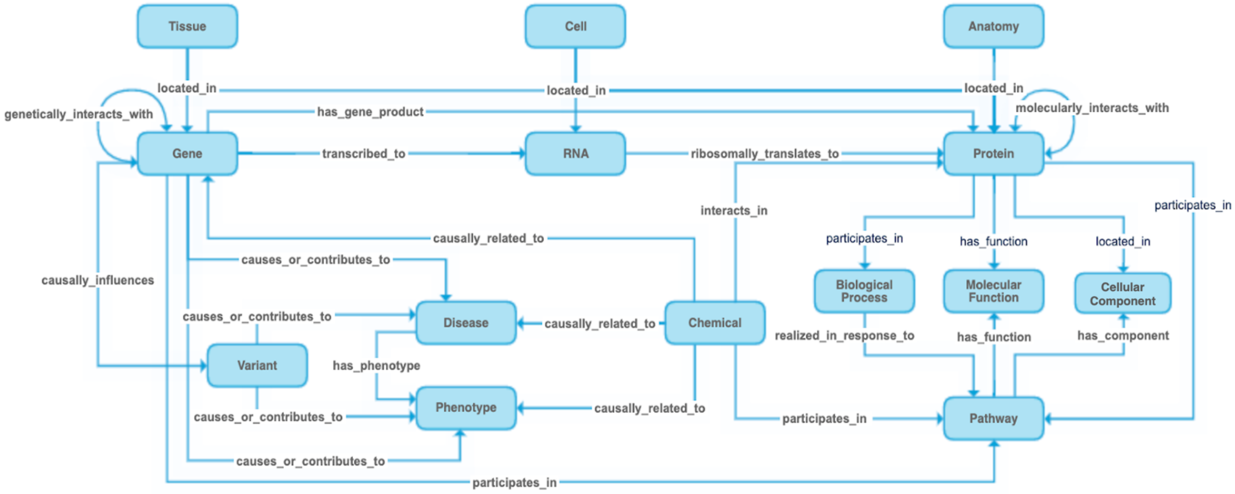


Figure 2. The knowledge model used to construct the PheKnowLator human disease mechanisms knowledge graph.

This metric ranges from 0-1, where 0 is assigned to distant node pairs and 1 is assigned to near pairs. Efficiency was calculated using an undirected version of PKT-KG; and (2) Shortest Paths. The shortest path length and count of shortest paths of this length. Shortest path descriptives were calculated using a directed version of PKT-KG.

Use Cases

Three different types of drug-outcome pairs were examined: (1) **Positive Pairs.** Known drug-outcome pairs. We examined lisinopril dihydrate and myocardial infarction; (2) **Negative Pairs.** Drug-outcome pairs known to not be related. We examined: (a) lisinopril dihydrate and contact dermatitis; (b) lisinopril dihydrate and ingrown toenail; and (c) lisinopril dihydrate and presbyopia; and (3) **Unknown Pairs.** Drug-outcome pairs with no known relationship. We examined ivermectin and neurotoxicity. All drug-outcome pairs were selected under the guidance of a domain expert.

Results

PKT-KG contained 743,829 nodes, 4,967,427 edges, 294 unique relations, 1 connected component, a density of 8.98E-06, and an average degree of 6.68. The counts by edge type are shown in **Table 1** and PKT-KG is visualized in **Figure 3**. The most prevalent edge types were protein-protein, rna-anatomy, and disease-phenotype.

Table 1. Counts of entities and relations by edge type.

Edge	Relation	Subjects	Objects	Edges
chemical-disease	causally related to	4,290	4,574	170,675
chemical-gene	interacts with	462	11,981	16,699
chemical-biological process	molecularly interacts with	1,338	1,584	288,921
chemical-cellular component	molecularly interacts with	1,086	250	44,553
chemical-molecular function	molecularly interacts with	1,105	208	26,165
chemical-pathway	participates in	2,105	2,213	28,691
chemical-phenotype	causally related to	4,055	1,721	108,452
chemical-protein	interacts with	4,179	6,389	65,124
disease-phenotype	has phenotype	11,746	9,717	414,193
gene-disease	causes or contributes to	5,035	4,429	12,735
gene-gene	genetically interacts with	247	263	1,668
gene-pathway	participates in	10,371	1,860	107,025
gene-phenotype	causes or contributes to	6,785	1,530	23,516
gene-protein	has gene product	19,327	19,143	19,534
gene-rna	transcribed to	25,529	179,870	182,736
biological process-pathway	realized in response to	471	665	665
pathway-cellular component	has component	11,134	99	15,846
pathway-molecular function	has function	2,412	726	2,416
protein-anatomy	located in	10,747	68	30,682
protein-catalyst/cofactor	molecularly interacts with	4,610	3,778	26,966
protein-cell	located in	10,045	128	75,318
protein-biological process	participates in	17,527	12,246	137,812
protein-cellular component	located in	18,427	1,757	81,602
protein-molecular function	has function	17,779	4,324	68,633
protein-pathway	participates in	10,886	2,480	117,585
protein-protein	molecularly interacts with	14,230	14,230	618,069
rna-anatomy	located in	29,115	103	444,668
rna-cell	located in	14,038	130	65,156
rna-protein	ribosomally translates to	44,144	19,200	44,147
variant-disease	causes or contributes to	13,297	3,621	38,129
variant-gene	causally influences	121,790	3,236	121,790
variant-phenotype	causes or contributes to	1,824	373	2,526

