Beyond Knowledge Graphs: PubMedBERT Embeddings as a Competitive Standalone Modality for Drug Repurposing

Rishik K. Kondadadi
University of Minnesota
Eastview High School
konda052@umn.edu

John E. Ortega
Northeastern University
j.ortega@northeastern.edu

Abstract

Drug repurposing methods rely heavily on knowledge graph (KG) embeddings, but building and curating these graphs takes considerable effort. We present two findings on the Hetionet drug-disease benchmark and an epilepsy ranking task. *First*, PubMedBERT text embeddings, fed through the same downstream classifiers and identical 10-fold splits as four re-trained KG baselines (TransE, ComplEx, DistMult, RotatE), reach AUROC 0.910, above all four (best: RotatE, 0.854); a Random Forest on the same vectors scores 0.880. The comparison is asymmetric in one important way: PubMedBERT was pretrained on the literature Hetionet was curated from, so the result is best read as "text-with-literature-supervision vs. graph-only," and a head-to-head with text-augmented KG methods (KG-BERT, TxGNN) is left as follow-up. *Second*, across all seven combinations of text, molecular (ECFP4), and gene expression (LINCS L1000) features, cross-attention fusion of weaker modalities into text consistently degrades performance, despite a gated mechanism intended to suppress unhelpful modalities; the residual path forces the strong modality to absorb noise. The model also ranks proconvulsants (amoxapine, flumazenil) near the top, because text embeddings encode strength of association with a disease but not its direction.

1 Introduction

Finding new therapeutic uses for existing drugs remains one of the more tractable problems in pharmaceutical development, in part because repurposed candidates already have established safety profiles (Pushpakom et al., 2019; Ashburn and Thor, 2004). The potential payoff is large for diseases like epilepsy, where roughly a third of patients are not adequately controlled by current medications (Löscher et al., 2020).

The computational side of drug repurposing has become dominated by knowledge graph (KG) methods, which represent biomedical knowledge as a network of drugs, diseases, genes, and pathways connected by typed edges (Himmelstein et al., 2017; Nickel et al., 2016). Embedding methods such as TransE (Bordes et al., 2013), ComplEx (Trouillon et al., 2016), DistMult (Yang et al., 2015), and RotatE (Sun et al., 2019) learn vectors for each node and score candidate drug-disease pairs. The catch is that all of them need the complete graph at both training and inference time, and assembling a graph like Hetionet (47,000+ nodes, 2.2M edges, 24 relationship types) is labor-intensive and inevitably reflects its curators' choices and blind spots.

We wondered whether the same information might already be latent in the biomedical literature. PubMedBERT (Gu et al., 2021), now rebranded as MSR BiomedBERT, captures rich biomedical knowledge through distributional patterns in text (Shtar et al., 2022). A drug like carbamazepine appears in thousands of abstracts that also mention seizures and sodium channel blockade; after enough co-occurrences, the model places it close to these concepts in embedding space. The result, we hypothesized, functions as an implicit knowledge graph that updates automatically and does not require manual curation. Prior work has mined text for relations and co-occurrences (Wei et al., 2019; Andronis et al., 2011), combined text with structure and expression in multi-modal architectures (Luo et al., 2024, 2023), and trained graph-foundation models on larger graphs (Huang et al., 2024); published numbers are not directly comparable because each uses different splits (HRGAT (Yu et al., 2021) reports AUROC 0.912 on Hetionet under a setup we could not reproduce). A controlled evaluation of text embeddings *alone*, with classifiers and folds held fixed against KG baselines re-trained from scratch,

Proceedings of the 25th Workshop on Biomedical Language Processing (BioNLP 2026), pages 135–140

has not, as far as we can tell, been done.

We designed a framework that can train on any subset of three modalities (text, molecular fingerprints, gene expression) and tested all seven combinations on the Hetionet benchmark and an epilepsy drug ranking task. The results were, in several respects, not what we anticipated:

- Combining text with molecular or gene expression features through cross-attention consistently makes performance *worse*, not better, despite a gated fusion mechanism that should, in theory, learn to ignore unhelpful modalities. We trace this to residual-path contamination: the strong modality is forced to absorb noise from the weak ones.
- Text embeddings alone reach AUROC 0.910 on Hetionet, above all four KG baselines we re-trained on the same data under the same protocol (the best, RotatE, 0.854). This holds for both our neural architecture and a Random Forest baseline; the predictive signal lives in PubMedBERT itself, not in any particular downstream model.
- Top-ranked epilepsy candidates mix genuinely promising compounds (sirolimus) and known proconvulsants (amoxapine, flumazenil), exposing a basic limitation of any approach grounded in literature co-occurrence.

