

# Mining Heterogenous Relationships from Pubmed Abstracts Using Weak Supervision

This manuscript ([permalink](#)) was automatically generated from [greenelab/text\\_mined\\_hetnet\\_manuscript@77307b4](#) on July 29, 2019.

## Authors

---

- **David N. Nicholson**

 [0000-0003-0002-5761](#) ·  [danich1](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552

- **Daniel S. Himmelstein**

 [0000-0002-3012-7446](#) ·  [dhimmel](#) ·  [dhimmel](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552

- **Casey S. Greene**

 [0000-0001-8713-9213](#) ·  [cgreene](#) ·  [GreeneScientist](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552 and R01 HG010067

## Abstract

---

Knowledge bases support multiple research efforts including providing contextual information for biomedical entities, constructing networks, and supporting the interpretation of high-throughput analyses. Some knowledge bases are automatically constructed, but most are populated via some form of manual curation. Manual curation is time consuming and difficult to scale in the context of an increasing publication rate. A recently described “data programming” paradigm seeks to circumvent this arduous process by combining distant supervision with simple rules and heuristics written as labeling functions that can be automatically applied to inputs. Unfortunately writing useful label functions requires substantial error analysis and is a non trivial task: in early efforts to use data programming we found that producing each label function could take a few days. Producing a biomedical knowledge base with multiple node and edge types could take hundreds or thousands of label functions. In this paper we sought to evaluate the extent to which label functions could be re-used across edge types. We used a subset of Hetionet v1 that centered on disease, compound, and gene nodes to evaluate this approach. We compare a baseline distant supervision model with the same distant supervision resources added to edge-type-specific label functions, edge-type-mismatch label functions, and all label functions. We confirmed that adding additional edge-type-specific label functions improves performance. We also found that adding one or a few edge-type-mismatch label functions also nearly always improves performance. Adding a large number of edge-type-mismatch label functions produces more variable performance that depends on the edge type being predicted and the edge type that is the source of the label function. Lastly, we show that this approach, even on this subgraph of Hetionet, could certain novel edges to Hetionet v1 with high confidence. We expect that its use in practice would include additional filtering and scoring methods which would further enhance precision.

## Introduction

---

Knowledge bases are important resources that hold complex structured and unstructured information. These resources have been used in important tasks such as network analysis for drug repurposing discovery [1,2,3] or as a source of training labels for text mining systems [4,5,6]. Populating knowledge bases often requires highly-trained scientists to read biomedical literature and summarize the results [7]. This manual curation process requires a significant amount of effort and time: in 2007 researchers estimated that filling in the missing annotations at that point would require approximately 8.4 years [8]). The rate of publications has continued to increase exponentially [9]. This has been recognized as a considerable challenge and leads to gaps in knowledge bases [8]. Relationship extraction has been studied as a solution towards handling this problem [7]. This process consists of creating a machine learning system to automatically scan and extract relationships from textual sources. Machine learning methods often leverage a large corpus of well-labeled training data, which still requires manual curation. Distant supervision is one technique to sidestep the requirement of well-annotated sentences: with distant supervision one makes the assumption that that all sentences containing an entity pair found in a selected database provide evidence for a relationship [4]. Distant supervision provides many labeled examples; however it is accompanied by a decrease in the quality of the labels. Ratner et al. [10] recently introduced “data programming” as a solution. Data programming combines distant supervision with the automated labeling of text using hand-written label functions. The distant supervision sources and label functions are integrated using a noise aware generative model, which is used to produce training labels. Combining distant supervision with label functions can dramatically reduce the time required to acquire sufficient training data. However, constructing a knowledge base of heterogeneous relationships through this framework still requires tens of hand-written label functions for each relationship type. Writing useful label functions requires significant error analysis, which can be a time-consuming process.

In this paper, we aim to address the question: to what extent can label functions be re-used across different relationship types? We hypothesized that sentences describing one relationship type may share information in the form of keywords or sentence structure with sentences that indicate other relationship types. We designed a series of experiments to determine the extent to which label function re-use enhanced performance over distant supervision alone. We examine relationships that indicate similar types of physical interactions (i.e., gene-binds-gene and compound-binds-gene) as well as different types (i.e., disease-associates-gene and compound-treats-disease). The re-use of label functions could dramatically reduce the number required to generate and update a heterogeneous knowledge graph.

## Recent Work

---

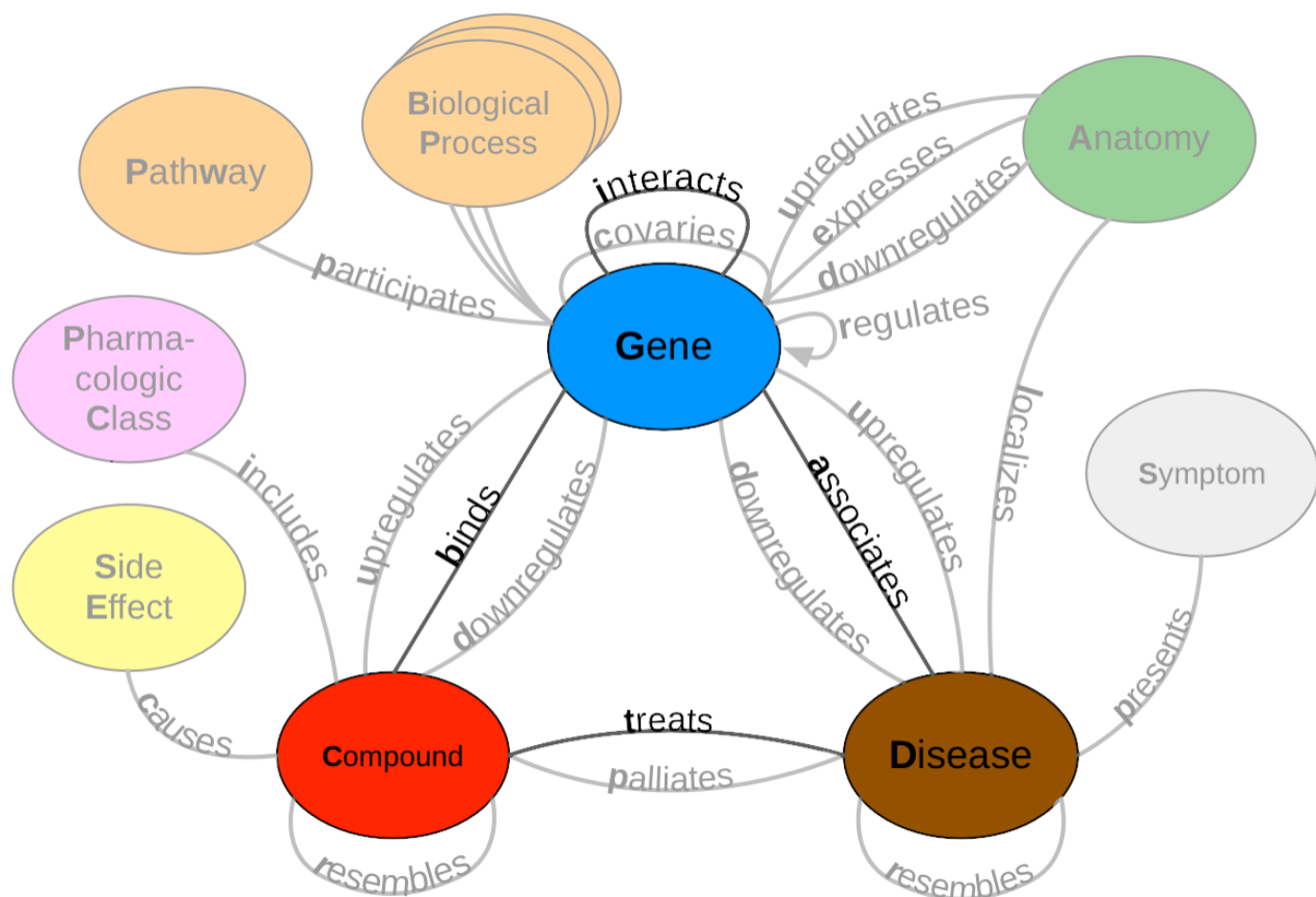
Talk about what has been done in the field in regards to text mining and knowledge base integration

## Materials and Methods

---

### Hetionet

Hetionet [\[3\]](#) is a large heterogeneous network that contains pharmacological and biological information. This network depicts information in the form of nodes and edges of different types: nodes that represent biological and pharmacological entities and edges which represent relationships between entities. Hetionet v1.0 contains 47,031 nodes with 11 different data types and 2,250,197 edges that represent 24 different relationship types (Figure [1](#)). Edges in Hetionet were obtained from open databases, such as the GWAS Catalog [\[11\]](#) and DrugBank [\[12\]](#). For this project, we analyzed performance over a subset of the Hetionet relationship types: disease associates with a gene (DaG), compound binds to a gene (CbG), gene interacts with gene (GiG) and compound treating a disease (CtD).



**Figure 1:** A metagraph (schema) of Hetionet where pharmacological, biological and disease entities are represented as nodes and the relationships between them are represented as edges. This project only focuses on the information shown in bold; however, we can extend this work to incorporate the faded out information as well.

## Dataset

We used PubTator [13] as input to our analysis. PubTator provides MEDLINE abstracts that have been annotated with well-established entity recognition tools including DNorm [14] for disease mentions, GeneTUKit [15] for gene mentions, Gnorm [16] for gene normalizations and a dictionary based look system for compound mentions [17]. We downloaded PubTator on June 30, 2017, at which point it contained 10,775,748 abstracts. Then we filtered out mention tags that were not contained in hetionet. We used the Stanford CoreNLP parser [18] to tag parts of speech and generate dependency trees. We extracted sentences with two or more mentions, termed candidate sentences. Each candidate sentence was stratified by co-mention pair to produce a training set, tuning set and a testing set (shown in Table 1). Each unique co-mention pair is sorted into four categories: (1) in hetionet and has sentences, (2) in hetionet and doesn't have sentences, (3) not in hetionet and does have sentences and (4) not in hetionet and doesn't have sentences. Within these four categories each pair receives their own individual partition rank (continuous number between 0 and 1). Any rank lower than 0.7 is sorted into training set, while any rank greater than 0.7 and lower than 0.9 is assigned to tuning set. The rest of the pairs with a rank greater than or equal to 0.9 is assigned to the test set. Sentences that contain more than one co-mention pair are treated as multiple individual candidates. We hand labeled five hundred to a thousand candidate sentences of each relationship to obtain a ground truth set (Table 1, dataset).

**Table 1:** Statistics of Candidate Sentences. We sorted each candidate sentence into a training, tuning and testing set. Numbers in parentheses show the number of positives and negatives that resulted from the hand-labeling process.