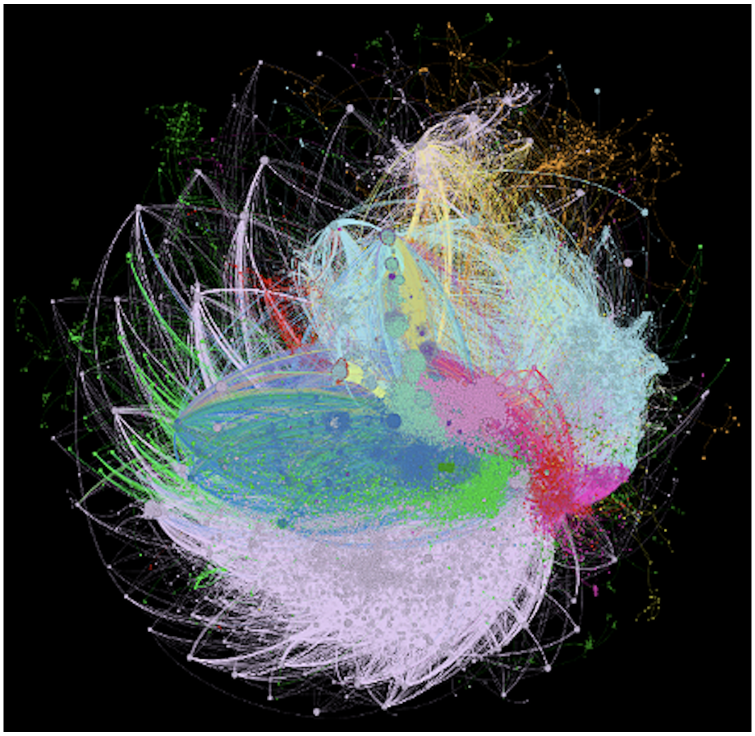


Figure 3. Visualization of the PheKnowLator molecular mechanisms of human disease knowledge graph. Nodes are colored by entity type: anatomical entities (light blue), chemicals (light purple), diseases (red), genes (purple), biological processes, cellular components, and molecular functions (light green), organisms (yellow), pathways (dark green), phenotypes (magenta), proteins (dark blue), sequence features (orange), transcripts (turquoise), and variants (light pink).

Drug-outcome pairs with at least 1 directed shortest path are shown in **Table 2**.

Positive Pairs. A single positive relationship was examined, lisinopril dihydrate-myocardial infarction. This pair had an efficiency of 1.0 and 1 shortest path of length 1.

Negative Pairs. Three negative relationships were examined: (1) lisinopril dihydrate-contact dermatitis. This pair had an efficiency of 0.33 and 14 shortest paths of length 4; (2) lisinopril dihydrate-ingrown toenail. This pair had an efficiency of 0.33 and no shortest paths; and (3) lisinopril dihydrate-presbyopia. This pair had an efficiency of 0.25 and no shortest paths.

Unknown Pairs. A single positive relationship was examined, ivermectin-neurotoxicity. This pair had an efficiency of 1.0 and 1 shortest path of length 1.

Conclusion

This work provides a demonstration of how a PheKnowLator KG can be traversed and used to construct features capable of discriminating positive, negative, and unknown drug-outcome pairs. These findings present several opportunities for future work. First, expanding this characterization to include additional drug-outcome pairs will be important to understanding if the identified patterns can be developed into robust and generalizable discriminatory heuristics. Second, exploring the utility of more complex network science and deep learning methods may provide addiitonal insight and lead to the development of more powerful heuristics. Finally, developing a PheknowLator KG for systems pharmacology, which includes a more detailed representation of systems biology as well as evidence and metadata may not only improve heuristics but may also yield more explainable molecular mechanisms.