2 Methods

2.1 Feature Representations

We query the NCBI E-utilities API for up to 20 PubMed abstracts per entity (drug or disease), run each through PubMedBERT (Gu et al., 2021) to obtain the `[CLS]` token vector (768d), and take the mean. Because PubMedBERT was pretrained on a large fraction of the biomedical literature, these vectors carry information about a compound's pharmacology, therapeutic context, and the biological processes it has been studied in connection with. Compounds with no indexed abstracts (3.4%) receive zero vectors and are retained in the dataset but have no text signal. Molecular features are 2048-bit ECFP4 fingerprints (Rogers and Hahn, 2010) generated with RDKit (RDKit, 2024), capturing chemical substructure. Gene expression features come from LINCS L1000 (Subramanian et al., 2017): 978-dimensional vectors recording how a compound perturbs landmark gene expression in treated cell lines.

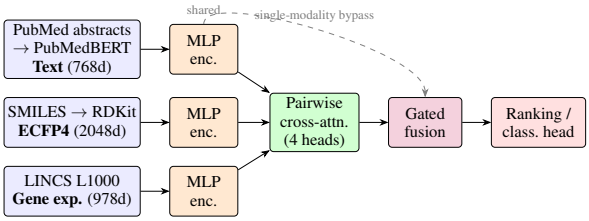

Figure 1: Model architecture. Per-modality MLP encoders project text (PubMedBERT `[CLS]`, 768d), molecular (ECFP4, 2048d), and gene expression (LINCS, 978d) inputs into a shared embedding space (128d on Hetionet, 256d on epilepsy); pairwise cross-attention (4 heads, Eq. (1)) and gated fusion combine them before a two-layer head. Single-modality runs skip both blocks (dashed).

2.2 Architecture

Figure 1 gives the overall architecture. Each modality passes through a two-layer MLP encoder (hidden 512, LayerNorm (Ba et al., 2016), ReLU, dropout) into a shared embedding space (128d on Hetionet, 256d on epilepsy). In multi-modal settings, pairwise cross-attention (Vaswani et al., 2017; Lu et al., 2019) with 4 heads updates each representation from the others:

$$\hat{\mathbf{h}}_{m_i \leftarrow m_j} = \mathrm{LN}(\mathbf{h}_{m_i} + \mathrm{MHA}(\mathbf{h}_{m_i}, \mathbf{h}_{m_j}, \mathbf{h}_{m_j})) \quad (1)$$

With three modalities, the two cross-attended versions per modality are averaged. Gated fusion (Arevalo et al., 2017) computes per-sample softmax weights over ℓ_2-normalized vectors, and a two-layer head (256, 128) outputs ranking scores or classification logits.

2.3 Datasets and Training

Epilepsy Drug Repurposing. From LINCS L1000 Phase 2 (Subramanian et al., 2017), we kept the 1,713 compounds with both gene expression profiles and ECFP4 fingerprints (PubMed coverage: 96.6%). Of these, 19 are established AEDs, including sodium channel blockers (carbamazepine, phenytoin), GABAergic agents (gabapentin), and multi-mechanism drugs (topiramate). We split compound-disjoint (1,027/343/343) with stratified positives. Training uses $\mathcal{L} = 0.7 \cdot \mathcal{L}_{\text{LambdaNDCG}} + 0.3 \cdot \mathcal{L}_{\text{BCE}}$ (Burges et al., 2006) with Adam (Kingma and Ba, 2015) (lr 5×10^{-4}, batch 128, 100 epochs, early stopping on validation NDCG@50). We report NDCG@50, which covers the top $\sim$15% of test compounds. All experiments use 3 random seeds.

Hetionet Benchmark. Hetionet (Himmelstein et al., 2017) has 755 curated drug-disease edges. We sampled negatives at 1:1 (1,509 pairs total) and trained with BCEWithLogitsLoss under 10-fold CV (embedding dim 128, early stopping on validation AUROC). For the KG baselines we trained TransE, ComplEx, DistMult, and RotatE on the full Hetionet graph with PyKEEN (Ali et al., 2021) (dim 128, margin-ranking loss, Adam, 200 epochs with early stopping), then fed entity embeddings through the same downstream classifiers (Random Forest (Breiman, 2001), Logistic Regression) on the same folds, isolating the representation. Gene expression is not added to this benchmark because LINCS L1000 covers only a fraction of Hetionet's 1,538 compound entities; we restrict Hetionet modality fusion to text and molecular, and use the epilepsy dataset for the full seven-way ablation. All runs use 3 seeds; we report mean ± std.