Relationship	Train	Tune	Test
--------------	-------	------	------

Relationship	Train	Tune	Test
Disease Associates Gene	2.35 M	31K (397+, 603-)	313K (351+, 649-)
Compound Binds Gene	1.7M	468K (37+, 463-)	227k (31+, 469-)
Compound Treats Disease	1.013M	96K (96+, 404-)	32K (112+, 388-)
Gene Interacts Gene	12.6M	1.056M (60+, 440-)	257K (76+, 424-)

## Label Functions for Annotating Sentences

A common challenge in natural language processing is having too few ground truth annotations, even when textual data are abundant. Data programming circumvents this issue by quickly annotating large datasets by using multiple noisy signals emitted by label functions [10]. Label functions are simple pythonic functions that emit: a positive label (1), a negative label (-1) or abstain from emitting a label (0). We combine these functions using a generative model to output a single annotation, which is a consensus probability score bounded between 0 (low chance of mentioning a relationship) and 1 (high chance of mentioning a relationship). We used these annotations to train a discriminator model that makes the final classification step. Our label functions fall into three categories: databases, text patterns and domain heuristics. We provide examples for the categories, described below, using the following candidate sentence: “PTK6 may be a novel therapeutic target for pancreatic cancer.”

**Databases:** These label functions incorporate existing databases to generate a signal, as seen in distant supervision [4]. These functions detect if a candidate sentence’s co-mention pair is present in a given database. If the pair is present, emit a positive label and abstain otherwise. If the pair isn’t present in any existing database, then a separate label function will emit a negative label. We use a separate label function to prevent the label imbalance problem. This problem occurs when candidates, that scarcely appear in databases, are drowned out by negative labels. The multitude of negative labels increases the likelihood of misclassification when training the generative model.

$$\Lambda_{DB}(D, G) = \begin{cases} 1 & (D, G) \in DB \\ 0 & otherwise \end{cases}$$

$$\Lambda_{\neg DB}(D, G) = \begin{cases} -1 & (D, G) \notin DB \\ 0 & otherwise \end{cases}$$

**Text Patterns:** These label functions are designed to use keywords and sentence context to generate a signal. For example, a label function could focus on the number of words between two mentions or focus on the grammatical structure of a sentence. These functions emit a positive or negative label depending on the situation. In general, those focused on keywords emit positives and those focused on negation emit negatives.

$$\Lambda_{TP}(D, G) = \begin{cases} 1 & "target" \in Candidate\ Sentence \\ 0 & otherwise \end{cases}$$

$$\Lambda_{TP}(D, G) = \begin{cases} -1 & "VB" \notin pos\_tags(Candidate\ Sentence) \\ 0 & otherwise \end{cases}$$

**Domain Heuristics:** These label functions use the other experiment results to generate a signal. For this category, we used dependency path cluster themes generated by Percha et al [19]. If a candidate sentence’s dependency path belongs to a previously generated cluster, then the label function will emit a positive label and abstain otherwise.

$$\Lambda_{DH}(D, G) = \begin{cases} 1 & \text{Candidate Sentence} \in \text{Cluster Theme} \\ 0 & \text{otherwise} \end{cases}$$

Roughly half of our label functions are based on text patterns, while the others are distributed across the databases and domain heuristics (Table 2).

**Table 2:** The distribution of each label function per relationship.

Relationship	Databases (DB)	Text Patterns (TP)	Domain Heuristics (DH)
Disease associates Gene (DaG)	7	20	10
Compound treats Disease (CtD)	3	15	7
Compound binds Gene (CbG)	9	13	7
Gene interacts Gene (GiG)	9	20	8

## Training Models

### Generative Model

The generative model is a core part of this automatic annotation framework. It integrates multiple signals emitted by label functions and assigns a training class to each candidate sentence. This model assigns training classes by estimating the joint probability distribution of the latent true class ( $Y$ ) and label function signals ( $\Lambda$ ),  $P(\Lambda, Y)$ . Assuming each label function is conditionally independent, the joint distribution is defined as follows:

$$P(\Lambda, Y) = \frac{\exp(\sum_{i=1}^m \theta^T F_i(\Lambda, y))}{\sum_{\Lambda'} \sum_{y'} \exp(\sum_{i=1}^m \theta^T F_i(\Lambda', y'))}$$

where  $m$  is the number of candidate sentences,  $F$  is the vector of summary statistics and  $\theta$  is a vector of weights for each summary statistic. The summary statistics used by the generative model are as follows:

$$F_{i,j}^{Lab}(\Lambda, Y) = \mathbb{1}\{\Lambda_{i,j} \neq 0\}$$

$$F_{i,j}^{Acc}(\Lambda, Y) = \mathbb{1}\{\Lambda_{i,j} = y_{i,j}\}$$

*Lab* is the label function's propensity (the frequency of a label function emitting a signal). *Acc* is the individual label function's accuracy given the training class. This model optimizes the weights ( $\theta$ ) by minimizing the negative log likelihood:

$$\hat{\theta} = \underset{\theta}{argmin} - \sum_{\Lambda} \log \sum_Y P(\Lambda, Y)$$

In the framework we used predictions from the generative model,  $\hat{Y} = P(Y | \Lambda)$ , as training classes for our dataset [20,21].

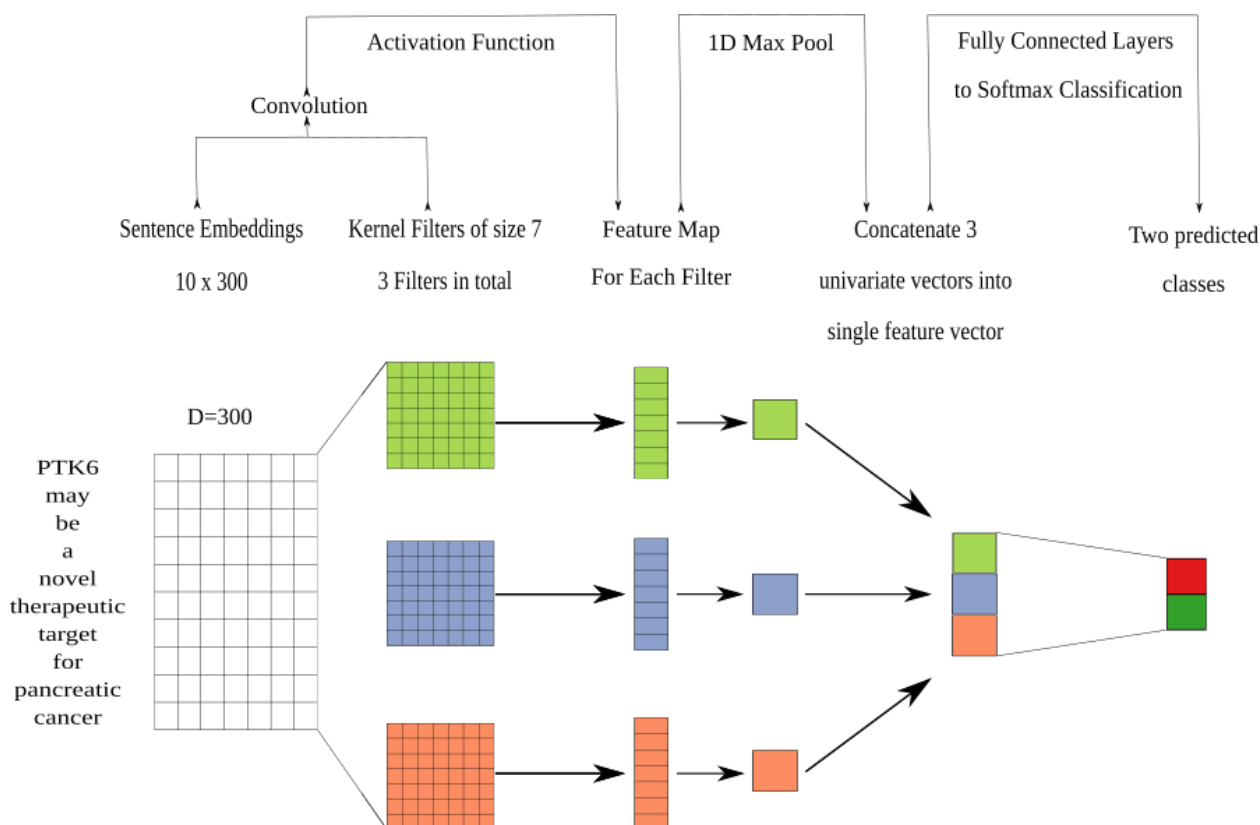
### Word Embeddings

Word embeddings are representations that map individual words to real valued vectors of user-specified dimensions. These embeddings have been shown to capture the semantic and syntactic information between words [22]. Using all candidate sentences for each individual relationship pair,

we trained facebook's fastText [23] to generate word embeddings. The fastText model uses a skipgram model [24] that aims to predict the context given a candidate word and pairs the model with a novel scoring function that treats each word as a bag of character n-grams. We trained this model for 20 epochs using a window size of 2 and generated 300-dimensional word embeddings. We use the optimized word embeddings to train a discriminative model.

## Discriminative Model

The discriminative model is a neural network, which we train to predict labels from the generative model. The expectation is that the discriminative model can learn more complete features of the text than the label functions that are used in the generative model. We used a convolutional neural network with multiple filters as our discriminative model. This network uses multiple filters with fixed widths of 300 dimensions and a fixed height of 7 (Figure 2), because this height provided the best performance in terms of relationship classification [25]. We trained this model for 20 epochs using the adam optimizer [26] with a learning rate of 0.001. This optimizer used pytorch's default parameter settings. We added a L2 penalty on the network weights to prevent overfitting. Lastly, we added a dropout layer ( $p=0.25$ ) between the fully connected layer and the softmax layer.



**Figure 2:** The architecture of the discriminative model is a convolutional neural network. We perform a convolution step using multiple filters. These filters generate a feature map that is sent into a maximum pooling layer. This layer extracts the largest feature in each of these maps. The extracted features are concatenated into a singular vector that is passed into a fully connected network. The fully connected network has 300 neurons for the first layer, 100 neurons for the second layer and 50 neurons for the last layer. From the fully connected network the last step is to generate predictions using the softmax layer.

## Calibration of the Discriminative Model

Often many tasks require a machine learning model to output reliable probability predictions. A model is well calibrated if the probabilities emitted from the model match the observed probabilities: a well-calibrated model that assigns a class label with 80% probability should have that class appear 80% of the time. Deep neural network models can often be poorly calibrated [27,28]. These models are usually over-confident in their predictions. As a result, we calibrated our convolutional neural



network using temperature scaling. Temperature scaling uses a parameter  $T$  to scale each value of the logit vector ( $z$ ) before being passed into the softmax (SM) function.

$$\sigma_{SM}\left(\frac{z_i}{T}\right) = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_i \exp\left(\frac{z_i}{T}\right)}$$

We found the optimal  $T$  by minimizing the negative log likelihood (NLL) of a held out validation set. The benefit of using this method is the model becomes more reliable and the accuracy of the model doesn't change [27].

## Experimental Design

Being able to re-use label functions across edge types would substantially reduce the number of label functions required to extract multiple relationship types from biomedical literature. We first established a baseline by training a generative model using only distant supervision label functions designed for the target edge type. As an example, for the gene-interacts-gene edge type we used label functions that returned a 1 if the pair of genes were included in the Human Interaction database [29], the iRefIndex database [30] or in the Incomplete Interactome database [31]. Then we compared models that also included text and domain-heuristic label functions. Using a sampling with replacement approach, we sampled these text and domain-heuristic label functions separately within edge types, across edge types, and from a pool of all label functions. We compared within-edge-type performance to across-edge-type and all-edge-type performance. For each edge type we sampled a fixed number of label functions consisting of five evenly-spaced numbers between one and the total number of possible label functions. We repeated this sampling process 50 times for each point. We evaluated both generative and discriminative models at each point, and we report performance of each in terms of the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR).