Table 2. Shortest paths by relationship type.

Relationship Type	Drug-Outcome Pair	Shortest Paths
Positive	lisinopril dihydrate, myocardial infarction	lisinopril dihydrate - causally related to - Myocardial infarction
Negative	lisinopril dihydrate, contact dermatitis	lisinopril dihydrate - interacts with - endothelin-1 endothelin-1 - molecularly interacts with - CCN family member 2 CCN family member 2 - has_gene_template - CCN2 CCN2 - causes or contributes to condition - contact dermatitis *** lisinopril dihydrate - interacts with - angiotensin-converting enzyme angiotensin-converting enzyme - molecularly interacts with - CCN family member 2 CCN family member 2 - has_gene_template - CCN2 CCN2 - causes or contributes to condition - contact dermatitis *** lisinopril dihydrate - interacts with - angiotensinogen angiotensinogen - molecularly interacts with - C-C chemokine receptor type 1 C-C chemokine receptor type 1 - has_gene_template - CCR1 CCR1 - causes or contributes to condition - contact dermatitis *** lisinopril dihydrate - interacts with - endothelin-1 endothelin-1 - molecularly interacts with - serum amyloid A-1 protein serum amyloid A-1 protein - has_gene_template - SAA1 SAA1 - causes or contributes to condition - contact dermatitis *** lisinopril dihydrate - interacts with - renin renin - molecularly interacts with - cytochrome P450 11B2 cytochrome P450 11B2 - molecularly interacts with - cortisol cortisol - causally related to - contact dermatitis *** lisinopril dihydrate - interacts with - angiotensinogen angiotensinogen - molecularly interacts with - cytochrome P450 11B2 cytochrome P450 11B2 - molecularly interacts with - cortisol cortisol- causally related to - contact dermatitis *** lisinopril dihydrate - interacts with - angiotensinogen angiotensinogen - molecularly interacts with - CCN family member 2 CCN family member 2 - has_gene_template - CCN2 CCN2 - causes or contributes to condition - contact dermatitis *** lisinopril dihydrate - causally related to - inherited aplastic anemia inherited aplastic anemia- has phenotype - Low levels of vitamin E Low levels of vitamin E - inheres in - vitamin E vitamin E - causally related to - contact dermatitis *** lisinopril dihydrate - causally related to - idiopathic aplastic anemia idiopathic aplastic anemia - has phenotype - Low levels of vitamin E Low levels of vitamin E - inheres in - vitamin E vitamin E - causally related to - contact dermatitis *** lisinopril dihydrate - interacts with - angiotensin-converting enzyme angiotensin-converting enzyme - molecularly interacts with - cytochrome P450 11B2 cytochrome P450 11B2 - molecularly interacts with - cortisol cortisol - causally related to - contact dermatitis
Unknown	ivermectin, neurotoxicity	ivermectin - causally related to - toxic encephalopathy