Method	Features	AUROC	AUPRC
KG embedding baselines (trained on full Hetionet)			
TransE	KG Embedding	0.822 ± 0.018	0.809 ± 0.030
DistMult	KG Embedding	0.833 ± 0.017	0.818 ± 0.030
ComplEx	KG Embedding	0.837 ± 0.019	0.828 ± 0.031
RotatE	KG Embedding	0.854 ± 0.019	0.849 ± 0.022
Our baselines (no graph)			
RF	Molecular	0.496 ± 0.047	0.342 ± 0.053
RF	Text	0.880 ± 0.015	0.875 ± 0.025
RF	Text + Mol	0.883 ± 0.015	0.708 ± 0.029
Our neural model (cross-attention, no graph)			
Ours	Molecular	0.581 ± 0.026	0.338 ± 0.046
Ours	Text	$\mathbf{0.910 \pm 0.017}$	0.727 ± 0.057
Ours	Text + Mol	0.817 ± 0.030	0.584 ± 0.063

Table 1: Drug-disease link prediction on Hetionet. KG methods were trained on the full graph using PyKEEN; all methods evaluated under the same 10-fold CV protocol on the same 1,509 drug-disease pairs. Gene-expression rows are omitted: LINCS L1000 covers only a fraction of Hetionet's 1,538 compound entities, so an apples-to-apples row is not available; the full seven-way modality ablation is run on the LINCS-aligned epilepsy dataset (Table 2).

3 Results and Analysis

3.1 Hetionet Benchmark

Table 1 summarizes the Hetionet results. Every row is run under the same 10-fold protocol on the same 1,509 pairs; we exclude published numbers from prior papers (e.g. HRGAT 0.912) because those use different splits. With text features only, our neural model reaches AUROC 0.910 ± 0.017. The KG baselines score lower in AUROC across the board: RotatE 0.854 ± 0.019, ComplEx 0.837 ± 0.019, DistMult 0.833 ± 0.017, TransE 0.822 ± 0.018. Only the entity-vector source differs. AUPRC is more nuanced: RF+Text (0.875) and RotatE (0.849) both exceed the neural model's 0.727. The neural model leads in AUROC but is less calibrated at the high-precision end; for a 1:1 balanced task the two metrics mostly agree, but under harder negative ratios the AUPRC gap between text and the strongest KG method is narrower than AUROC suggests.

A plain Random Forest on the same PubMedBERT vectors scores 0.880 ± 0.015. The embeddings, not our neural model, are the main source of predictive power. To rule out a dimensionality artifact (PubMedBERT 768d vs. KG 128d) we ran PCA-reduced text embeddings at 128d: AUROC drops to 0.818, below RotatE (0.854). The full 768d is needed; the higher dimensionality reflects PubMedBERT's richer training signal (billions of tokens vs. a single graph). Molecular fingerprints alone are near chance for RF (AUROC 0.496) and only modestly above for the neural model (0.581): structure-to-indication requires a chain (drug→target→pathway→disease) that fingerprints do not encode, but PubMedBERT picks up much of it from co-occurrence patterns in text.

3.2 Cross-modal Contamination

Adding molecular features to text makes things worse, not better. AUROC drops from 0.910 to 0.817 in the neural model, while Random Forest barely moves (0.883 vs. 0.880). The residual connection in Eq. (1) blends each modality with the others regardless of quality, so when one modality is mostly noise, the stronger one gets contaminated; tree-based models simply ignore uninformative features.

3.3 Epilepsy Case Study

Table 2 gives the full seven-way modality ablation for the epilepsy ranking task. Text alone hits AUROC 0.980 ± 0.006 and the highest P@10 (0.267) and NDCG@50 (0.680) of any neural configuration; the precision and ranking metrics track the AUROC ordering, which is reassuring given how few positives there are. Molecular fingerprints land at AUROC 0.662 and gene expression at 0.512, both close to what you would get from random ordering, and both produce P@10 of zero. The bi-modal results are worth a closer

Config	P@10	P@50	NDCG@50	AUROC
Single modality				
T	0.267 ± 0.094	0.080 ± 0.000	0.680 ± 0.113	0.980 ± 0.006
M	0.000 ± 0.000	0.040 ± 0.016	0.169 ± 0.083	0.662 ± 0.109
G	0.000 ± 0.000	0.013 ± 0.019	0.049 ± 0.070	0.512 ± 0.088
Bi-modal				
TM	0.067 ± 0.047	0.053 ± 0.025	0.282 ± 0.134	0.891 ± 0.061
TG	0.200 ± 0.141	0.073 ± 0.009	0.595 ± 0.195	0.932 ± 0.040
MG	0.033 ± 0.047	0.007 ± 0.009	0.038 ± 0.053	0.507 ± 0.091
Tri-modal				
TMG	0.100 ± 0.141	0.053 ± 0.025	0.414 ± 0.323	0.779 ± 0.140
Baselines				
RF + T	0.400	–	0.895	0.994
RF + M	0.000	–	0.049	0.603
RF + G	0.000	–	0.046	0.518
RF + TMG	0.300	–	0.882	0.989
CMap	0.000	–	–	0.437

