Drug Repositioning using Consilience of Knowledge Graph Completion Methods

Roger Tu†, Meghamala Sinha†, Carolina González, Eric Hu, Shehzaad Dhuliawala, Andrew McCallum and Andrew I. Su

***Abstract*—While link prediction methods in knowledge graphs have been increasingly utilized to locate potential associations between compounds and diseases, they suffer from lack of sufficient evidence to prove why a drug and a disease may be indicated. This is especially true for knowledge graph embedding (KGE) based methods where a drug-disease indication is linked only by information gleaned from a vector representation. Complementary pathwalking algorithms can increase the confidence of drug repositioning candidates by traversing a knowledge graph. However, these methods heavily weigh the relatedness of drugs, through their targets, pharmacology or shared diseases. Furthermore, these methods rely on arbitrarily extracted paths as evidence of a compound to disease indication and lack the ability to make predictions on rare diseases. In this paper, we evaluate seven link prediction methods on a large biomedical knowledge graph on a drug repositioning task. We follow the principle of consilience, and combine the reasoning paths and predictions provided by probabilistic case-based reasoning (probCBR) with those of a KGE method, TransE, to identify putative drug repositioning indications. Finally, we highlight the utility of our approach through two potential repurposing indications.**

***Index Terms*—Drug repositioning, heterogenous network, case-based reasoning, knowledge graph embedding**

I. Introduction

THE estimated cost to bring a drug to market increased from $802 million to $2.7 billion between 2003 and 2013 due to drug attrition, longer development timelines and changing regulatory requirements, adjusting for inflation [1]. Drug repositioning, the process of identifying a new indication for an advanced clinical compound or approved drug, has become increasingly more attractive by leveraging prior work characterizing a drug candidate's safety and efficacy profile; resulting in a concomitant decrease in time to market, risk of failure, and investment costs [2].

There are two main classes of drug repositioning methods, experimental approaches may involve high throughput screens of drug ligand affinity binding, whereas computational approaches may utilize -omic scale fingerprints to conduct signature matching between a compound and a disease [3]. An emerging area in computational drug repositioning utilizes knowledge graphs, networks consisting of a set of entities bound together by relationships. In a drug repositioning knowledge graph, entities can consist of but are not limited to drugs, genes and diseases, and the link between the entities can signify an association or "treat" relationship.

Drug repositioning can be conducted by exploiting features in the knowledge graph to identify new relationships between a drug and a disease. A common method is to identify drug-drug and disease-disease similarities to find a repositioning candidate [4]–[6]. An alternative approach utilized graph traversal algorithms like shortest path and random walks, or path ranking algorithms like degree weighted path count to prioritize paths linking a drug and a disease [5], [7]–[11]. More specific graph traversal methods utilize deep learning methods by formalizing drug repositioning as a reinforcement learning task for link prediction [12], [13].

Link prediction methods like Knowledge Graph Embedding (KGE) and Case based reasoning (CBR) have enjoyed success in identifying relationships in large semantic based knowledge graphs and can be utilized to drug repositioning knowledge graphs to establish links between a drug and a disease. KGE algorithms learn a vector representation of each entity and relation in n-dimensional space for link prediction in a knowledge graph, and represent a class of novel methods for computational drug repositioning [14]–[18]. As knowledge graphs by nature are inherently incomplete, KGE are a powerful method for drug repositioning because they can extrapolate relationships between compounds and diseases in embedding space and without the restriction of traversing a knowledge graph. This is particularly useful for identifying drug repositioning candidates in rare diseases as they are not well characterized.

Case based reasoning (CBR) is an established technique from artificial intelligence modeled after human ability to retrieve and apply prior experience to tackling a new but similar challenge, and represents another promising approach to conduct drug repositioning [19]–[23]. While not specifically applied to drug repositioning, [24] demonstrated the utility and effectiveness of this simple approach for link prediction on knowledge graphs. Applied to drug repositioning, pathways from other similar compounds can be leveraged as evidence to identify a potential disease treatment given a compound of interest. This is beneficial as the analogy to other drugs provides mechanistic insight for the application of a compound to a disease.

In this paper, we apply the concept of consilience, that evidence from independent provenience can converge on a conclusion, to drug repositioning through the use of various

Roger Tu, Meghamala Sinha, Carolina González, Eric Hu and Andrew I. Su are with the Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA. E-mail: {rogertu, msinha, agonzalez, erichu, asu}@scripps.edu.

Shehzaad Dhuliawala and Andrew McCallum are with the College of Information of Computer Sciences, University of Massachusetts Amherst, Amherst, MA 01003 USA. E-mail: {dhuliawala, mccallum}@cs.umass.edu.

† denotes equal contribution

Manuscript received May XX, 2023

KG Completion algorithms like KGE and CBR. We highlight the performance of seven KGE, CBR and path based reasoning approaches on a large biomedical knowledge graph using approved drug-disease indications, and conduct a thorough pairwise analysis of prediction results between each algorithm evaluated. Finally, we demonstrate the effectiveness of our approach through manual curation and identifying plausible drug repositioning indications by combining the results of two complementary link prediction methods.

II. Methods

A. Mechanistic Repurposing Network with Indications Knowledge Graph

Mechanistic Repositioning Network with Indications (MIND) is a knowledge graph incorporating two biomedical resources: MechRepoNet and DrugCentral. MechRepoNet is a knowledge graph comprising of 18 biomedical resources that reflects and expands on important drug mechanism relationships identified from a curated biomedical drug mechanism dataset. MechRepoNet consists of 9,652,116 edges, 250,035 nodes, 9 node types and 22 relations [25]. DrugCentral is a curated, open-access, online resource that integrates structure, bioactivity, regulatory, pharmacologic actions and approved drug indications from the FDA and other regulatory agencies [26]. As of October 2021, DrugCentral has 11,292 indications, consisting of 2,494 unique compounds and 1,459 unique diseases.

MIND was created by mapping DrugCentral edges to existing MechRepoNet nodes through the Unified Medical Language System (UMLS) and Medical Subject Headings (MeSH). A total of 5,379 indication edges made up of 1,308 unique compounds and 1,030 unique diseases were mapped to MechRepoNet. DrugCentral edges incorporated in MIND are labeled have the relationship indication. The indication edge is differentiated from the treat edge in that the latter represents a weaker link between a drug and a disease as some treat edges are not approved by a regulatory body in contrast to the former. In MIND, when both indication and treat edges exist, indication superseded and replaced treat edges.

The MIND training set has a total of 9,651,040 edges. Figure 1 (upper) highlights total node to node and (lower) node to relation counts in MIND as a whole. MIND was separated into a train, and a test set based on DrugCentral indications. The indications were divided into a training set (80%; 4303 triples) and a test set (20%; 1076 triples).