References/Citations

1. Medication safety basics [Internet]. 2022. Available from: https://www.cdc.gov/medicationsafety/basics.html
2. da Silva BA, Krishnamurthy M. The alarming reality of medication error: a patient case and review of Pennsylvania and National data. J Community Hosp Intern Med Perspect. 2016;6(4):31758
3. Boland MR, Jacunski A, Lorberbaum T, Romano JD, Moskovitch R, Tatonetti NP. Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms. Wiley Interdiscip Rev Syst Biol Med. 2016;8(2):104–22
4. Berlin JA, Glasser SC, Ellenberg SS. Adverse event detection in drug development: recommendations and obligations beyond phase 3. Am J Public Health. 2008;98(8):1366–71
5. Raj N, Fernandes S, Charyulu NR, Dubey A, G S R, Hebbar S. Postmarket surveillance: a review on key aspects and measures on the effective functioning in the context of the United Kingdom and Canada. Ther Adv Drug Saf. 2019;10:2042098619865413
6. Berger SI, Iyengar R. Role of systems pharmacology in understanding drug adverse events. Wiley Interdiscip Rev Syst Biol Med. 2011;3(2):129–35
7. Trame MN, Biliouris K, Lesko LJ, Mettetal JT. Systems pharmacology to predict drug safety in drug development. Eur J Pharm Sci. 2016;94:93–5
8. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. Inf Fusion. 2019;50:71–91
9. Xie L, Draizen EJ, Bourne PE. Harnessing big data for systems pharmacology. Annu Rev Pharmacol Toxicol. 2017;57:245–62
10. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci Data. 2016;3:160026
11. Lorberbaum T, Nasir M, Keiser MJ, Vilar S, Hripcsak G, Tatonetti NP. Systems pharmacology augments drug safety surveillance. Clin Pharmacol Ther. 2015;97(2):151–8
12. Wu Q, Taboureau O, Audouze K. Development of an adverse drug event network to predict drug toxicity. Curr Res Toxicol. 2020;1:48–55
13. Bai JPF, Fontana RJ, Price ND, Sangar V. Systems pharmacology modeling: an approach to improving drug safety. Biopharm Drug Dispos. 2014;35(1):1–14
14. Berger SI, Ma'ayan A, Iyengar R. Systems pharmacology of arrhythmias. Sci Signal. 2010;3(118):ra30
15. Xie L, Li J, Xie L, Bourne PE. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. PLoS Comput Biol. 2009;5(5):e1000387
16. Zhao S, Nishimura T, Chen Y, Azeloglu EU, Gottesman O, Giannarelli C, et al. Systems pharmacology of adverse event mitigation by drug combinations. Sci Transl Med. 2013;5(206):206ra140
17. Bordbar A, McCloskey D, Zielinski DC, Sonnenschein N, Jamshidi N, Palsson BO. Personalized Whole-Cell Kinetic Models of Metabolism for Discovery in Genomics and Pharmacodynamics. Cell Syst. 2015;1(4):283–92
18. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. PLoS Comput Biol. 2009;5(7):e1000423
19. Ng C, Hauptman R, Zhang Y, Bourne PE, Xie L. Anti-infectious drug repurposing using an integrated chemical genomics and structural systems biology approach. Pac Symp Biocomput. 2014;136–47
20. Ho Sui SJ, Lo R, Fernandes AR, Caulfield MDG, Lerman JA, Xie L, et al. Raloxifene attenuates

Pseudomonas aeruginosa pyocyanin production and virulence. Int J Antimicrob Agents. 2012;40(3):246–51

21. Wilson JL, Racz R, Liu T, Adeniyi O, Sun J, Ramamoorthy A, et al. PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development. PLoS Comput Biol. 2018;14(12):e1006614
22. PheKnowLator: Heterogeneous Biomedical Knowledge Graphs and Benchmarks Constructed Under Alternative Semantic Models [Internet]. 2022. Github. Available from: https://github.com/callahantiff/PheKnowLator
23. Callahan TJ, Tripodi IJ, Wyrwa JM, Unni D, Reese J, Bennett TD, et al. Phenotype Knowledge Translator: A FAIR Ecosystem for Representing Large-Scale Biomedical Knowledge. Zenodo [Preprint]. DOI:10.5281/zenodo.5716383
24. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(1):1-9.
25. Release v3 Knowledge Graph Data Sources [Internet]. 2022. Github Wiki. Available from: https://github.com/callahantiff/PheKnowLator/wiki/v3-Data-Sources
26. Martin S, Brown WM, Klavans R, Boyack KW. OpenOrd: an open-source toolbox for large graph layout. In Visualization and Data Analysis. 2011;7868:45-55
27. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In Proceedings of the international AAAI conference on web and social media. 2009;3(1): 361-362
28. Rivas-Barragan D, Mubeen S, Guim Bernat F, Hofmann-Apitius M, Domingo-Fernández D. Drug2ways: Reasoning over causal paths in biological networks for drug discovery. PLoS Comput Biol. 2020;16(12):e1008464
29. Guney E, Menche J, Vidal M, Barábasi AL. Network-based in silico drug efficacy screening. Nat Commun. 2016;7:10331
30. Jiang Y, Li Y, Kuang Q, Ye L, Wu Y, Yang L, Li M. Predicting putative adverse drug reaction related proteins based on network topological properties. Anal Methods. 2014;6(8):2692–2698
31. Matsunaga S, Kishi T, Iwata N. Combination therapy with cholinesterase inhibitors and memantine for Alzheimer’s disease: a systematic review and meta-analysis. Int J Neuropsychopharmacol. 2014;18(5):pyu115
32. Callahan T J. Overview of the PheKnowLator Ecosystem (v1.0) [Internet]. Zenodo. 2022. DOI:10.5281/zenodo.6685133