Table 2: Ablation study on epilepsy drug repurposing dataset. T = PubMedBERT text, M = ECFP4 molecular, G = LINCS gene expression. Results are mean $\pm$ std over 3 random seeds.

look. Text plus gene expression (TG: 0.932 AUROC, P@10 0.200) holds up better than text plus molecular (TM: 0.891, P@10 0.067). Gene expression profiles record cellular-level responses to a compound, so they carry at least some biological signal that fingerprints lack. Despite this, text alone still outperforms both combinations on every ranking metric. The tri-modal configuration drops to AUROC 0.779 and P@10 0.100, substantially worse than text alone.

Random Forest with text outperforms everything else here (AUROC 0.994, NDCG@50 0.895), ahead of the neural model (0.980 and 0.680). RF on molecular is only weakly predictive (AUROC 0.603) and on gene expression is at chance (AUROC 0.518; Table 2), confirming the modality asymmetry holds independent of architecture. With only 19 AEDs among 1,713 compounds and about 4 in the test fold, RF's built-in regularization matters more than cross-attention's flexibility. A Connectivity Map baseline (Lamb et al., 2006) that correlates compound expression profiles with an epilepsy disease signature came in below chance (AUROC 0.437). Being "graph-free" is not by itself sufficient; the particular information in PubMedBERT embeddings is doing real work.

3.4 Biological Plausibility

All 4 AEDs in the test set end up ranked in the top 11 out of 343 compounds (rufinamide at rank 1, phenytoin at rank 6), consistent across 3 seeds; the hypergeometric probability of this by chance is $\approx 5.8 \times 10^{-7}$. Sirolimus (rank 15 in the RF+T model) inhibits mTOR, and its analog everolimus was approved for treating seizures associated with tuberous sclerosis complex (French et al., 2016), a condition driven by mTOR pathway hyperactivation. Clomipramine (rank 7 in our neural model) is a tricyclic antidepressant that potently inhibits serotonin reuptake. There is clinical evidence that serotonergic enhancement can reduce seizure frequency (Favale et al., 2003), so this prediction has some mechanistic grounding.

The false positives tell us more. Amoxapine at rank 2 lowers seizure thresholds (Pisani et al., 1999), and flumazenil at rank 4 is a benzodiazepine antagonist used to *provoke* seizures diagnostically (Schulze-Bonhage and Elger, 2000). Why does the model rank them so highly? Because the seizure literature discusses them extensively, just as it discusses actual anticonvulsants. PubMedBERT's embeddings reflect how strongly a compound is associated with a disease area, but carry no information about the polarity of that association. A drug that treats epilepsy and one that triggers seizures both end up near epilepsy-related concepts. This is the central limitation of any repurposing approach built purely on literature.

4 Conclusion

PubMedBERT text embeddings, without fine-tuning, outperform four re-trained KG baselines on Hetionet under a matched protocol; both neural and tree-based classifiers agree. Cross-attention fusion of unequal-quality modalities hurts the strong one via residual contamination.

Limitations. The comparison is asymmetric: PubMedBERT was pretrained on the PubMed literature Hetionet was curated from, so this is best read as "text-with-literature-supervision vs. graph-only." 1:1 negative sampling may flatter AUROC, and gaps to the strongest KG method may close under harder ratios; with only 19 epilepsy positives ($\sim$4 per fold), P@k and NDCG carry wide intervals (the top-11 ranking, however, has $p \sim 6 \times 10^{-7}$).

Future Work. Two natural follow-ups: (i) a unified-protocol head-to-head against KG-BERT (Yao et al., 2019), TxGNN (Huang et al., 2024), and HRGAT; (ii) re-training KG baselines on a literature-augmented Hetionet.

References

Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Sahand Sharifzadeh, Volker Tresp, and Jens Lehmann. 2021. PyKEEN 1.0: A Python library for training and evaluating knowledge graph embeddings. *Journal of Machine Learning Research*, 22(82):1–6.