### Adding Random Noise to Generative Model

We discovered in the course of this work that adding a single label function from a mismatched type would often improve the performance of the generative model (see Results). We designed an experiment to test whether adding a noisy label function also increased performance. This label function emitted a positive or negative label at varying frequencies, which were evenly spaced from zero to one. Zero is the same as distant supervision alone. We trained the generative model with these label functions added and report results in terms of AUROC and AUPR.

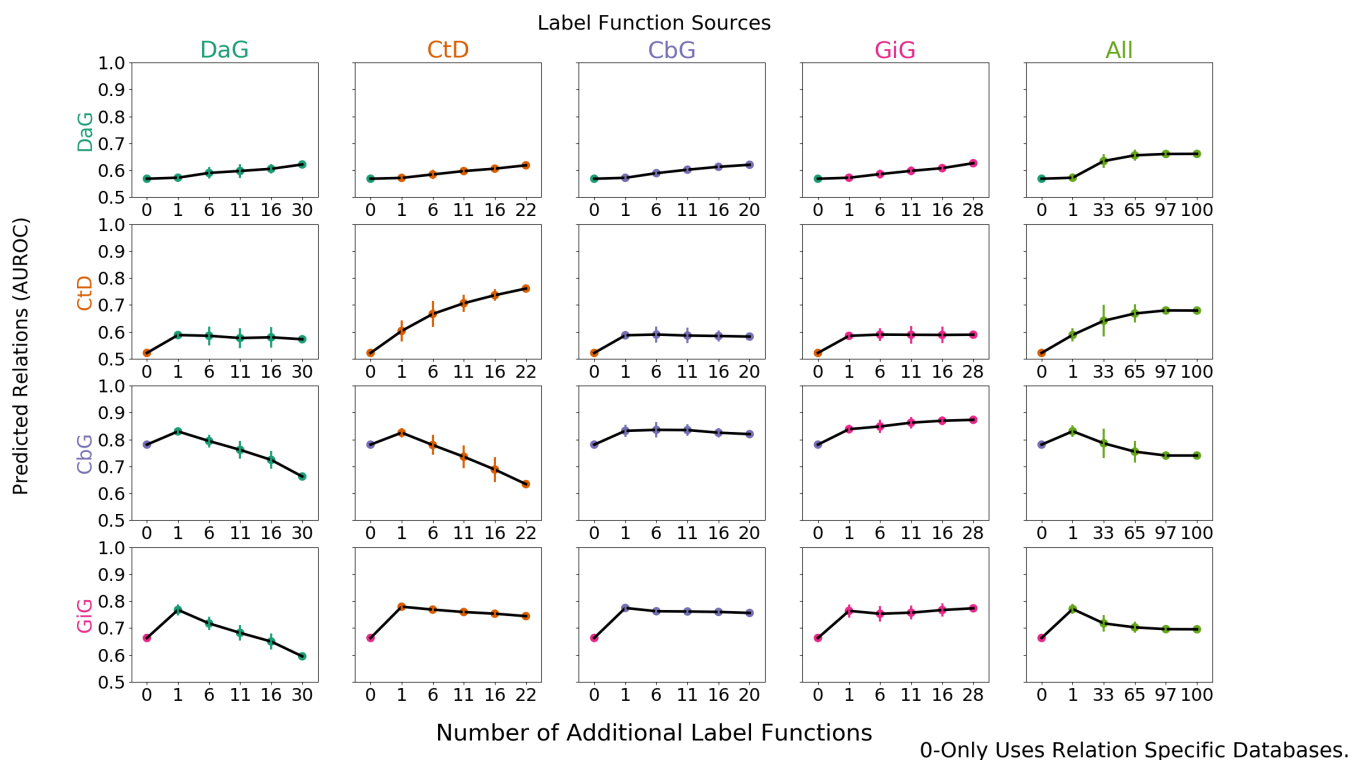
## Results

---

### Generative Model Using Randomly Sampled Label Functions



### Label Sampling Generative Model Assessment (Test Set)



**Figure 3:** Grid of Area Under the Receiver Operating Curve (AUROC) scores for each generative model trained on randomly sampled label functions. The rows depict the relationship each model is trying to predict and the columns are the relationship specific sources each label function is sampled from. For example, the top-left most square depicts the generative model predicting Disease associates Gene (DaG) sentences, while randomly sampling label functions designed to predict the DaG relationship. The square towards the right depicts the generative model predicting DaG sentences, while randomly sampling label functions designed to predict the Compound treats Disease (CtD) relationship. This pattern continues filling out the rest of the grid. The last most column consists of pooling every relationship specific label function and proceeding as above.

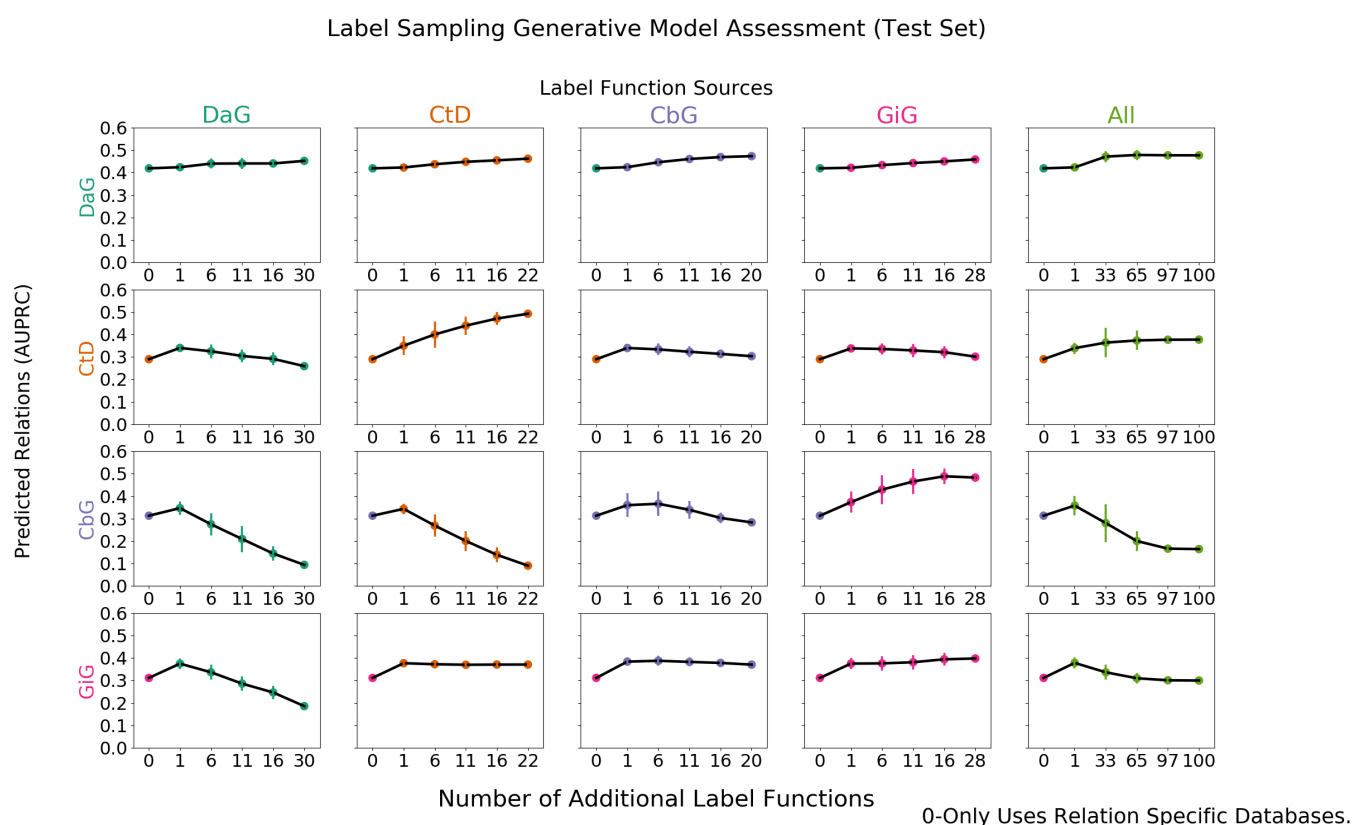
We added randomly sampled label functions to a baseline for each edge type to evaluate the feasibility of label function re-use. Our baseline model consisted of a generative model trained with only the edge type's distant supervision label functions. We report the results in the form of area under the precision recall curve (AUPR) (Figure 4) and area under the receiver operating curve (AUROC) (Figure 3).

The on-diagonal plots of figure 4) and figure 4 show performance when edge-specific label functions are added on top of edge-specific baselines. The general trend is performance increases in this setting. The Compound-treats-Disease (CtD) edge type is a quintessential example of this trend. The baseline model starts off with an AUROC score of 52% and an AUPRC of 28%, which increase to 76% and 49% respectively as more CtD label functions are included. Disease-associates-Gene (DaG) edges have a similar trend: performance starting off with a AUROC of 56% and AUPRC of 41%, which increase to 62% and 45% respectively. Both the Compound-binds-Gene (CbG) and Gene-interacts-Gene (GiG) edges have an increasing trend but plateau after a few label functions are added.

The off-diagonals in figure 4) and figure 4 show how performance varies when label functions from one edge type are added to a different edge type's baseline. In certain cases (apparent for DaG), performance increases regardless of the edge type used for label functions. In other cases (apparent with CtD), one label function appears to improve performance; however, adding more label functions does not improve performance (AUROC) or decreases it (AUPRC). In certain cases, the source of the label functions appear to be important: for CbG edges performance decreases when using label functions from the DaG and CtD categories.

Our initial hypothesis was based on the idea that certain edge types capture similar physical relationships and that these cases would be particularly amenable for label function transfer. For example, Compound-binds-Gene (CbG) and Gene-interacts-Gene (GiG) both describe physical interactions. We observed that performance increased as assessed by both AUPRC and AUPRC when using label functions from the GiG edge type to predict CbG edges. A similar trend was observed when predicting the GiG edge; however, the performance differences were small for this edge type making the importance difficult to assess.