B. Pathwalking Methods for Drug Repositioning

A variety of pathwalking algorithms have been utilized for knowledge graph based drug repositioning. Shortest path algorithms identify the smallest distance between two nodes in a graph while random walk algorithms search the space between two nodes to generate a probability distribution of the paths traveled. In the context of drug repositioning, the intuition of shortest path algorithms is that a shorter path between a compound and a disease signifies a more likely repositioning candidate than a longer path between a compound and a different disease. In contrast, random pathwalking algorithms identify a repositioning candidate based on the probability the path extending from a compound reaches a particular disease. The Rephetio Project repositioned compounds by calculating the prevalence of metapaths, a path based on node type [10]. Utilizing Degree Weighted Path Count, an algorithm that penalizes paths that travel through high-degree nodes when calculating metapath prevalence, Rephetio incorporated each metapath into a logistic regression and calculated the expected probability a compound treated a disease. Mayers et. al. expanded on this approach and incorporated rules based path exclusions and hyperparameter optimization schemes [25]. This pathwalking approach not only eliminated effects of high-degree nodes, but also removed similarity based metapaths; further improving the path interpretability for drug repositioning.

C. Knowledge Graph Completion Methods for Drug Repositioning

In this paper, we utilize and evaluate a variety of Knowledge Graph Completion algorithms for drug repositioning. These algorithms fall in two classes: knowledge graph embeddings, and case based reasoning.

1) Knowledge Graph Embeddings: Knowledge graph embedding algorithms model missing links in a graph by defining a scoring function for each triple. A knowledge graph is made up of a collection of triples (head, relation, tail), where $h, t \in \mathcal{E}$, the set of all entities, and r $\in \mathcal{R}$, the set of all relations in the graph. Each entity is represented by a vector embedding, and is modeled by a score function $f_r(h, t)$. The goal of knowledge graph embedding algorithms is to score true triplets (h, r, t) higher than corrupted (untrue) triplets (h', r, t) or (h, r, t'). In this paper, we train knowledge graph embedding models with the following algorithms: TransE, DistMult, ComplEx and RotatE. The scoring functions can be observed in Table I.

In brief, TransE, short for Translation Embedding, was developed in 2013 by Bordes et al., aimed to model relationships as translations in embedding space; where t should be close to the embedding of h when added to a vector, r [17]. DistMult, modeled relationships using a bilinear diagonal matrix and trilinear dot product as a scoring function [27]. ComplEx modified the DistMult model by embedding additional complex vector components but maintained the trilinear dot product as a scoring function [28]. Finally, RotatE modeled each relation as an element-wise rotation in complex vector spaces [29].

2) Case Based Reasoning: Case based reasoning is an elegant problem solving approach that leverages past observations with similar challenges to identify a solution. The intuition behind case based reasoning is similar to how a doctor may treat a patient; by relying on previous experiences treating similar patients, the doctor can propose a treatment for a new patient that draws upon their prior experience.

Case based reasoning models first retrieve similar entities to the query entity, h_q, that have the specified query relation, r_q given a query (h_q, r_q). Next, the set of relations connecting the similar entities to their answer via r_q is collected. Finally, the paths obtained are applied to the query entity h_q to identify the

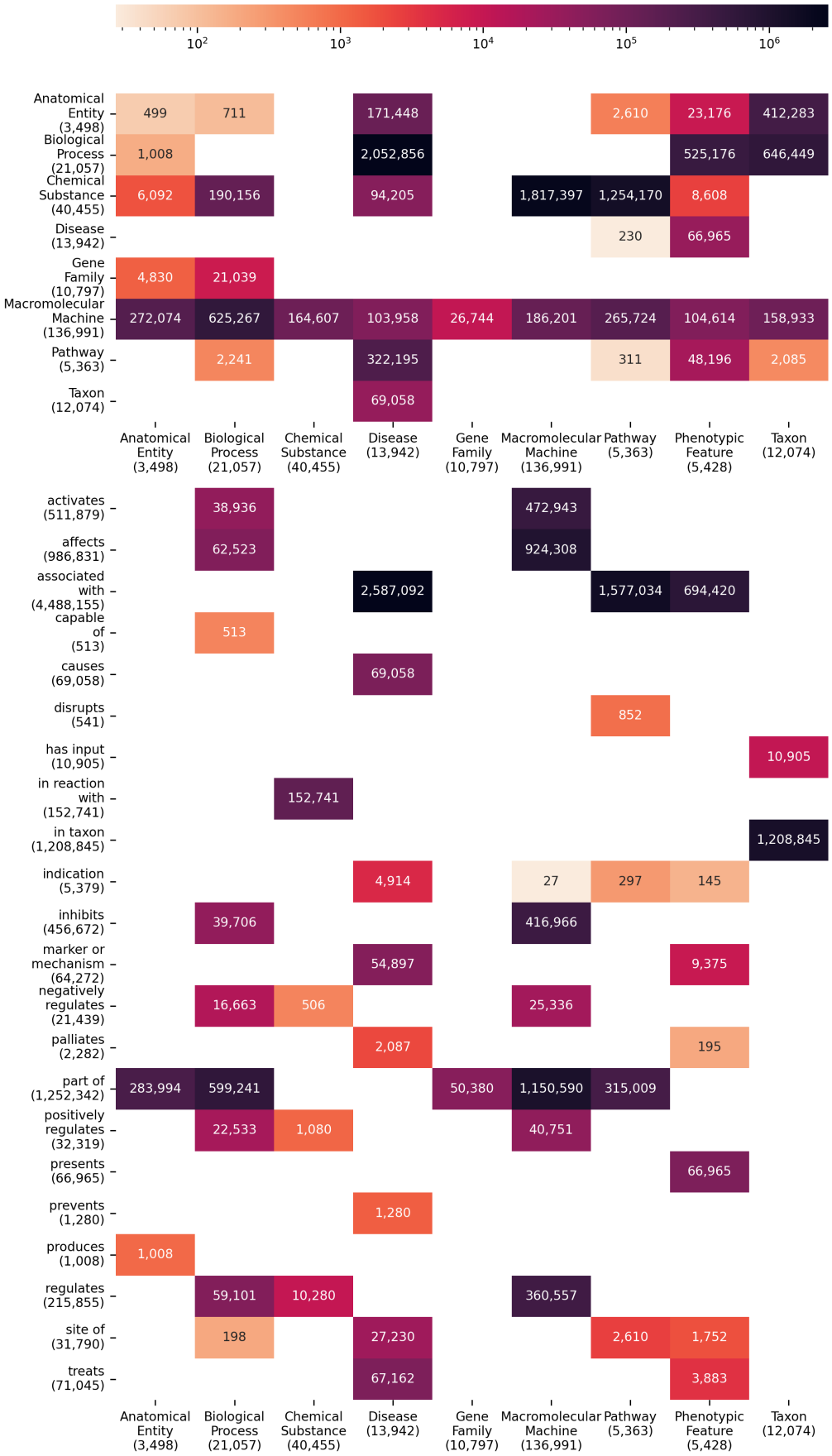

Fig. 1. Describes the composition of nodes and edges in the MIND dataset as a whole. (Upper): Heatmap of node type to node type counts. The y-axis and the x-axis correspond to the head and tail node respectively. (Lower): Heatmap of Node type (both head and tail) to relation type counts; Heatmap colors are in exponential scale where darker signifies more and lighter less, respectively. Parentheses highlight the node type and relation type counts.

TABLE I
SCORING FUNCTIONS FOR KNOWLEDGE GRAPH EMBEDDING ALGORITHMS. THE ANGLE BRACKETS IN DISTMULT AND COMPLEX DENOTE A TRILINEAR DOT PRODUCT. $\odot$ DENOTES THE HADAMARD PRODUCT (ELEMENT WISE PRODUCT) BETWEEN THE h AND r EMBEDDINGS.