Christos Andronis, Anuj Sharma, Vassilis Virvilis, Spyros Deftereos, and Aris Persidis. 2011. Literature mining, ontologies and information visualization for drug repurposing. *Briefings in Bioinformatics*, 12(4):357–368.

John Arevalo, Thamar Solorio, Manuel Montes-y Gómez, and Fabio A González. 2017. Gated multimodal units for information fusion. In *International Conference on Learning Representations Workshop*.

Ted T Ashburn and Karl B Thor. 2004. Drug repositioning: Identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8):673–683.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26.

Leo Breiman. 2001. Random forests. *Machine Learning*, 45:5–32.

Christopher J C Burges, Robert Ragno, and Quoc V Le. 2006. Learning to rank with nonsmooth cost functions. In *Advances in Neural Information Processing Systems*, volume 19.

Emilio Favale, Daniela Audenino, Leonardo Cocito, and Carmelo Albano. 2003. The anticonvulsant effect of citalopram as an indirect evidence of serotonergic impairment in human epileptogenesis. *Seizure*, 12(5):316–318.

Jacqueline A French, John A Lawson, Zuhal Yapici, Hiroko Ikeda, Tilman Polster, Rima Nabbout, Paolo Curatolo, Petrus J de Vries, Dennis J Dlugos, Nicola Berkowitz, and 1 others. 2016. Adjunctive everolimus therapy for treatment-resistant focal-onset seizures associated with tuberous sclerosis (EXIST-3): A phase 3, randomised, double-blind, placebo-controlled study. *The Lancet*, 388(10056):2153–2163.

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 3(1):1–23.

Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. 2017. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, 6:e26726.

Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish N Nadkarni, Benjamin S Glicksberg, Nils Gehlenborg, and Marinka Zitnik. 2024. A foundation model for clinician-centered drug repurposing. *Nature Medicine*, 30:3601–3613.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*.

Justin Lamb, Emily D Crawford, David Peck, Joshua W Modell, Irene C Blat, Matthew J Wrobel, Jim Lerner, Jean-Philippe Brunet, Aravind Subramanian, Kenneth N Ross, and 1 others. 2006. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935.

Wolfgang Löscher, Heidrun Potschka, Sanjay M. Sisodiya, and Annamaria Vezzani. 2020. Drug resistance in epilepsy: Clinical impact, potential mechanisms, and new innovative treatment options. *Pharmacological Reviews*, 72(3):606–638.

Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In *Advances in Neural Information Processing Systems*, volume 32.

Yizhen Luo, Xing Yi Liu, Kai Yang, Kui Huang, Massimo Hong, Jiahuan Zhang, Yushuai Wu, and Zaiqing Nie. 2024. Toward unified AI drug discovery with multimodal knowledge. *Health Data Science*, 4:0113.

Yizhen Luo, Kai Yang, Massimo Hong, Xing Yi Liu, and Zaiqing Nie. 2023. MolFM: A multimodal molecular foundation model. *arXiv preprint arXiv:2307.09484*.

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33.

F Pisani, E Spina, and G Oteri. 1999. Antidepressant drugs and seizure susceptibility: From in vitro data to clinical practice. *Epilepsia*, 40(Suppl 10):S48–S56.

Sudeep Pushpakom, Francesco Iorio, Patrick A Eyers, K Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, and 1 others. 2019. Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1):41–58.

RDKit. 2024. RDKit: Open-source cheminformatics. https://www.rdkit.org.

David Rogers and Mathew Hahn. 2010. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754.

A Schulze-Bonhage and C E Elger. 2000. Induction of partial epileptic seizures by flumazenil. *Epilepsia*, 41(2):186–192.

Guy Shtar, Asnat Greenstein-Messica, Eyal Mazuz, Lior Rokach, and Bracha Shapira. 2022. Predicting drug characteristics using biomedical text embedding. *BMC Bioinformatics*, 23(1):526.

Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, and 1 others. 2017. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171(6):1437–1452.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge graph embedding by relational rotation in complex space. In *Proceedings of the 7th International Conference on Learning Representations*.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 2071–2080.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.

Chih-Hsuan Wei, Alexis Allot, Robert Leaman, and Zhiyong Lu. 2019. PubTator Central: Automated concept annotation for biomedical full text articles. *Nucleic Acids Research*, 47(W1):W587–W593.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the 3rd International Conference on Learning Representations*.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for knowledge graph completion. *arXiv preprint arXiv:1909.03193*.

Zhouxin Yu, Feng Huang, Xiaohan Zhao, Wenjie Xiao, and Wen Zhang. 2021. Predicting drug-disease associations via heterogeneous graph attention networks. *Briefings in Bioinformatics*, 22(4):bbaa244.