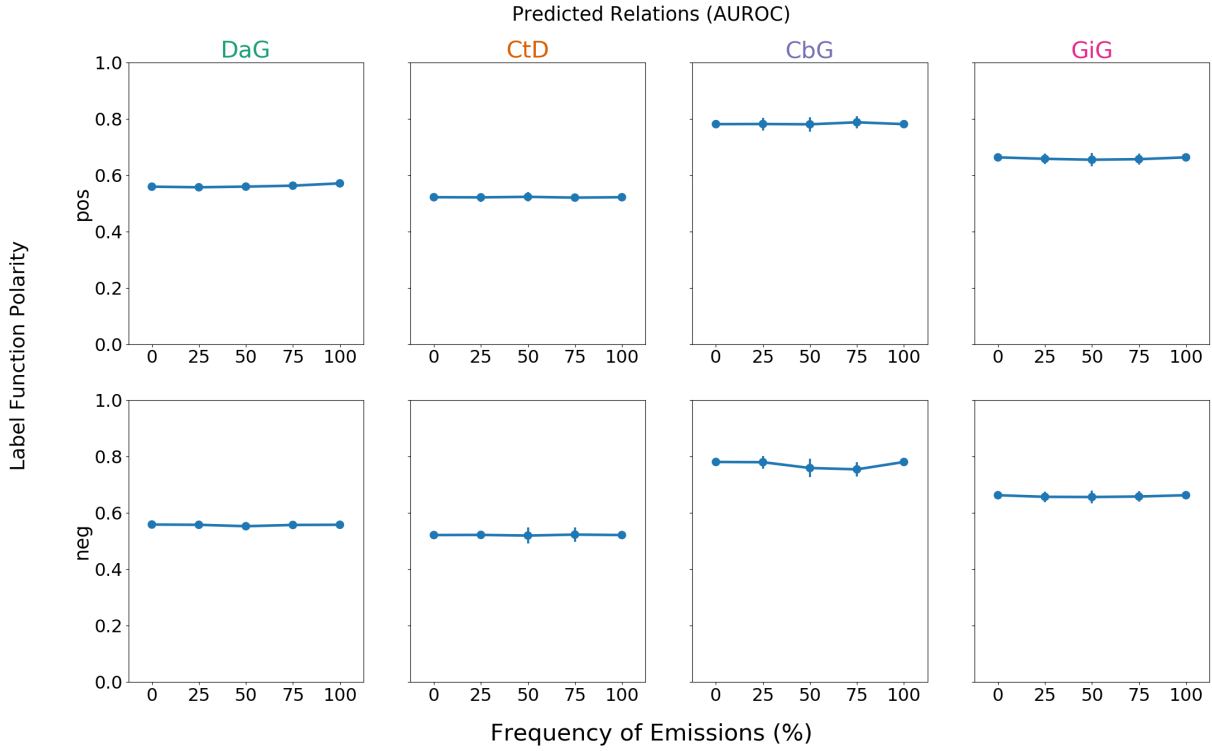
The last column shows performance when sampling from all label functions. Performance increased (AUROC and AUPRC) for both DaG and CtD, when sampling from the full pool of label functions. CbG and GiG also had increased performance when one random label function was sampled, but performance decreased drastically as more label functions were added. It is possible that a small number of irrelevant label functions are able to overwhelm the distant supervision label functions in these cases.



**Figure 4:** Grid of Area Under the Precision Recall Curve (AUPRC) scores for each generative model trained on randomly sampled label functions. The rows depict the relationship each model is trying to predict and the columns are the relationship specific sources each label function is sampled from. For example, the top-left most square depicts the generative model predicting Disease associates Gene (DaG) sentences, while randomly sampling label functions designed to predict the DaG relationship. The square towards the right depicts the generative model predicting DaG sentences, while randomly sampling label functions designed to predict the Compound treats Disease (CtD) relationship. This pattern continues filling out the rest of the grid. The last most column consists of pooling every relationship specific label function and proceeding as above.

## Random Label Function Gen Model Analysis

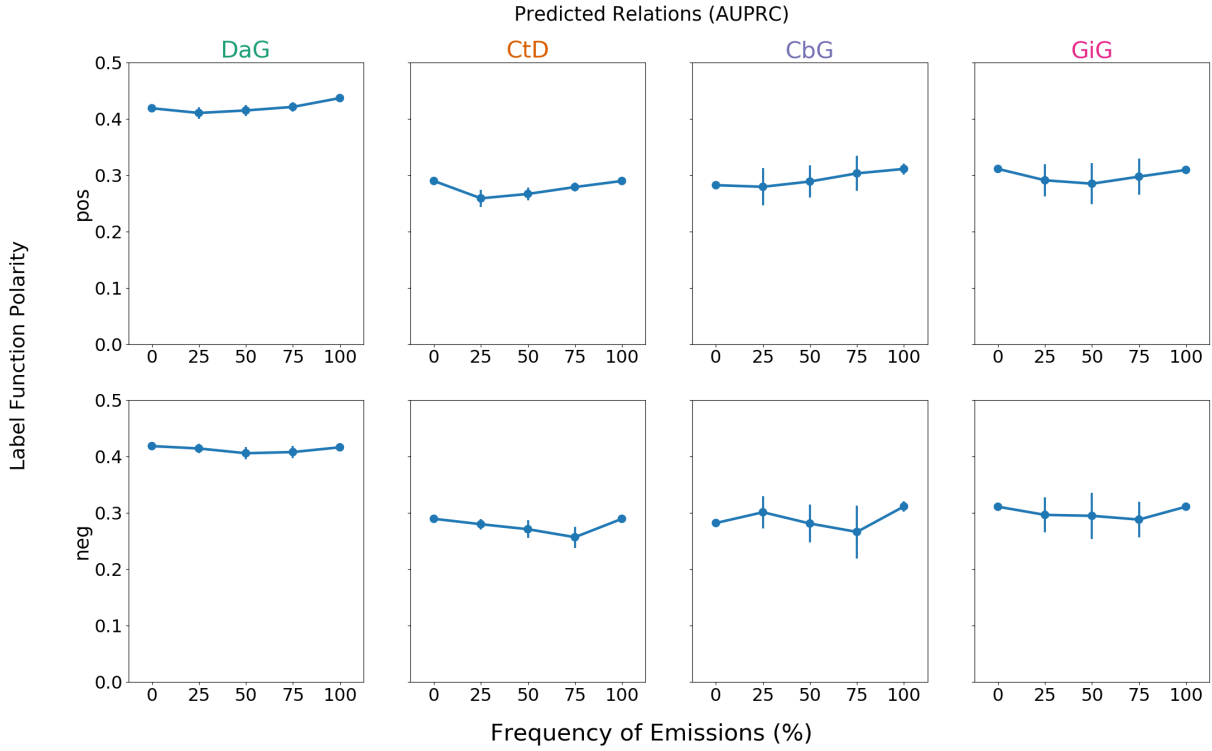
### Random Label Function Generative Model Assessment (Test Set)



**Figure 5:** A grid of area under the receiver operating curve (AUROC) for each edge type. Each plot consists of adding a single label function on top of the baseline model. This label function emits a positive (top row) or negative (bottom row) label at specified frequencies, and performance at zero is equivalent to not having a randomly emitting label function. The error bars represent 95% confidence intervals for AUROC (y-axis) at each emission frequency.

We observed that including one label function of a mismatched type to distant supervision often improved performance, so we evaluated the effects of adding a random label function in the same setting. We found that adding random noise did not usually improve performance (Figures 6 and 6). For the CbG edge type we did observe slightly increased performance via AUPR (Figure 6). However, in general the performance changes were smaller than those observed with mismatched label types.

## Random Label Function Generative Model Assessment (Test Set)



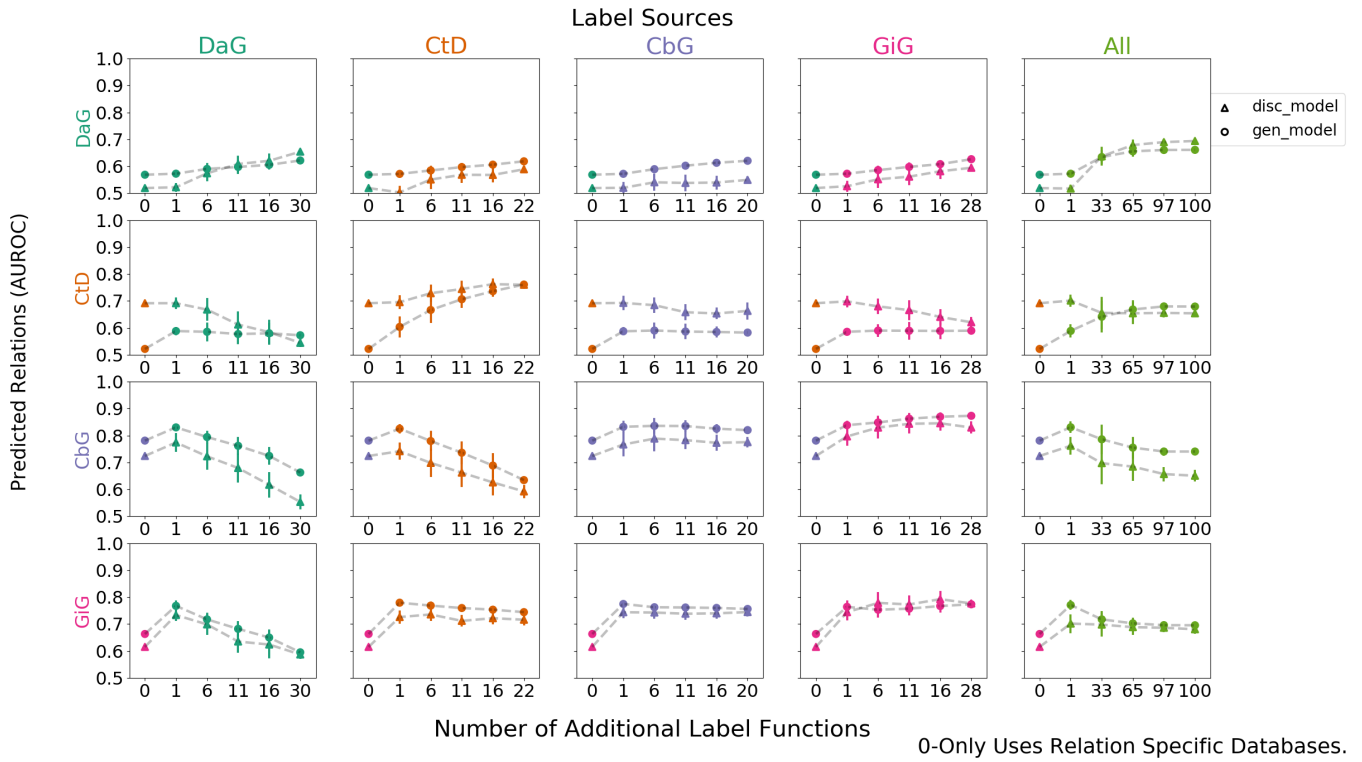
**Figure 6:** A grid of area under the precision recall curve (AUPR) for each edge type. Each plot consists of adding a single label function on top of the baseline model. This label function emits a positive (top row) or negative (bottom row) label at specified frequencies. The error bars represent 95% confidence intervals for AUPR (y-axis) at emission frequency.

## Discriminative Model Performance

In this framework we used a generative model trained over label functions to produce probabilistic training labels for each sentence. Then we train a discriminative model, which has full access to a representation of the text of the sentence, to predict the generated labels. The discriminative model is a convolutional neural network trained over word embeddings. We report the results of the discriminative model using AUPR (Figure 8) and AUROC (Figure 7).