Algorithm	Score Function
TransE	$-\|\|h + r - t\|\|$
DistMult	$\langle h, r, t\rangle$
ComplEx	$Re\langle h, r, t\rangle$
RotatE	$-\|\|h \odot r - t\|\|$

TABLE II
MEDIAN FIVE-FOLD CROSS VALIDATION PERFORMANCE RESULTS.

Algorithm	MRR	Hits@1	Hits@3	Hits@10
CBR	0.013781	0.004655	0.009311	0.027002
probCBR	**0.094672**	**0.069326**	0.101545	0.173195
TransE	0.093047	0.004646	**0.122676**	**0.251859**
RotatE	0.091986	0.056691	0.097584	0.147769
DistMult	0.041012	0.014870	0.035316	0.089219
ComplEx	0.053792	0.018587	0.049257	0.118030
Rephetio	0.087318	0.037529	0.104563	0.168251

answer. When multiple answers arise, the answers are ranked by the frequency of paths that lead to them [24]. Probabilistic case based reasoning extends the original case based reasoning method by improving the identification of similar entities with k-nearest neighbors approach, and by creating a probabilistic model that estimates the likelihood a retrieved path leads to a correct answer given query relation [30].

All algorithm hyperparameters were optimized using Optuna software [31]. Hyperparameter optimization parameters and search space, and results can be found at Supplementary Tables I and II, respectively. Next, each KGE model was trained on formerly mentioned datasets using the RotatE KGE library written by Sun et al [29]. Finally, knowledge graph completion prediction performance was evaluated on a subset of DrugCentral indications.

III. RESULTS

In this section, we applied several knowledge graph completion algorithms on the MIND dataset to identify drug repositioning candidates using the concept of consilience. First, we trained and evaluated the performance of each knowledge graph completion algorithm independently on DrugCentral indications in MIND. Next, we studied pairwise overlapping predictions between the algorithms to identify drug repositioning candidates. Finally, we gathered supporting evidence through manual curation on the identified drug repositioning candidates for supporting evidence in literature.

A. Knowledge Graph Completion Comparison

In order to identify the best drug repositioning knowledge graph completion algorithm, we evaluated the drug repositioning performance of each knowledge graph completion algorithm using a five-fold cross validation on the MIND dataset. For each fold, we trained each algorithm on 80% of compounds to disease indications, and tested the performance on the remaining 20%. Algorithms were evaluated using hits at k, which measures the success rate of identifying the correct answer among the top k predictions, and mean reciprocal rank (MRR), the average of the reciprocal of the positional rank for all correct answers. Overall, probCBR performed the best on average, with a median MRR score of 0.094672. The best KGE method on average was TransE, with a median MRR of 0.093047 across all five folds. The median results of the five-fold cross validation can be observed in Table 2 and the distribution of the results can be seen in Supplementary Figure 1.

B. Ranked Results Comparison

After identifying each algorithm's performance on the test dataset, we evaluated the similarity of each algorithm's predictions. All known answers for each indication edge were removed. Following the principle of consilience, that a given indication prediction is more likely to be true when independent approaches converge on a target, we evaluated the pairwise prediction overlaps between each algorithm to identify potential drug repositioning indications.

In order to assess prediction overlaps between algorithms, the top 100 edge predictions for every edge in the testing set were isolated and the overlap frequency distribution was calculated between each algorithm pair. The median of the overlaps are shown in Figure 2 where the maximum and minimum median overlap between each algorithm pair was observed to be 0.23 and 0.05, respectively. The KGE method, TransE, had the largest average pairwise overlap at 0.17. In contrast, DistMult had the lowest average pairwise overlap at 0.10. Notably, KGE methods overlap more within the class than outside the class of algorithms; the average overlap between KGE methods, between path-based methods and between a KGE and a path-based methods was 0.16, 0.14, and 0.11 respectively. As a whole, these results suggest that the top 100 edge results for each algorithm largely do not overlap.

C. Select diseases for Annotation

As the top 100 test edge predictions largely do not overlap between each algorithm, we sought to understand the drug repositioning prediction performance on identifying new indications. The two best performing algorithms in our five-fold cross validation study as shown in Table II were TransE and probCBR. We combined the results of those two algorithms with a manual literature search to assess the plausibility of each predicted indication.

In order to identify drug repositioning candidates, disease candidates were isolated for each given drug predicted by TransE and probCBR and the mean overlapping rank was calculated. Next, we filtered for drug to disease predictions

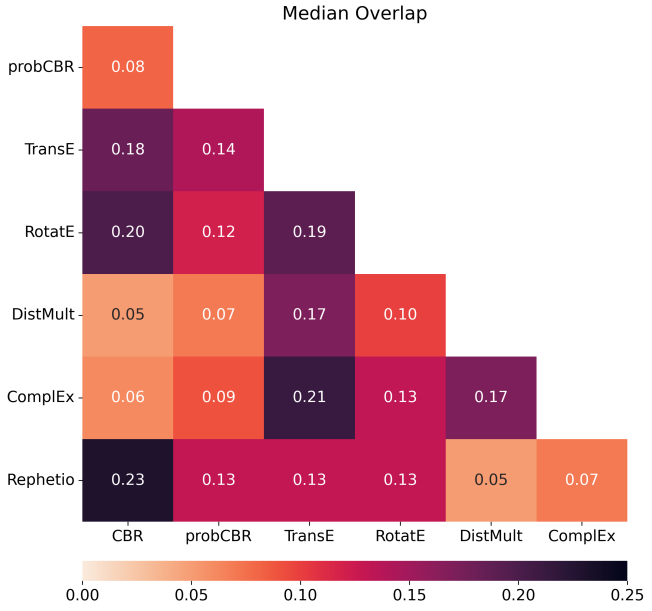

Fig. 2. Median of pairwise overlaps and rank difference for each edge in the test set. For each testing edge, the overlap frequency of the top 100 predictions between algorithm pairs were calculated; the median of the overlaps are shown.

with a mean overlapping rank of less than five. In the case a given drug has multiple indications with a mean overlapping rank of less than five, we keep the highest ranked indication. We identified 115 drug to disease indication edge predictions with a mean overlap rank of less than 5. These predictions were curated into three categories: reported, side-effect, and "no prior link". Indication predictions were labeled as "reported" if literature was found to support a drug-disease indication; these predicted edges have the potential to become drug indications, and warrant further studies. "Side-effects" are predicted edges that might induce or exacerbate the effects of the compound to cause a disease. Finally, "no prior link" are predicted edges where no literature support was found. Of the 115 potential indications filtered from our test set, 86.1% (99) were categorized as "reported", 3.5% (4) as "side-effects" and 10.4% (12) as "no prior link". The curation results can be found in Supplementary worksheet.

As a rough estimate of prediction performance, the first ranked predictions were additionally curated for TransE, and probCBR. We found that 76.5% (88) indications were "reported" in the literature and 5.2% (6) were "side-effects" and 18.3% (21) as "no prior link" for TransE. For probCBR, we found 73% (84) were "reported", 6.1% (7) were "side-effects" and 20.9% (24) as "no prior link". Plausible predictions from combining TransE and probCBR can be seen in Table III. We observe that by combining TransE and probCBR we decrease the amount of "side-effect" and "no prior link" indications and increase both "reported" indications. Through this comparison, combining multiple methods for drug repositioning improved the prediction performance compared to using each algorithm independently.

1) Case Study: Telmisartan as a potential treatment for congestive heart failure: Congestive heart failure (CHF) is a chronic condition in which the heart fails to adequately circulate blood to meet the body's needs due to the heart's inability to pump or fill; this results in fluid accumulation in the periphery of the body [48]. The onset of CHF is mainly attributed to genetic disposition, and the long term effects of coronary artery disease and hypertension [49]. There are various classes of compounds designed to prevent the onset of congestive heart failure. These compounds primarily have anti-hypertensive properties like decreasing heart rate (β blockers), decreasing blood volume (thiazides), and relaxing the blood vessels (angiotensin-converting enzyme (ACE) inhibitors and angiotensin II receptor blockers (ARBs)). Telmisartan, an ARBs and a PPAR-γ agonist, is approved to treat hypertension and reduce the risks of cardiovascular diseases like myocardial infarction, stroke or death for those who are unable to take ACE inhibitors; this includes but is not limited to patients experiencing pregnancy, impaired renal function, and angioedema [50], [51]. As telmisartan specifically targets and blocks angiotensin II receptors, it mediates vasoconstriction, resulting in lower blood pressure and improving the hearts' ability to transport blood throughout the body [52]. In the clinic, telmisartan successfully demonstrated significant mediation of both systolic and diastolic blood pressure with a lower incidence of treatment related side effects when compared to ramipril, an ACE inhibitor [53]–[56]. Utilizing our method, we propose an additional mechanism by which telmisartan mediates congestive heart failure through NPPB, also known as B-type natriuretic peptides. NPPB is both a marker and modulator of cardiac hypertrophy and is a hormone secreted by cardiomyocytes in both the atria and ventricles that act locally to reduce ventricular fibrosis in response to increased ventricular blood volume [57], [58]. As telmisartan directly affects NPPB expression in the heart when applied to a diabetic rat model, it is plausible that telmisartan's down regulation of NPPB, decreases the potential of cardiac hypertrophy and concomitant mediation of congestive heart failure [41]. We illustrate probCBRs predicted reasoning chain superimposed over mechanisms proposed by Rephetio in Figure 3. It is worth noting that TransE predicted telmisartan potentially treats diabetes mellitus. While this is not the primary indication of telmisartan, the use for treating diabetes is a popular off label use of the drug in combination with insulin in order to improve diabetic patients insulin sensitivity [39]. In contrast, probCBR predicted that telmisartan treats impotence. There is no evidence for this particular indication. A study found that telmisartan alone or in combination with an ACE inhibitor did not exhibit a beneficial effect on erectile dysfunction [59].

2) Case Study: Eseridine hydrochloride as a potential treatment for Alzheimer's disease: Alzheimer's disease is the most prevalent form of dementia, affects an estimated 11.3% of Americans and costs $1 trillion globally in 2020 [60]. Patients with Alzheimer's disease suffer from a decline in memory and independence due to progressive degradation and damage to the cells in the brain; this can lead to an increased economic and emotional burden for caretakers and the patient's family [61]. There are currently no FDA-approved cures for

TABLE III
SELECTED EXAMPLES OF TOP MOST PREDICTED 'DRUG INDICATION DISEASE' BY TRANSE, PROBCBR AND TRANSE & PROBCBR. PREDICTIONS WERE CATEGORIZED INTO "REPORTED", **"SIDE-EFFECT"** AND *"no prior link"*. THE PROBCBR PATH LINKING THE COMPOUND TO THE TRANSE & PROBCBR DISEASE IS SHOWN.

Compound	Topmost predicted answers by individual methods			Mechanistic path supporting the overlapping method
	TransE	probCBR	Overlapping Method, Mean rank	
Almotriptan	Persistent aura without infarction	**Major depressive disorder** [32]	**Major depressive disorder, 2.5** [32]	Almotriptan → associated_with → Metabolism → part_of_inv → SREBF1 → regulates → cholesterol biosynthetic process → associated_with → major depressive disorder
Clonidine hydrochloride	morphine dependence [37]	primary open angle glaucoma [38]	residual stage of open angle glaucoma, 4 [38]	Clonidine hydrochloride → affects_CafG → ADRA1D → in_reaction_with_GrxC → Epinephrine → indication → residual stage of open angle glaucoma
Eseridine	Organophosphorus poisoning [42]	*neuroblastoma*	Alzheimer's disease, 4 [43]	Eseridine → affects_CafG → BCHE → marker_or_mechanism_GmD → Alzheimer's disease
Lisdexamfetamine	narcolepsy [33]	epilepsy [34]	major depressive disorder, 3.5 [35]	Lisdexamfetamine → affects_CafG → SLC6A3 → inhibits_CinG_inv → Escitalopram → indication → major depressive disorder
Telavancin	*Chlamydia pneumonia*	*bronchitis*	Streptococcus pneumonia, 4.5 [44]	Telavancin → affects_CafG → ADRA2B → part_of_GpoBP → platelet activation → associated_with_BPawD → Streptococcus pneumonia
Telmisartan	Diabetes Mellitus [39]	*impotence* [40]	congestive heart failure, 2.5 [41]	Telmisartan → inhibits_CinG → NPPB → marker_or_mechanism_GmD" → congestive heart failure
Teriflunomide	Multiple endocrine neoplasia type 4 [46]	megakaryocytic leukemia [47]	megakaryocytic leukemia, 2 [47]	Teriflunomide → associated_with_CawPW → FoxO signaling pathway → part_of_GpoPW_inv → PRKAG1 → regulates_GrBP → protein kinase activity → associated_with_BPawD → megakaryocytic leukemia
Thiothixene	schizophreniform disorder [45]	*Autistic Disorder*	schizophreniform disorder, 4 [45]	Thiothixene → inhibits_CinG → HRH1 → marker_or_mechanism_GmD → schizophreniform disorder