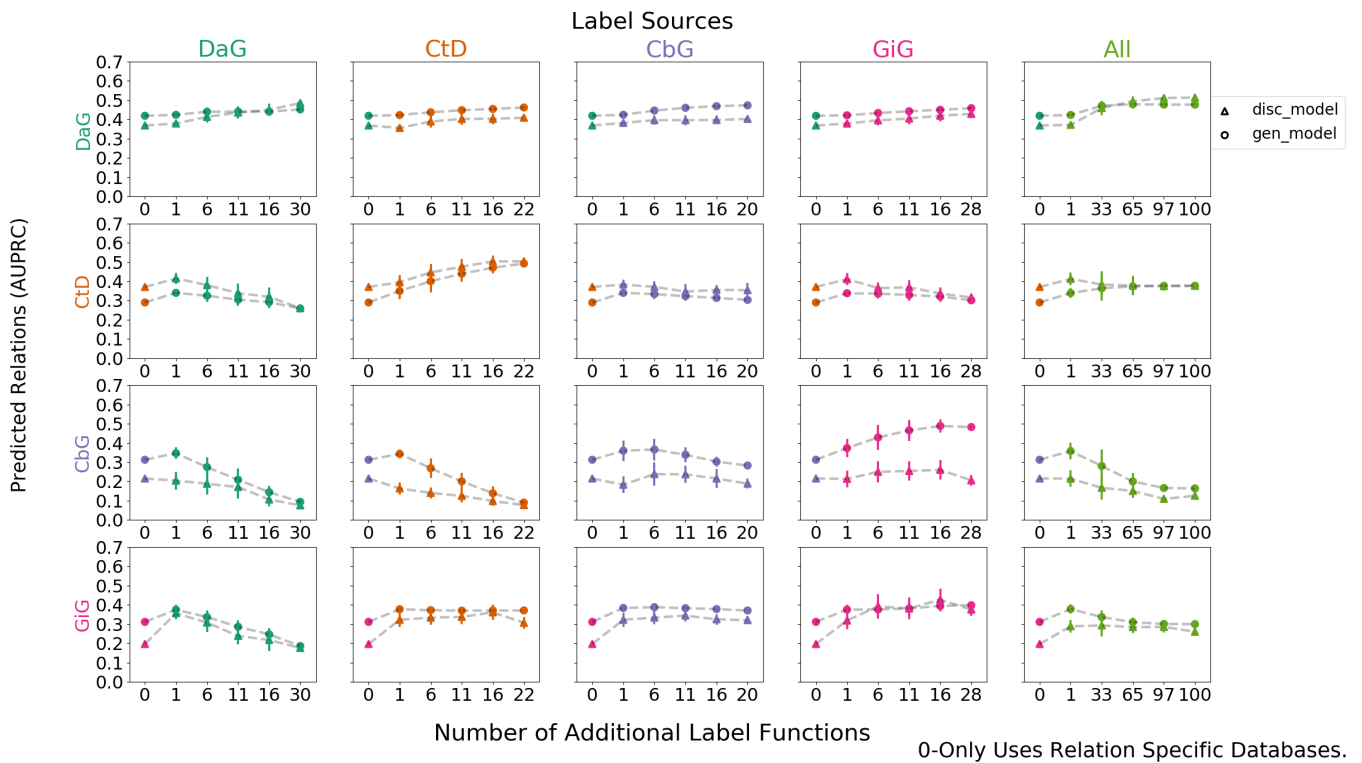
We found that the discriminative model under-performed the generative model in most cases. Only for the CtD edge does the discriminative model appear to provide performance above the generative model and that increased performance is only with modest numbers of label functions. With the full set of label functions, the performance of both remains similar. The trend observed in the generative model that one or a few mismatched label functions (off-diagonal) improves performance is retained despite the limited performance of the discriminative model.

### Label Sampling Discriminator Model Assessment (Test Set)



**Figure 7:** Grid of Area Under the Receiver Operating Curve (AUROC) scores for each discriminative model trained using generated labels from the generative models. The rows depict the relationship each model is trying to predict and the columns are the relationship specific sources each label function was sampled from. For example, the top-left most square depicts the discriminator model predicting Disease associates Gene (DaG) sentences, while randomly sampling label functions designed to predict the DaG relationship. The error bars over the points represents the standard deviation between sampled runs. The square towards the right depicts the discriminative model predicting DaG sentences, while randomly sampling label functions designed to predict the Compound treats Disease (CtD) relationship. This pattern continues filling out the rest of the grid. The last most column consists of pooling every relationship specific label function and proceeding as above.

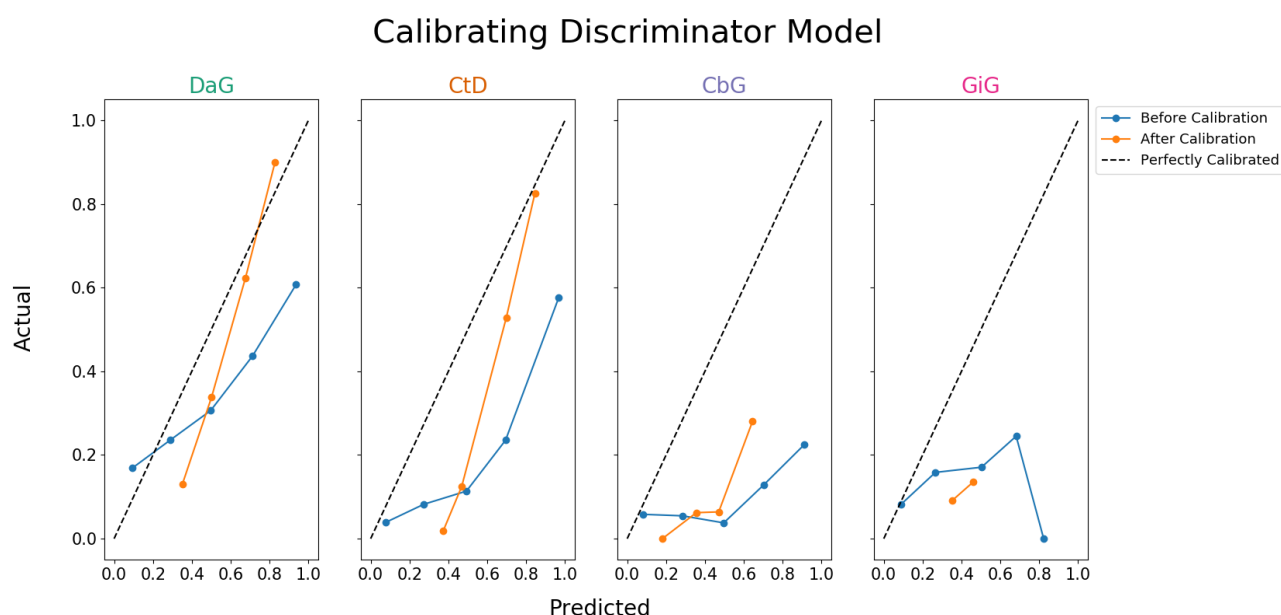
### Label Sampling Discriminator Model Assessment (Test Set)



**Figure 8:** Grid of Area Under the Receiver Operating Curve (AUROC) scores for each discriminative model trained using generated labels from the generative models. The rows depict the relationship each model is trying to predict and the columns are the relationship specific sources each label function was sampled from. For example, the top-left most square depicts the discriminator model predicting Disease associates Gene (DaG) sentences, while randomly sampling label functions designed to predict the DaG relationship. The error bars over the points represents the standard deviation between sampled runs. The square towards the right depicts the discriminative model predicting DaG sentences, while randomly sampling label functions designed to predict the Compound treats Disease (CtD) relationship. This pattern continues filling out the rest of the grid. The last most column consists of pooling every relationship specific label function and proceeding as above.

## Discriminative Model Calibration

Even deep learning models with high precision and recall can be poorly calibrated, and the overconfidence of these models has been noted [27,28]. We attempted to calibrate the best performing discriminative model so that we could directly use the emitted probabilities. We examined the calibration of our existing model (Figure 9, blue line). We found that the DaG and CtG edge types were, though not perfectly calibrated, were somewhat aligned with the ideal calibration lines. The CbG and GiG edges were poorly calibrated and increasing model certainty did not always lead to an increase in precision. Applying the calibration algorithm (orange line) did not appear to bring predictions in line with the ideal calibration line, but did capture some of the uncertainty in the GiG edge type. For this reason we use the measured precision instead of the predicted probabilities when determining how many edges could be added to existing knowledge bases with specified levels of confidence.



**Figure 9:** Calibration plots for the discriminative model. A perfectly calibrated model would follow the dashed diagonal line. The blue line represents the predictions before calibration and the orange line shows predictions after calibration.

## Reconstructing Hetionet

place figure of number of new edges that can be added to hetionet as well as edges we can reconstruct using this method

## Discussion

Here mention why performnace increases in the beginning for the generative model then decreases

Discuss discriminator model performance given generative model

Mention Take home messages

1. have a centralized set of negative label functions and focus more on constructing positive label functions

## **Conclusion and Future Direction**

Recap the original problem - takes a long time to create useful label function

Proposed solution - reuse label functions

Mention incorporating more relationships Mention creating a centralized multitask text extractor using this method.



## References

---

### 1. Graph Theory Enables Drug Repurposing – How a Mathematical Model Can Drive the Discovery of Hidden Mechanisms of Action

Ruggero Gramatica, T. Di Matteo, Stefano Giorgetti, Massimo Barbiani, Dorian Bevec, Tomaso Aste  
*PLoS ONE* (2014-01-09) <https://doi.org/gf45zp>  
DOI: [10.1371/journal.pone.0084912](https://doi.org/10.1371/journal.pone.0084912) · PMID: [24416311](https://pubmed.ncbi.nlm.nih.gov/24416311/) · PMCID: [PMC3886994](https://pubmed.ncbi.nlm.nih.gov/PMC3886994/)

### 2. Drug repurposing through joint learning on knowledge graphs and literature

Mona Alshahrani, Robert Hoehndorf  
*Cold Spring Harbor Laboratory* (2018-08-06) <https://doi.org/gf45zk>  
DOI: [10.1101/385617](https://doi.org/10.1101/385617)

### 3. Systematic integration of biomedical knowledge prioritizes drugs for repurposing

Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, Sergio E Baranzini  
*eLife* (2017-09-22) <https://doi.org/cdfk>  
DOI: [10.7554/elife.26726](https://doi.org/10.7554/elife.26726) · PMID: [28936969](https://pubmed.ncbi.nlm.nih.gov/28936969/) · PMCID: [PMC5640425](https://pubmed.ncbi.nlm.nih.gov/PMC5640425/)

### 4. Distant supervision for relation extraction without labeled data

Mike Mintz, Steven Bills, Rion Snow, Dan Jurafsky  
*Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - ACL-IJCNLP '09* (2009)  
<https://doi.org/fg9q43>  
DOI: [10.3115/1690219.1690287](https://doi.org/10.3115/1690219.1690287)

### 5. CoCoScore: Context-aware co-occurrence scoring for text mining applications using distant supervision

Alexander Junge, Lars Juhl Jensen  
*Cold Spring Harbor Laboratory* (2018-10-16) <https://doi.org/gf45zm>  
DOI: [10.1101/444398](https://doi.org/10.1101/444398)