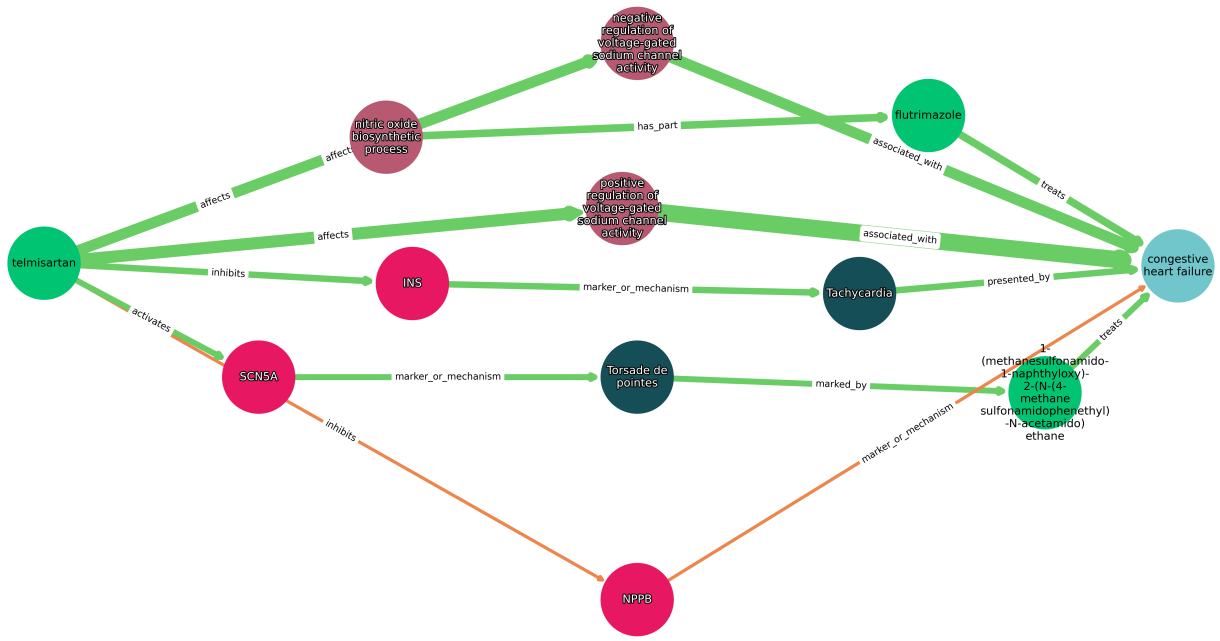

Fig. 3. Schematic illustrating the relationship between telmisartan and congestive heart failure. The path highlighted in orange is the top ranked path suggested by both TransE and probCBR. The remaining paths are paths suggested by Rephetio. All paths proposed by Rephetio between telmisartan and congestive heart failure had low percentile ranks; notably, the path through NPPB was not among the top 10 paths identified by Rephetio.

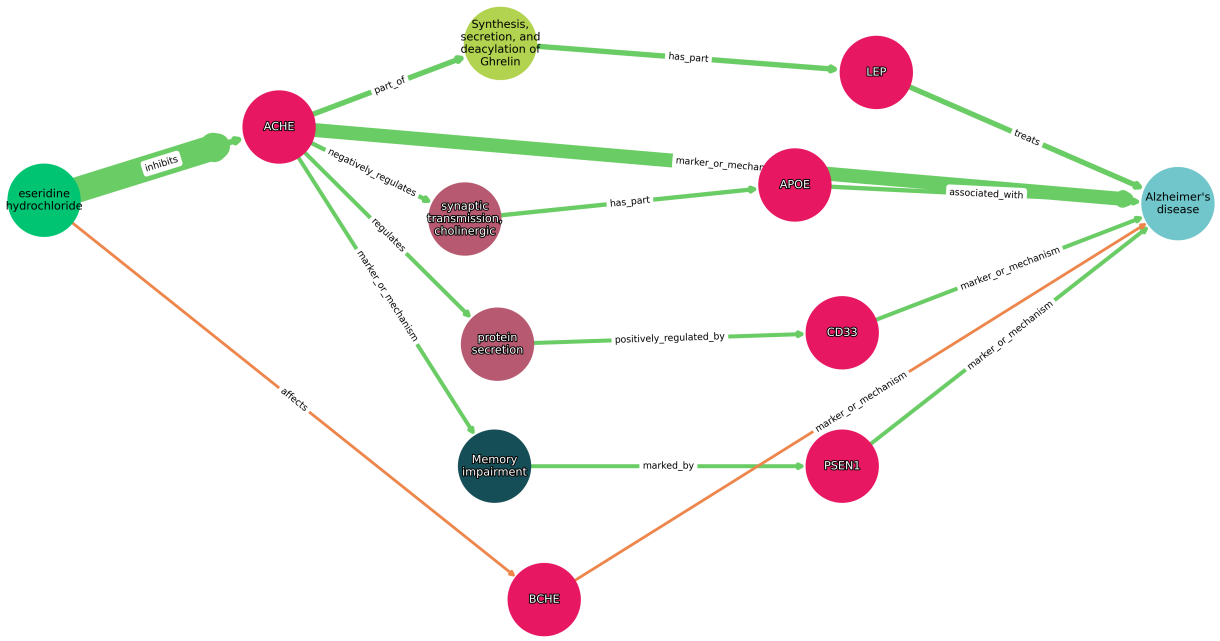

Fig. 4. Schematic illustrating the relationship between eseridine hydrochloride and Alzheimer's disease. The path highlighted in orange is the top ranked path suggested by both TransE and probCBR. The remaining paths are paths suggested by Rephetio. Notably, the path through BCHE was not identified in the top paths provided by Rephetio.

Alzheimer's disease. While immunotherapies like lecanemab and aducanumab have been demonstrated to moderately slow cognitive decline in Alzheimer's disease by targeting and reducing amyloid plaques (a marker of the disease), the long term safety and efficacy of the therapies remain to be evaluated [62], [63]. Some cholinesterase inhibitors like galantamine and rivastigmine, have been approved to manage Alzheimer disease symptoms. However, similar to the aforementioned immunotherapies, these drugs also lack evidence in mediating the onset of Alzheimer's disease. Utilizing our method, we identify eserdine to be a potential treatment of Alzheimer's disease through butyrylcholinesterase inhibition. Eseridine, an alkaloid isolated from Physostigma venenosum, is a potent neurotoxin and inhibitor of acetylcholinesterase and butyrylcholinesterase [42]. As deficiency of acetylcholine or increased acetylcholinesterase interferes with cholinergic neurotransmission, cholinergic intervention is believed to improve cognitive abilities in Alzheimer's disease [64]. Butyrylcholinesterase (BCHE) can hydrolyze acetylcholine and compensate for decreased levels of acetylcholinesterase activity in the progression of Alzheimer's disease [43]. Illustrated in Figure 4, we highlight probCBR's predicted reasoning chain superimposed on alternative mechanisms identified by Rephetio.

IV. Discussion

While link prediction methods have been largely successful in identifying drug repositioning targets, each approach suffers from drawbacks by design. As knowledge graphs are inherently incomplete, path based methods like Rephetio and CBR may not have the required edges to identify drug repositioning targets. In contrast, KGE methods are more resilient to missing edges but lack the ability to provide a reasoning chain for a potential drug repositioning indication. Even though CBR methods encounter the same problems that plague path walking algorithms like Rephetio, it avoids arbitrarily extracted paths by leveraging observations of relationships between similar drugs and the diseases they treat, before proposing a path to a predicted disease. Combining CBR and KGE approaches can synergistically improve drug repositioning performance by not only validating KGE derived predictions and highlighting a potential reasoning chain for a particular predicted drug disease indication but also allow CBR predictions to prioritize diseases with minimal similar cases in the graph.

Although this approach viably identifies potential repositioning candidates, we recognize some challenges in our implementation. Our results demonstrate that no disease predictions could be identified when overlapping predictions between all algorithms. This is likely due to limiting each algorithm's predictions to the top 100 ranked diseases. In addition, only compounds in our test set were evaluated (538 of 40,455 compounds) to limit computational expense. As a result, our analysis was limited to a pairwise study between TransE and probCBR.

Furthermore, our approach overlapping two algorithm predictions gives each algorithm an equal weight in its consideration of an indication irrespective of a disease prediction rank. Weighting each algorithm's contributions by the predicted indication rank or by incorporating an algorithm's cross validation performance into a logistic regression would potentially improve the prediction efficacy of the algorithm. These modifications would further fuel the addition of diverse algorithms into our implementation.

Finally, our utilization of CBR does not always avoid similarity based drug-disease associations, even when the results are filtered by those that also occur in TransE. For example, the pathway identified by our approach between lisdexamfetamine and major depressive disorder, traverses through escitalopram and its indication to major depressive disorder. Adding simple path similarity filters or penalties to the CBR retrieval mechanism would likely improve both the CBR and our own approach.

Regardless of variation of the consilience strategy applied, our results demonstrate that even with a simplistic approach, we were able to identify putative drug repositioning candidates that are supported by existing literature. Our strategy synergizes the advantages of both KGE and CBR based algorithms by blending the combined predictive power of TransE and probCBR. Through manual literature curation, we demonstrate that our method is more likely to identify a drug repurposing indication than when utilizing TransE or probCBR alone. Moreover, our method enables human interpretable reasoning chains derived from probCBR to support a putative compound to disease indication that would otherwise not be present with TransE alone.

Acknowledgments

The authors would like to thank Dr. Laura Hughes for visualization feedback, and Dr. Chunlei Wu, Dr. Jean-Christophe Ducom and Dr. Rajarshi Das for computational support. This work was supported by the National Institute of Health (R01 AG066750).

V. References Section

References

[1] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs," *J. Health Econ.*, vol. 47, pp. 20–33, May 2016, doi: 10.1016/j.jhealeco.2016.01.012.

[2] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nat. Rev. Drug Discov.*, vol. 3, no. 8, pp. 673–683, Aug. 2004, doi: 10.1038/nrd1468.

[3] S. Pushpakom et al., "Drug repurposing: progress, challenges and recommendations," *Nat. Rev. Drug Discov.*, vol. 18, no. 1, pp. 41–58, Jan. 2019, doi: 10.1038/nrd.2018.168.

[4] J. Li and Z. Lu, "A new method for computational drug repositioning using drug pairwise similarity," in *2012 IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, PA, USA: IEEE, Oct. 2012, pp. 1–4. doi: 10.1109/BIBM.2012.6392722.

[5] H. Luo et al., "Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, Sep. 2016, doi: 10.1093/bioinformatics/btw228.

[6] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinformatics*, vol. 34, no. 11, pp. 1904–1912, Jun. 2018, doi: 10.1093/bioinformatics/bty013.

[7] H. Luo et al., "Computational Drug Repositioning with Random Walk on a Heterogeneous Network," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 16, no. 6, pp. 1890–1900, Nov. 2019, doi: 10.1109/TCBB.2018.2832078.

[8] T. Lee and Y. Yoon, "Drug repositioning using drug-disease vectors based on an integrated network," *BMC Bioinformatics*, vol. 19, no. 1, p. 446, Dec. 2018, doi: 10.1186/s12859-018-2490-x.

[9] D. Emig et al., "Drug Target Prediction and Repositioning Using an Integrated Network-Based Approach," *PLoS ONE*, vol. 8, no. 4, p. e60618, Apr. 2013, doi: 10.1371/journal.pone.0060618.

[10] D. S. Himmelstein et al., "Systematic integration of biomedical knowledge prioritizes drugs for repurposing," *eLife*, vol. 6, p. e26726, Sep. 2017, doi: 10.7554/eLife.26726.

[11] M. Mayers, T. S. Li, N. Queralt-Rosinach, and A. I. Su, "Time-resolved evaluation of compound repositioning predictions on a text-mined knowledge network," *BMC Bioinformatics*, vol. 20, no. 1, p. 653, Dec. 2019, doi: 10.1186/s12859-019-3297-0.

[12] Y. Liu, M. Hildebrandt, M. Joblin, M. Ringsquandl, R. Raissouni, and V. Tresp, "Neural Multi-hop Reasoning with Logical Rules on Biomedical Knowledge Graphs," in *The Semantic Web*, R. Verborgh, K. Hose, H. Paulheim, P.-A. Champin, M. Maleshkova, O. Corcho, P. Ristoski, and M. Alam, Eds., in Lecture Notes in Computer Science, vol. 12731. Cham: Springer International Publishing, 2021, pp. 375–391. doi: 10.1007/978-3-030-77385-4_22.

[13] H. Wang, F. Huang, Z. Xiong, and W. Zhang, "A heterogeneous network-based method with attentive meta-path extraction for predicting drug–target interactions," *Brief. Bioinform.*, vol. 23, no. 4, p. bbac184, Jul. 2022, doi: 10.1093/bib/bbac184.

[14] S. Bonner et al., "Understanding the performance of knowledge graph embeddings in drug discovery," *Artif. Intell. Life Sci.*, vol. 2, p. 100036, Dec. 2022, doi: 10.1016/j.ailsci.2022.100036.

[15] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, and M. Dumontier, "Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings," *BMC Bioinformatics*, vol. 20, no. 1, p. 726, Dec. 2019, doi: 10.1186/s12859-019-3284-5.

[16] S. K. Mohamed, A. Nounu, and V. Nováček, "Biological applications of knowledge graph embedding models," *Brief. Bioinform.*, vol. 22, no. 2, pp. 1679–1693, Mar. 2021, doi: 10.1093/bib/bbaa012.

[17] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, in NIPS'13. Red Hook, NY, USA: Curran Associates Inc., Dec. 2013, pp. 2787–2795.

[18] K. Toutanova, V. Lin, W. Yih, H. Poon, and C. Quirk, "Compositional Learning of Embeddings for Relation Paths in Knowledge Base and Text," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany: Association for Computational Linguistics, 2016, pp. 1434–1444. doi: 10.18653/v1/P16-1136.

[19] R. C. Schank, *Dynamic memory: A theory of reminding and learning in computers and people*. cambridge university press, 1983.

[20] J. L. Kolodner, "Maintaining organization in a dynamic long-term memory," *Cogn. Sci.*, vol. 7, no. 4, pp. 243–280, 1983.

[21] E. L. Rissland, "Examples in Legal Reasoning: Legal Hypotheticals.," in *IJCAI*, 1983, pp. 90–93.

[22] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994.

[23] D. B. Leake, *CBR in context: the present and future. Case based reasoning experiences-lessons and future experiences.* D. Leake. Cambridge, MIT Press, 1996.

[24] R. Das, A. Godbole, S. Dhuliawala, M. Zaheer, and A. McCallum, "A Simple Approach to Case-Based Reasoning in Knowledge Bases," *arXiv*, vol. abs/2006.14198, 2020, doi: 10.48550/arXiv.2006.14198.

[25] M. Mayers, R. Tu, D. Steinecke, T. S. Li, N. Queralt-Rosinach, and A. I. Su, "Design and application of a knowledge network for automatic prioritization of drug mechanisms," *Bioinforma. Oxf. Engl.*, vol. 38, no. 10, pp. 2880–2891, May 2022, doi: 10.1093/bioinformatics/btac205.

[26] O. Ursu et al., "DrugCentral: online drug compendium," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D932–D939, Jan. 2017, doi: 10.1093/nar/gkw993.

[27] B. Yang, W. Yih, X. He, J. Gao, and L. Deng, "Embedding Entities and Relations for Learning and Inference in Knowledge Bases." *arXiv*, Aug. 29, 2015. Accessed: Sep. 12, 2022. [Online]. Available: http://arxiv.org/abs/1412.6575

[28] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, in ICML'16. New York, NY, USA: JMLR.org, Jun. 2016, pp. 2071–2080.