### 6. Knowledge-guided convolutional networks for chemical-disease relation extraction

Huiwei Zhou, Chengkun Lang, Zhuang Liu, Shixian Ning, Yingyu Lin, Lei Du  
*BMC Bioinformatics* (2019-05-21) <https://doi.org/gf45zn>  
DOI: [10.1186/s12859-019-2873-7](https://doi.org/10.1186/s12859-019-2873-7) · PMID: [31113357](https://pubmed.ncbi.nlm.nih.gov/31113357/) · PMCID: [PMC6528333](https://pubmed.ncbi.nlm.nih.gov/PMC6528333/)

### 7. Facts from text: can text mining help to scale-up high-quality manual curation of gene products with ontologies?

R. Winnenburg, T. Wachter, C. Plake, A. Doms, M. Schroeder  
*Briefings in Bioinformatics* (2008-07-11) <https://doi.org/bfsnwg>  
DOI: [10.1093/bib/bbn043](https://doi.org/10.1093/bib/bbn043) · PMID: [19060303](https://pubmed.ncbi.nlm.nih.gov/19060303/)

### 8. Manual curation is not sufficient for annotation of genomic databases

William A. Baumgartner Jr, K. Bretonnel Cohen, Lynne M. Fox, George Acquaaah-Mensah, Lawrence Hunter  
*Bioinformatics* (2007-07-01) <https://doi.org/dtck86>  
DOI: [10.1093/bioinformatics/btm229](https://doi.org/10.1093/bioinformatics/btm229) · PMID: [17646325](https://pubmed.ncbi.nlm.nih.gov/17646325/) · PMCID: [PMC2516305](https://pubmed.ncbi.nlm.nih.gov/PMC2516305/)

### 9. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references

Lutz Bornmann, Rüdiger Mutz

**10. Data Programming: Creating Large Training Sets, Quickly**

Alexander Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, Christopher Ré  
*arXiv* (2016-05-25) <https://arxiv.org/abs/1605.07723v3>

**11. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)**

Jacqueline MacArthur, Emily Bowler, Maria Cerezo, Laurent Gil, Peggy Hall, Emma Hastings, Heather Junkins, Aoife McMahon, Annalisa Milano, Joannella Morales, ... Helen Parkinson  
*Nucleic Acids Research* (2016-11-29) <https://doi.org/f9v7cp>  
DOI: [10.1093/nar/gkw1133](https://doi.org/10.1093/nar/gkw1133) · PMID: [27899670](https://pubmed.ncbi.nlm.nih.gov/27899670/) · PMCID: [PMC5210590](https://pubmed.ncbi.nlm.nih.gov/PMC5210590/)

**12. DrugBank 5.0: a major update to the DrugBank database for 2018**

David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, ... Michael Wilson  
*Nucleic Acids Research* (2017-11-08) <https://doi.org/gcwtzk>  
DOI: [10.1093/nar/gkx1037](https://doi.org/10.1093/nar/gkx1037) · PMID: [29126136](https://pubmed.ncbi.nlm.nih.gov/29126136/) · PMCID: [PMC5753335](https://pubmed.ncbi.nlm.nih.gov/PMC5753335/)

**13. PubTator: a web-based text mining tool for assisting biocuration**

Chih-Hsuan Wei, Hung-Yu Kao, Zhiyong Lu  
*Nucleic Acids Research* (2013-05-22) <https://doi.org/f475th>  
DOI: [10.1093/nar/gkt441](https://doi.org/10.1093/nar/gkt441) · PMID: [23703206](https://pubmed.ncbi.nlm.nih.gov/23703206/) · PMCID: [PMC3692066](https://pubmed.ncbi.nlm.nih.gov/PMC3692066/)

**14. DNorm: disease name normalization with pairwise learning to rank**

R. Leaman, R. Islamaj Dogan, Z. Lu  
*Bioinformatics* (2013-08-21) <https://doi.org/f5gj9n>  
DOI: [10.1093/bioinformatics/btt474](https://doi.org/10.1093/bioinformatics/btt474) · PMID: [23969135](https://pubmed.ncbi.nlm.nih.gov/23969135/) · PMCID: [PMC3810844](https://pubmed.ncbi.nlm.nih.gov/PMC3810844/)

**15. GeneTUKit: a software for document-level gene normalization**

M. Huang, J. Liu, X. Zhu  
*Bioinformatics* (2011-02-08) <https://doi.org/dng2cb>  
DOI: [10.1093/bioinformatics/btr042](https://doi.org/10.1093/bioinformatics/btr042) · PMID: [21303863](https://pubmed.ncbi.nlm.nih.gov/21303863/) · PMCID: [PMC3065680](https://pubmed.ncbi.nlm.nih.gov/PMC3065680/)

**16. Cross-species gene normalization by species inference**

Chih-Hsuan Wei, Hung-Yu Kao  
*BMC Bioinformatics* (2011-10-03) <https://doi.org/dnmvds>  
DOI: [10.1186/1471-2105-12-s8-s5](https://doi.org/10.1186/1471-2105-12-s8-s5) · PMID: [22151999](https://pubmed.ncbi.nlm.nih.gov/22151999/) · PMCID: [PMC3269940](https://pubmed.ncbi.nlm.nih.gov/PMC3269940/)

**17. Collaborative biocuration-text-mining development task for document prioritization for curation**

T. C. Wieggers, A. P. Davis, C. J. Mattingly  
*Database* (2012-11-22) <https://doi.org/gbb3zw>  
DOI: [10.1093/database/bas037](https://doi.org/10.1093/database/bas037) · PMID: [23180769](https://pubmed.ncbi.nlm.nih.gov/23180769/) · PMCID: [PMC3504477](https://pubmed.ncbi.nlm.nih.gov/PMC3504477/)

**18. The Stanford CoreNLP Natural Language Processing Toolkit**

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, David McClosky  
*Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (2014) <https://doi.org/gf3xhp>  
DOI: [10.3115/v1/p14-5010](https://doi.org/10.3115/v1/p14-5010)

**19. A global network of biomedical relationships derived from text**

Bethany Percha, Russ B Altman

*Bioinformatics* (2018-02-27) <https://doi.org/gc3ndk>  
DOI: [10.1093/bioinformatics/bty114](https://doi.org/10.1093/bioinformatics/bty114) · PMID: [29490008](https://pubmed.ncbi.nlm.nih.gov/29490008/) · PMCID: [PMC6061699](https://pubmed.ncbi.nlm.nih.gov/PMC6061699/)

## 20. **Snorkel MeTaL**

Alex Ratner, Braden Hancock, Jared Dunnmon, Roger Goldman, Christopher Ré  
*Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning - DEEM'18* (2018) <https://doi.org/gf3xk7>  
DOI: [10.1145/3209889.3209898](https://doi.org/10.1145/3209889.3209898) · PMID: [30931438](https://pubmed.ncbi.nlm.nih.gov/30931438/) · PMCID: [PMC6436830](https://pubmed.ncbi.nlm.nih.gov/PMC6436830/)

## 21. **Snorkel**

Alexander Ratner, Stephen H. Bach, Henry Ehrenberg, Jason Fries, Sen Wu, Christopher Ré  
*Proceedings of the VLDB Endowment* (2017-11-01) <https://doi.org/ch44>  
DOI: [10.14778/3157794.3157797](https://doi.org/10.14778/3157794.3157797) · PMID: [29770249](https://pubmed.ncbi.nlm.nih.gov/29770249/) · PMCID: [PMC5951191](https://pubmed.ncbi.nlm.nih.gov/PMC5951191/)

## 22. **Distributed Representations of Words and Phrases and their Compositionality**

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean  
*arXiv* (2013-10-16) <https://arxiv.org/abs/1310.4546v1>

## 23. **Enriching Word Vectors with Subword Information**

Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov  
*arXiv* (2016-07-15) <https://arxiv.org/abs/1607.04606v2>

## 24. **Efficient Estimation of Word Representations in Vector Space**

Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean  
*arXiv* (2013-01-16) <https://arxiv.org/abs/1301.3781v3>

## 25. **A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification**

Ye Zhang, Byron Wallace  
*arXiv* (2015-10-13) <https://arxiv.org/abs/1510.03820v4>

## 26. **Adam: A Method for Stochastic Optimization**

Diederik P. Kingma, Jimmy Ba  
*arXiv* (2014-12-22) <https://arxiv.org/abs/1412.6980v9>

## 27. **On Calibration of Modern Neural Networks**

Chuan Guo, Geoff Pleiss, Yu Sun, Kilian Q. Weinberger  
*arXiv* (2017-06-14) <https://arxiv.org/abs/1706.04599v2>

## 28. **Accurate Uncertainties for Deep Learning Using Calibrated Regression**

Volodymyr Kuleshov, Nathan Fenner, Stefano Ermon  
*arXiv* (2018-07-01) <https://arxiv.org/abs/1807.00263v1>

## 29. **A Proteome-Scale Map of the Human Interactome Network**

Thomas Rolland, Murat Taşan, Benoit Charleatoux, Samuel J. Pevzner, Quan Zhong, Nidhi Sahni, Song Yi, Irma Lemmens, Celia Fontanillo, Roberto Mosca, ... Marc Vidal  
*Cell* (2014-11) <https://doi.org/f3mn6x>  
DOI: [10.1016/j.cell.2014.10.050](https://doi.org/10.1016/j.cell.2014.10.050) · PMID: [25416956](https://pubmed.ncbi.nlm.nih.gov/25416956/) · PMCID: [PMC4266588](https://pubmed.ncbi.nlm.nih.gov/PMC4266588/)

## 30. **iRefIndex: A consolidated protein interaction database with provenance**

Sabry Razick, George Magklaras, Ian M Donaldson  
*BMC Bioinformatics* (2008) <https://doi.org/b99bjj>  
DOI: [10.1186/1471-2105-9-405](https://doi.org/10.1186/1471-2105-9-405) · PMID: [18823568](https://pubmed.ncbi.nlm.nih.gov/18823568/) · PMCID: [PMC2573892](https://pubmed.ncbi.nlm.nih.gov/PMC2573892/)

31. **Uncovering disease-disease relationships through the incomplete interactome**

J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabasi

*Science* (2015-02-19) <https://doi.org/f3mn6z>

DOI: [10.1126/science.1257601](https://doi.org/10.1126/science.1257601) · PMID: [25700523](https://pubmed.ncbi.nlm.nih.gov/25700523/) · PMCID: [PMC4435741](https://pubmed.ncbi.nlm.nih.gov/PMC4435741/)

## Supplemental Figures

### Confidence Scores for Each Edge Type

Contains the top ten Disease-associates-Gene confidence scores before and after model calibration. Disease mentions are highlighted in **brown** and Gene mentions are highlighted in **blue**.{#tbl:dg\_top\_ten\_table}

Disease Name	Gene Symbol	Text	Before Calibration	After Calibration
adrenal gland cancer	TP53	the mechanisms of adrenal tumorigenesis remain poorly established ; the r337h germline mutation in the <b>p53</b> gene has previously been associated with <b>acts</b> in brazilian children .	1.0	0.882
breast cancer	ERBB2	in <b>breast cancer</b> , overexpression of <b>her2</b> is associated with an aggressive tumor phenotype and poor prognosis .	1.0	0.845
lung cancer	TP53	however , both adenine ( a ) and guanine ( g ) mutations are found in the <b>p53</b> gene in cr exposure-related <b>lung cancer</b> .	1.0	0.83
malignant glioma	BAX	these data suggest that the combination of tra-8 treatment with specific overexpression of <b>bax</b> using advegf <b>bax</b> may be an effective approach for the treatment of human <b>malignant gliomas</b> .	0.999	0.827
polycystic ovary syndrome	SERPINE1	4 g allele in <b>pai-1</b> gene was more frequent in <b>pcos</b> and the 4g/4 g genotype was associated with increased pai-1 levels .	0.999	0.814
systemic lupus erythematosus	PRL	results : <b>sle</b> patients showed a significantly higher serum level of <b>prl</b> than healthy subjects , which was especially obvious in the active stage of the disease ( p = 0.000 .	0.999	0.813
hematologic cancer	TNF	the mean <b>tnf-alpha</b> plasma concentration in the patients with <b>cll</b> was significantly higher than in the healthy control population ( 16.4 versus 8.7 pg/ml ; p < .0001 ) .	0.999	0.81

Disease Name	Gene Symbol	Text	Before Calibration	After Calibration
lung cancer	MUC16	the mean concentration of <b>ca 125</b> was higher in patients with <b>lung cancer</b> ( 37 + / - 81 u/ml ) than in those with nonmalignant disease ( 4.2 + / - 5.7 u/ml ) ( p less than 0.01 ) .	0.999	0.806
prostate cancer	AR	the <b>androgen receptor</b> was expressed in all primary and metastatic <b>prostate cancer</b> tissues and no mutations were identified .	0.999	0.801
breast cancer	ERBB2	the results of multiple linear regression analysis , with her2 as the dependent variable , showed that family history of <b>breast cancer</b> was significantly associated with elevated <b>her2</b> levels in the tumors ( p = 0.0038 ) , after controlling for the effects of age , tumor estrogen receptor , and dna index .	0.999	0.8

**Table 3:** Contains the bottom ten Disease-associates-Gene confidence scores before and after model calibration. Disease mentions are highlighted in **brown** and Gene mentions are highlighted in **blue**.

Disease Name	Gene Symbol	Text	Before Calibration	After Calibration
breast cancer	NAT2	[ the relationship between passive smoking , <b>breast cancer</b> risk and <b>n-acetyltransferase 2</b> ( nat2 ) ] .	0.012	0.287
schizophrenia	EP300	ventricle size and <b>p300</b> in <b>schizophrenia</b> .	0.012	0.286

Di se as e N a m e	G e n e s y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib rai to n
he m a t o l o g i c a n c e r	C D 3 3	in the 2 ( nd ) study of <b>cd33</b> + <b>sr-aml</b> 2 doses of go ( 4.5 - 9 mg/m ( 2 ) ) were administered > = 60d post reduced intensity conditioning ( ric ) allosct ( 8 wks apart ) .	0. 01	0. 28 1
Cr oh n's di se as e	P T P N 2	in this sample , we were able to confirm an association between <b>cd</b> and <b>ptpn2</b> ( genotypic p = 0.019 and allelic p = 0.011 ) , and phenotypic analysis showed an association of this snp with late age at first diagnosis , inflammatory and penetrating cd behaviour , requirement of bowel resection and being a smoker at diagnosis .	0. 00 8	0. 26 8
br ea st ca nc e r	E R B B 2	long-term efficacy and safety of adjuvant trastuzumab for <b>her2-positive</b> early [breast cancer ] .	0. 00 7	0. 26 2
he m a t o l o g i c a n c e r	C D 4 0 L G	we examined the direct effect of lenalidomide on <b>cli-cell proliferation</b> induced by <b>cd154-expressing</b> accessory cells in media containing interleukin-4 and -10 .	0. 00 6	0. 25 9
he m a t o l o g i c a n c e r	M L A N A	methods : the sln sections ( n = 214 ) were assessed by qrt assay for 4 established messenger rna biomarkers : <b>mart-1</b> , mage-a3 , <b>galnac-t</b> , and pax3 .	0. 00 5	0. 25 2
br ea st ca nc e r	E R B B 2	the keywords erbb2 or her2 or erbb-2 or <b>her-2</b> and <b>breast cancer</b> and ( country ) were used to search pubmed , international and local conference abstracts and local-language journals from the year 2000 onwards .	0. 00 3	0. 22 5



Disease Name	Gene Symbol	Text	Before Calibration	After Calibration
hepatitis B	PKD2	conversely , a significant enhancement of activation was observed for afb1 in cases of mild cah and especially for <b>trp-p-2</b> in <b>hepatitis b</b> virus carriers , irrespective of their histologic diagnosis .	0.002	0.217
hematologic cancer	C7	serum antibody responses to four haemophilus influenzae type b capsular polysaccharide-protein conjugate vaccines ( prp-d , hboc , <b>c7p</b> , and <b>prp-t</b> ) were studied and compared in 175 infants , 85 adults and 140 2-year-old children .	0.002	0.208

Contains the top ten Compound-treats-Disease confidence scores after model calibration. Disease mentions are highlighted in **brown** and Compound mentions are highlighted in **red**.{#tbl:cd\_top\_ten\_table}

Compound Name	Disease Name	Text	Before Calibration	After Calibration
Methylprednisolone	asthma	use of tao without <b>methylprednisolone</b> in the treatment of severe <b>asthma</b> .	1.0	0.895
Methyldopa	hypertension	atenolol and <b>methyldopa</b> in the treatment of <b>hypertension</b> .	1.0	0.888
Prednisone	asthma	<b>prednisone</b> and beclomethasone for treatment of <b>asthma</b> .	1.0	0.885
Prazosin	hypertension	experience with <b>prazosin</b> in the treatment of <b>hypertension</b> .	1.0	0.883
Prazosin	hypertension	<b>prazosin</b> in the treatment of <b>hypertension</b> .	1.0	0.878
Prazosin	hypertension	<b>prazosin</b> in the treatment of [hypertension] .	1.0	0.878
Methyldopa	hypertension	oxprenolol plus cyclopenthiiazide-kcl versus <b>methyldopa</b> in the treatment of <b>hypertension</b> .	1.0	0.877
Prednisolone	lymphatic system cancer	peptichemio : a new oncolytic drug in combination with vincristine and <b>prednisolone</b> in the treatment of <b>non-hodgkin lymphomas</b> .	1.0	0.871
Methyldopa	hypertension	methyldopate , the ethyl ester hydrochloride salt of <b>alpha-methyldopa</b> ( alpha-md ) , is used extensively in the treatment of severe <b>hypertension</b> .	1.0	0.851

Compound Name	Disease Name	Text	Before Calibration	After Calibration
Haloperidol	Gilles de la Tourette syndrome	a comparison of pimozide and <b>haloperidol</b> in the treatment of gilles de la tourette 's syndrome .	1.0	0.839

Contains the bottom ten Compound-treats-Disease confidence scores before and after model calibration. Disease mentions are highlighted in **brown** and Compound mentions are highlighted in **red**.{#tbl:cd\_bottom\_ten\_table}

Compound Name	Disease Name	Text	Before Calibration	After Calibration
Dexamethasone	hypertension	<b>dexamethasone</b> and <b>hypertension</b> in preterm infants .	0.011	0.034
Reserpine	hypertension	<b>reserpine</b> in <b>hypertension</b> : present status .	0.01	0.036
Creatine	coronary artery disease	scintiphotographic findings were compared with the size of <b>myocardial infarcts</b> calculated from measurements of the activity of mb isoenzymes of <b>creatine</b> kinase ( ck-mb ) in serum and in the myocardium at autopsy , as described by sobel 's method .	0.009	0.034
Hydrocortisone	brain cancer	to explore the effects of repeated episodes of hypercortisolemia on <b>hypothalamic-pituitary-adrenal axis</b> regulation , we studied plasma acth and <b>cortisol</b> ( cort ) responses to 100 micrograms human crh ( hcrh ) in 10 dexamethasone ( 1.5 mg ) - pretreated elderly endurance athletes who had abstained from physical activity for at least 48 h before testing and 13 sedentary age-matched controls .	0.009	0.033

Compound Name	Disease Name	Text	Before Calibration	After Calibration
Hydrocortisone	brain cancer	basal activity of the <b>hypothalamic-pituitary-adrenal axis</b> was estimated by determinations of 24-h urinary free cortisol-excretion , evening basal plasma total and free <b>cortisol</b> concentrations , and the cortisol binding globulin-binding capacity .	0 . 0 0 8	0 . 3 2 8
Creatine	coronary artery disease	during successful and uncomplicated angioplasty ( ptca ) , we studied the effect of a short lasting <b>myocardial ischemia</b> on plasma creatine kinase , creatine kinase mb-activity , and <b>creatine</b> kinase mm-isoforms ( mm1 , mm2 , mm3 ) in 23 patients .	0 . 0 0 6	0 . 3 1 8
Benzylpenicillin	epilepsy	it was shown in experiments on cats under nembutal anesthesia that a lesion of the medial forebrain bundle ( mfb ) and partly of the preoptic region at the side of local penicillin application on the cerebral cortex ( g. suprasylvius medius ) results in depression of the <b>epileptiform activity</b> in the <b>penicillin-induced</b> focus , as well as in the secondary " mirror " focus , which appeared in the symmetrical cortex area of the other hemisphere .	0 . 0 0 5	0 . 3 1 5
Indomethacin	hypertension	effects of <b>indomethacin</b> in rabbit <b>renovascular hypertension</b> .	0 . 0 0 4	0 . 3 0 8
Cyclic Adenosine Monophosphate	ovarian cancer	the hormonal regulation of steroidogenesis and <b>adenosine 3 ' :5 ' - cyclic monophosphate</b> in <b>embryonic-chick ovary</b> .	0 . 0 0 2	0 . 2 9 2

Compound Name	Disease Name	Text	Before Calibration	After Calibration
Dobutamine	coronary artery disease	two-dimensional echocardiography can detect regional wall motion abnormalities resulting from myocardial ischemia produced by dobutamine infusion .	0.002	0.287

Contains the top ten Compound-treats-Disease confidence scores before and after model calibration. Gene mentions are highlighted in blue and Compound mentions are highlighted in red.