[29] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "ROTATE: KNOWLEDGE GRAPH EMBEDDING BY RELA- TIONAL ROTATION IN COMPLEX SPACE," *arXiv*, vol. abs/1902.10197, p. 18, 2019, doi: 10.48550/arXiv.1902.10197.

[30] R. Das, A. Godbole, N. Monath, M. Zaheer, and A. McCallum, "Probabilistic case-based reasoning for open-world knowledge graph completion," *ArXiv Prepr. ArXiv201003548*, 2020.

[31] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework."

arXiv, Jul. 25, 2019. Accessed: Sep. 01, 2022. [Online]. Available: http://arxiv.org/abs/1907.10902

[32] X. Cabarrocas, R. Esbri, F. Peris, and P. Ferrer, "Long-term Efficacy and Safety of Oral Almotriptan: Interim Analysis of a 1-Year Open Study," *Headache J. Head Face Pain*, vol. 41, no. 1, pp. 57–62, 2001, doi: 10.1046/j.1526-4610.2001.111006057.x.

[33] A. C. Aguilar, C. Frange, L. H. Pimentel Filho, M. J. Reis, S. Tufik, and F. M. S. Coelho, "Lisdexamfetamine to improve excessive daytime sleepiness and weight management in narcolepsy: a case series," *Braz. J. Psychiatry*, vol. 42, no. 3, pp. 314–316, Dec. 2019, doi: 10.1590/1516-4446-2019-0544.

[34] K. K. Wiggs et al., "Attention-deficit/hyperactivity disorder medication and seizures," *Neurology*, vol. 90, no. 13, pp. e1104–e1110, Mar. 2018, doi: 10.1212/WNL.0000000000005213.

[35] M. Madhoo et al., "Lisdexamfetamine Dimesylate Augmentation in Adults With Persistent Executive Dysfunction After Partial or Full Remission of Major Depressive Disorder," *Neuropsychopharmacology*, vol. 39, no. 6, Art. no. 6, May 2014, doi: 10.1038/npp.2013.334.

[36] P. Sonneveld, "Teniposide in lymphomas and leukemias," *Semin. Oncol.*, vol. 19, no. 2 Suppl 6, pp. 59–64, Apr. 1992.

[37] J. L. Katz, "Effects of clonidine and morphine on opioid withdrawal in rhesus monkeys," *Psychopharmacology (Berl.)*, vol. 88, no. 3, pp. 392–397, Mar. 1986, doi: 10.1007/BF00180844.

[38] R. Harrison and C. S. Kaufmann, "Clonidine: Effects of a Topically Administered Solution on Intraocular Pressure and Blood Pressure in Open-Angle Glaucoma," *Arch. Ophthalmol.*, vol. 95, no. 8, pp. 1368–1373, Aug. 1977, doi: 10.1001/archopht.1977.04450080078007.

[39] N. Suksomboon, N. Poolsup, and T. Prasit, "Systematic review of the effect of telmisartan on insulin sensitivity in hypertensive patients with insulin resistance or diabetes," *J. Clin. Pharm. Ther.*, vol. 37, no. 3, pp. 319–327, 2012, doi: 10.1111/j.1365-2710.2011.01295.x.

[40] S. B. Ismail, N. M. Noor, N. H. N. Hussain, Z. Sulaiman, M. A. Shamsudin, and M. Irfan, "Angiotensin Receptor Blockers for Erectile Dysfunction in Hypertensive Men: A Brief Meta-Analysis of Randomized Control Trials," *Am. J. Mens Health*, vol. 13, no. 6, p. 1557988319892735, Dec. 2019, doi: 10.1177/1557988319892735.

[41] V. Malek and A. B. Gaikwad, "Telmisartan and thiorphan combination treatment attenuates fibrosis and apoptosis in preventing diabetic cardiomyopathy," *Cardiovasc. Res.*, vol. 115, no. 2, pp. 373–384, Feb. 2019, doi: 10.1093/cvr/cvy226.

[42] Q. Yu, X. Zhu, H. W. Holloway, N. F. Whittaker, A. Brossi, and N. H. Greig, "Anticholinesterase Activity of Compounds Related to Geneserine Tautomers. N -Oxides and 1,2-Oxazines," *J. Med. Chem.*, vol. 45, no. 17, pp. 3684–3691, Aug. 2002, doi: 10.1021/jm010491d.

[43] A. Nordberg, C. Ballard, R. Bullock, T. Darreh-Shori, and M. Somogyi, "A Review of Butyrylcholinesterase as a Therapeutic Target in the Treatment of Alzheimer's Disease," *Prim. Care Companion CNS Disord.*, vol. 15, no. 2, p. 26731, Mar. 2013, doi: 10.4088/PCC.12r01412.

[44] B. Das, C. Sarkar, D. Das, A. Gupta, A. Kalra, and S. Sahni, "Telavancin: a novel semisynthetic lipoglycopeptide agent to counter the challenge of resistant Gram-positive pathogens," *Ther. Adv. Infect. Dis.*, vol. 4, no. 2, pp. 49–73, Mar. 2017, doi: 10.1177/2049936117690501.

[45] C. Xin, W. Lihong, L. Qiuyuan, and L. Hongzhuo, "Injectable long-term control-released in situ gels of hydrochloric thiothixene for the treatment of schizophrenia: Preparation, in vitro and in vivo evaluation," *Int. J. Pharm.*, vol. 469, no. 1, pp. 23–30, Jul. 2014, doi: 10.1016/j.ijpharm.2014.04.044.

[46] L. Jiang, W. Zhang, W. Li, C. Ling, and M. Jiang, "Anti-inflammatory drug, leflunomide and its metabolite teriflunomide inhibit NSCLC proliferation in vivo and in vitro," *Toxicol. Lett.*, vol. 282, pp. 154–165, Jan. 2018, doi: 10.1016/j.toxlet.2017.10.013.

[47] J. T. Madak, A. Bankhead, C. R. Cuthbertson, H. D. Showalter, and N. Neamati, "Revisiting the role of dihydroorotate dehydrogenase as a therapeutic target for cancer," *Pharmacol. Ther.*, vol. 195, pp. 111–131, Mar. 2019, doi: 10.1016/j.pharmthera.2018.10.012.

[48] W. Huckabee, G. Casten, and T. R. Harrison, "Experimental Hypervolemic Heart Failure: Its Bearing on Certain General Principles of Heart Failure," *Circulation*, vol. 1, no. 3, pp. 343–356, Mar. 1950, doi: 10.1161/01.CIR.1.3.343.

[49] E. Braunwald and M. R. Bristow, "Congestive Heart Failure: Fifty Years of Progress," *Circulation*, vol. 102, no. suppl_4, p. Iv–14, Nov. 2000, doi: 10.1161/circ.102.suppl_4.IV-14.

[50] "Micardis (telmisartan) tablets".

[51] S. C. Benson et al., "Identification of Telmisartan as a Unique Angiotensin II Receptor Antagonist With Selective PPARγ–Modulating Activity," *Hypertension*, vol. 43, no. 5, pp. 993–1002, May 2004, doi: 10.1161/01.HYP.0000123072.34629.57.