Compound Name	Gene Symbol	Text	Before Calibration	After Calibration
Hydrocortisone	SHBG	serum concentrations of testicular and adrenal androgens and androgen precursors , cortisol , unconjugated ( e1 ) and total estrone ( te1 ; greater than or equal to 85 % e1 sulfate ) , pituitary hormones , sex hormone binding globulin ( shbg ) and albumin were measured in 14 male patients with non-diabetic end stage renal disease and in 28 age-matched healthy controls .	0.997	0.745

C o m p o u n d N a m e	G e n e s y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib ra tio n
M i n o x i d i l	E G F R	direct measurement of the ability of minoxidil to compete for binding to the egf receptor indicated that <b>minoxidil</b> probably does not bind to the <b>egf receptor</b> .	0. 99	0. 70 6
H y d r o c o r t i s o n e	S H B G	gonadotropin , testosterone , sex hormone binding globulin ( <b>shbg</b> ) , dehydroepiandrosterone sulphate , androstenedione , estradiol , prolactin , <b>cortisol</b> , thyrotropin , and free thyroxine levels were determined .	0. 98 8	0. 7
C h o l e c a l c i f e r o l	D B P	<b>cholecalciferol</b> ( vitamin d3 ) and its 25-hydroxy metabolite are transported in plasma bound to a specific protein , the binding protein for cholecalciferol and its metabolites ( <b>dbp</b> ) .	0. 98 3	0. 68 5
I n d o m e t h a c i n	A G T	<b>indomethacin</b> , a potent inhibitor of prostaglandin synthesis , is known to increase the maternal blood pressure response to <b>angiotensin ii</b> infusion .	0. 98 2	0. 68

C o m p o u n d N a m e	G e n e S y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib ra tio n
T r e t i n o i n	R X R A	the vitamin a derivative <b>retinoic acid</b> exerts its effects on transcription through two distinct classes of nuclear receptors , the retinoic acid receptor ( rar ) and the <b>retinoid x receptor</b> ( rxr ) .	0. 97 5	0. 66 8
D o p a m i n e	N T S	<b>neurotensin</b> binding was not modified by the addition of <b>dopamine</b> .	0. 97	0. 65 9
D - T y r o s i n e	P L C G 1	epidermal growth factor ( egf ) or platelet-derived growth factor binding to their receptor on fibroblasts induces tyrosine phosphorylation of plc gamma 1 and stable association of <b>plc gamma 1</b> with the receptor protein <b>tyrosine</b> kinase .	0. 96 9	0. 65 9
D - T y r o s i n e	P L C G 1	<b>tyrosine</b> phosphorylation of plc-ii was stimulated by low physiological concentrations of egf ( 1 nm ) , was quantitative , and was already maximal after a 30 sec incubation with 50 nm egf at 37 degrees c. interestingly , antibodies specific for plc-ii were able to coimmunoprecipitate the egf receptor and antibodies against egf receptor also coimmunoprecipitated <b>plc-ii</b> .	0. 96 4	0. 65 1
K e t a m i n e	C 5	additionally , reduction of glycine binding by the c-5 antagonists was reversed by both nmda receptor agonists and c-7 competitive <b>nmda</b> antagonists , providing evidence that the site of action of these <b>c-5</b> antagonists is the nmda recognition site , resulting in indirect modulation of the glycine site .	0. 95 7	0. 64 3

Contains the bottom ten Compound-binds-Gene confidence scores before and after model calibration. Gene mentions are highlighted in **blue** and Compound mentions are highlighted in **red**.{#tbl:cg\_bottom\_ten\_table}

C o m p o u n d N a m e	G e n e s y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib ra tio n
I r o n	N D U F B 3	since gastric acid plays an important role in the absorption process of <b>iron</b> and vitamin b12 , we determined levels of iron , ferritin , vitamin <b>b12</b> , and folic acid in 75 serum samples obtained during continuous omeprazole therapy ( 6-48 months after start of therapy ) from 34 patients with peptic diseases ( primarily reflux esophagitis ) .	0. 00 6	0. 27 6
D - T y r o s i n e	P L A U	either the 55 kda u-pa form and the lower mw form ( 33 kda ) derived from the 55 kda <b>u-pa</b> are tyrophosphorylated also the u-pa secreted in the culture media of human fibrosarcoma cells ( ht-1080 ) is phosphorylated in <b>tyrosine</b> as well as u-pa present in tissue extracts of tumors induced in nude mice by ht-1080 cells .	0. 00 6	0. 27 6
D - L e u c i n e	P O M C	cross-reactivities of <b>leucine-enkephalin</b> and <b>beta-endorphin</b> with the eia were less than 0.1 % , while that with gly-gly-phe-met and oxidized gly-gly-phe-met were 2.5 % and 10.2 % , respectively .	0. 00 6	0. 27 3
E p r a z i n o n e	G A S T	in patients with renal failure there exists the inhibition of the gastrin acid secretion which is the cause of the weakening of the mechanism of the feedback connection between <b>hcl</b> and <b>gastrin</b> , while because of a permanent stimulation of g-cells , the hyperplasia of these cells develops , as well as the increased secretory activity , and hypergastrinemia .	0. 00 5	0. 27 1
H y d r o c o r t i s o n e	G H 1	luteinizing hormone responses to luteinizing hormone releasing hormone , and <b>growth hormone</b> and <b>cortisol</b> responses to insulin induced hypoglycaemia in functional secondary amenorrhoea .	0. 00 5	0. 27 1



C o m p o u n d N a m e	G e n e s y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib ra tio n
H y d r o c o r t i s o n e	G H 1	group iv patients had normal basal levels of lh and normal lh , gh and cortisol responses .	0. 00 5	0. 26 9
B u p i v a c a i n e	A V P	plasma renin activity and vasopressin concentration , arterial pressure , and serum osmolality were measured in 17 patients before and after random epidural injection of either 6.7 ml of 0.75 % bupivacaine ( n = 7 ) or the same volume of saline ( n = 10 ) .	0. 00 4	0. 26
E p i n e p h r i n e	I N S	thermogenic effect of thyroid hormones : interactions with epinephrine and insulin .	0. 00 4	0. 25 9
H y d r o c o r t i s o n e	G H 1	cortisol and growth hormone ( gh ) secretion ( spontaneous variations at night and the release induced by insulin hypoglycaemia ) were investigated in 69 children and adolescents .	0. 00 2	0. 24 1

C o m p o u n d N a m e	G e n e S y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib ra tio n
E s t r i o l	L G A L S 1	[ diagnostic value of serial determination of <b>estriol</b> and <b>hpl</b> in plasma and of total estrogens in 24-h-urine compared to single values for diagnosis of fetal danger ] .	0. 0	0. 18 1

Contains the top ten Gene-interacts-Gene confidence scores before and after model calibration. Both gene mentions highlighted in **blue** .{#tbl:gg\_top\_ten\_table}

G e n e 1 S y m b o l	G e n e 2 S y m b o l	Text	Be fo re Ca lib ra tio n	Af te r Ca lib ra tio n
IN S	H SP A4	conclusions : intact insulin only weakly interacts with the <b>hsp70</b> chaperone dnak whereas monomeric proinsulin and peptides from 3 distinct <b>proinsulin</b> regions show substantial chaperone binding .	0.83 4	0.57 4
N M T1	S1 00 B	values for k ( cat ) indicated that , once gag or <b>nef</b> binds to the enzyme , myristoylation by <b>nmt1</b> and nmt2 proceeds at comparable rates .	0.82 6	0.57 1
VE GF A	HI F1 A	mechanistically , we demonstrated that resveratrol inhibited <b>hif-1alpha</b> and <b>vegf</b> expression through multiple mechanisms .	0.82	0.56 9
IT G AV	PE CA M 1	antigens expressed on emp and ec were assayed flow cytometrically and included constitutive markers ( <b>cd31</b> , <b>cd51/61</b> , cd105 ) , inducible markers ( cd54 , cd62e and cd106 ) , and annexin v binding .	0.81	0.56 6
F1 0	PF 4	these compounds inhibit both <b>factor xa</b> and thrombin , in the presence of antithrombin , while they are devoid of undesirable non-specific interactions , particularly with <b>platelet factor 4</b> ( pf4 ) .	0.76 6	0.55 4
NF KB 2	RE LB	the results indicate that dystrophic muscle is characterized by increases in the whole cell expression of ikappab-alpha , p65 , p50 , <b>relb</b> , <b>p100</b> , p52 , ikk , and traf-3 .	0.76	0.55 3
SS SC A1	C D K N 1B	conclusion : hl-60 / ht cells have lower <b>p27 ( kip1 )</b> expression compared with hl-60 cells .	0.75 7	0.55 2
PT H 2R	PT H 2	thus , the juxtamembrane receptor domain specifies the signaling and binding selectivity of <b>tip39</b> for the <b>pth2 receptor</b> over the pth1 receptor .	0.74 9	0.55

Gene1 Symbol	Gene2 Symbol	Text	Before Calibration	After Calibration
MM P9	MM P2	all these factors markedly influenced the secretion and/or activation of <b>mmp-2</b> and <b>mmp-9</b> .	0.738	0.547
CND 1	ABL1	synergy with <b>v-abl</b> depended on a motif in <b>cyclin d1</b> that mediates its binding to the retinoblastoma protein , suggesting that abl oncogenes in part mediate their mitogenic effects via a retinoblastoma protein-dependent pathway .	0.736	0.547

Contains the bottom ten Gene-interacts-Gene confidence scores before and after model calibration. Both gene mentions highlighted in **blue** .{#tbl:gg\_bottom\_ten\_table}

Gene1 Symbol	Gene2 Symbol	Text	Before Calibration	After Calibration
IFNG	IL6	in the control group , the positive rate for il-4 , <b>il-6</b> , il-10 were 0/10 , 2/10 and 1/10 , respectively , and those for il-2 and <b>ifn-gamma</b> were both 1/10 .	0.012	0.306
ACHE	BCHE	anticholinesterase activity was determined against acetylcholinesterase ( <b>ache</b> ) and butyrylcholinesterase ( <b>bche</b> ) , the enzymes vital for alzheimer 's disease , at 50 , 100 and 200 g ml ( -1 ) .	0.011	0.306
CC L2	AGT	we found no significant increase in <b>mcp-1</b> concentrations by <b>ang ii</b> alone ; but it enhanced the tnf-alpha-induced mcp-1 mrna expression in a dose-dependent manner .	0.011	0.306
CXCL8	IL1B	furthermore , somatostatin completely abrogated the increased secretion of <b>il-8</b> and <b>il-1beta</b> after invasion by salmonella .	0.011	0.303
SULT1A2	SULT1A3	to date , the laboratory has cloned seven unique human sulfotransferases ; five aryl sulfotransferases ( hast1 , hast2 , <b>hast3</b> , <b>hast4</b> and hast4v ) , an estrogen sulfotransferase and a dehydroepiandrosterone sulfotransferase .	0.009	0.295
IFNG	IL10	results : we found weak mrna expression of interleukin-4 ( il-4 ) and il-5 , and strong expression of il-6 , <b>il-10</b> and <b>ifn-gamma</b> before therapy .	0.008	0.292
IL2	IFNG	prostaglandin e2 at priming of naive cd4 + t cells inhibits acquisition of ability to produce <b>ifn-gamma</b> and <b>il-2</b> , but not il-4 and il-5 .	0.007	0.289
IL2	IFNG	the detailed distribution of lymphokine-producing cells showed that <b>il-2</b> and <b>ifn-gamma-producing</b> cells were located mainly in the follicular areas .	0.007	0.287
IL2	IFNG	pbl of ms patients produced more pro-inflammatory cytokines , <b>il-2</b> , <b>ifn-gamma</b> and tnf/lymphotoxin , and less anti-inflammatory cytokine , tgf-beta , during wk 2 to 4 in culture than pbl of normal controls .	0.006	0.283

Gene 1 Symbol	Gene 2 Symbol	Text	Before Calibration	After Calibration
NFKB1	TNFR	<a href="#">nf-kappab-dependent</a> reporter gene transcription activated by <a href="#">tnf</a> was also suppressed by calagualine .	0.005	0.276