[52] K. J. McClellan and A. Markham, "Telmisartan," *Drugs*, vol. 56, no. 6, pp. 1039–44; discussion 1045-6, Dec. 1998, doi: 10.2165/00003495-199856060-00007.

[53] B. Williams, P. Gosse, L. Lowe, R. Harper, and on behalf of the P. I. S. Group, "The prospective, randomized investigation of the safety and efficacy of telmisartan versus ramipril using ambulatory blood pressure monitoring (PRISMA I)," *J. Hypertens.*, vol. 24, no. 1, p. 193, Jan. 2006, doi: 10.1097/01.hjh.0000194364.11516.ab.

[54] Lacourcière et al., "A Multicenter, 14-Week Study of Telmisartan and Ramipril in Patients With Mild-to-Moderate Hypertension Using Ambulatory Blood Pressure Monitoring," *Am. J. Hypertens.*, vol. 19, no. 1, pp. 104–112, Jan. 2006, doi: 10.1016/j.amjhyper.2005.10.001.

[55] Yusuf et al., "Telmisartan, Ramipril, or Both in Patients at High Risk for Vascular Events," *N. Engl. J. Med.*, vol. 358, no. 15, pp. 1547–1559, Apr. 2008, doi: 10.1056/NEJMoa0801317.

[56] J. Redon et al., "Safety and Efficacy of Low Blood Pressures Among Patients With Diabetes," *J. Am. Coll. Cardiol.*, vol. 59, no. 1, pp. 74–83, Jan. 2012, doi: 10.1016/j.jacc.2011.09.040.

[57] L. R. Potter, A. R. Yoder, D. R. Flora, L. K. Antos, and D. M. Dickey, "Natriuretic peptides: their structures, receptors, physiologic functions and therapeutic applications," *Handb. Exp. Pharmacol.*, no. 191, pp. 341–366, 2009, doi: 10.1007/978-3-540-68964-5_15.

[58] D. G. Gardner, "Natriuretic peptides: markers or modulators of cardiac hypertrophy?," *Trends Endocrinol. Metab.*, vol. 14, no. 9, pp. 411–416, Nov. 2003, doi: 10.1016/S1043-2760(03)00113-9.

[59] V. Javaroni and M. F. Neves, "Erectile Dysfunction and Hypertension: Impact on Cardiovascular Risk and Treatment," *Int. J. Hypertens.*, vol. 2012, p. e627278, May 2012, doi: 10.1155/2012/627278.

[60] K. B. Rajan, J. Weuve, L. L. Barnes, E. A. McAninch, R. S. Wilson, and D. A. Evans, "Population Estimate of People with Clinical AD and Mild Cognitive Impairment in the United States (2020–2060)," *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 17, no. 12, pp. 1966–1975, Dec. 2021, doi: 10.1002/alz.12362.

[61] P. Winston Wong, "Economic Burden of Alzheimer Disease and Managed Care Considerations," vol. 26, Aug. 2020, Accessed: Apr. 27, 2023. [Online]. Available: https://www.ajmc.com/view/economic-burden-of-alzheimer-disease-and-managed-care-considerations

[62] C. H. van Dyck et al., "Lecanemab in Early Alzheimer's Disease," *N. Engl. J. Med.*, vol. 388, no. 1, pp. 9–21, Jan. 2023, doi: 10.1056/NEJMoa2212948.

[63] S. B. Haeberlein et al., "Emerge and Engage topline results: Phase 3 studies of aducanumab in early Alzheimer's disease," *Alzheimers Dement.*, vol. 16, no. S9, p. e047259, 2020, doi: 10.1002/alz.047259.

[64] M. P. Arce et al., "Neuroprotective and Cholinergic Properties of Multifunctional Glutamic Acid Derivatives for the Treatment of Alzheimer's Disease," *J. Med. Chem.*, vol. 52, no. 22, pp. 7249–7257, Nov. 2009, doi: 10.1021/jm900628z.

VI. Biography Section

Roger Tu received a BS degree in biological sciences from the University of California - Irvine, in 2015. From 2015 to 2018 he was employed with Seattle Genetics Inc. as a research associate in Process Sciences. Since 2018 he has been studying the application of link prediction methods for drug repositioning under the guidance of professor Andrew I. Su at The Scripps Research Institute.

Meghamala Sinha completed her PhD from Oregon State University with a major in computer science and minor in biological data science. Her dissertation research centered around using fundamentals of Causality to differentiate true cause-effect relationships from mere associations in data and building a more robust and reliable inference model. She is currently working as a staff scientist at The Scripps Research Institute where she is working towards mining large scale knowledge graphs to explore drug repurposing.

Carolina González earned her BS degree in Biotechnology and Genomics from Universidad Autónoma de Nuevo León, Mexico. During her senior year, she participated in the Visiting Undergraduate Research Fellowship Program at California Institute of Technology, division of Biology and Biological Engineering. Currently, she is pursuing her PhD degree at The Scripps Research Institute, under the supervision of Professor Andrew I. Su. Her research interests focus on leveraging state-of-the-art graph neural network approaches to model biomedical knowledge graphs for drug repurposing.

Eric Hu received his MSc degree in Computer Science in 2014 and PhD in Biomedical Engineering in 2018, both from University of Southern California. He worked as a CV data scientist for 2 years before joining the Scripps Research Institute in 2020 to take up a postdoctoral position in investigating drug repositioning targets for Alzheimer's Disease through analysis of electronic health record data. He is currently a Biostatistician at the Scripps Health Hospital.

Shehzaad Dhuliawala received his MSc in Computer Science from the University of Massachusetts, Amherst in 2018 studying under Andrew McCallum. From 2018 to 2020, he was a Research Engineer at Microsoft Research. Since 2020, Dhuliawla has been a PhD student at ETH Zürich. He is interested in building reasoning systems that are explainable, trustable, and robust to distributional shifts. Broadly, he works in machine learning and its applications in natural language processing.

Andrew McCallum graduated summa cum laude with his BA from Dartmouth College in 1989 and received his PhD in Computer Science in 1995 from University of Rochester, New York. He completed his post-doctoral studies at Carnegie Mellon in 1996 and continued to become a Research Scientist and Coordinator at Just Research while concurrently serving as an Adjunct Professor at Carnegie Mellon University Center for Automated Learning and Discovery & Language Technologies Institute. From 2000 to 2002, he served as the VP of Research and Development at WhizBang! Labs as a director of its Pittsburgh office. Since 2002 he has been a Professor at the University of Massachusetts, Amherst, where he is focused on researching statistical machine learning applied to text, with a specific interest in information extraction, entity resolution, information integration, structured prediction, clustering, deep neural networks, semi-supervised learning, especially applied to accelerating natural sciences and engineering.

Andrew I. Su eceived his BS degree in both Chemistry and Computer Science from Northwestern University, Evanston, Illinois in 1998 and PhD in Chemistry from The Scripps Research Institute in 2002. From 2002 to 2011, Andrew led a bioinformatics group at The Genomics Institute of the Novartis Foundation. Since 2011 he has been a professor at The Scripps Research Institute, where he is focused on building and applying bioinformatics infrastructure for biomedical discovery with a particular emphasis on leveraging crowdsourcing for genetics, genomics and drug repurposing.