



Title	Literature-derived, context-aware gene regulatory networks improve biological predictions and mathematical modeling
Author(s)	筒井, 真人
Citation	大阪大学, 2026, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/105665
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

A Doctorial Thesis

**Literature-derived, context-aware gene regulatory
networks improve biological predictions and
mathematical modeling**

(文献由来文脈依存的な遺伝子制御ネットワークによる生物学的予測と数理モデリングの改善)

Department of Biological Sciences

Graduate School of Science

The University of Osaka

大阪大学大学院理学研究科生物科学専攻

Masato Tsutsui

筒井 真人

January 2026

令和 8 年 1 月

Abstract

Many diseases are best understood not as abnormalities of single molecules, but as failures of gene regulatory networks (GRNs) and intracellular signaling circuits. Capturing disease-specific network structures and their dynamics is therefore essential for designing therapies and making accurate predictions. However, existing database- and literature-derived networks often treat regulations as fixed, context-independent interactions, failing to reflect dependence on disease, cell type, or experimental conditions. Moreover, although mechanistic mathematical models (e.g., ODE models) are powerful for understanding mechanisms and predicting drug responses, building them typically requires extensive literature review and manual network design, making the process labor-intensive and susceptible to subjective bias.

I propose a framework that assigns quantitative context-dependent weights to gene regulations extracted from the literature. A biological context is provided as a query, and each regulation is weighted by the semantic similarity between the query and the source literature. Network embedding is then applied to the weighted GRN to produce node embeddings that serve as prior knowledge for downstream analyses.

I validated the approach by showing that (1) differentially expressed genes cluster more tightly in the context-specific embedding space in large-scale transcriptome analyses, and (2) context-aware embeddings improve drug-target prediction over context-independent knowledge bases using L1000/Connectivity Map datasets. To bridge the gap between GRNs and ODE models, I further developed a semi-automated workflow that uses the context-specific GRN to retrieve and prioritize executable reaction candidates from resources such as BioModels, and employs a large language model (LLM) to reconcile naming and notation inconsistencies while assembling an executable ODE model. A breast cancer-specific signaling network case study demonstrates reduced modeling burden and workable model generation.

Overall, this study provides a unified framework that transforms literature corpora into context-dependent prior knowledge and links omics analysis, machine-learning prediction, and mechanistic modeling.

Abbreviation

AP-1: Activator Protein 1
API: Application Programming Interface
AUPRC: Area Under the Precision-Recall Curve
AUROC: Area Under the Receiver Operating Characteristic Curve
BERT: Bidirectional Encoder Representations from Transformers
DEG: Differentially Expressed Gene
EGF: Epidermal Growth Factor
EGFR: Epidermal Growth Factor Receptor
ERBB: ErbB receptor tyrosine kinase family
ERK: Extracellular Signal-Regulated Kinase
GEO: Gene Expression Omnibus
GO: Gene Ontology
GRN: Gene Regulatory Network
GSEA: Gene Set Enrichment Analysis
HER2/3: Human Epidermal growth factor Receptor 2/3 (=ERBB2/3)
HGNC: HUGO Gene Nomenclature Committee
KEGG: Kyoto Encyclopedia of Genes and Genomes
LINCS: Library of Integrated Network-based Cellular Signatures
LLM: Large Language Model
MAPK: Mitogen-Activated Protein Kinase
MEK: Mitogen-activated protein kinase kinase
MeSH: Medical Subject Headings
NER: Named Entity Recognition
NLP: Natural Language Processing
ODE: Ordinary Differential Equation
PCA: Principal Component Analysis
PI3K: Phosphoinositide 3-Kinase
PPI: Protein-Protein Interaction
RE: Relation Extraction
ROC: Receiver Operating Characteristic
RSK: Ribosomal S6 Kinase
UMAP: Uniform Manifold Approximation and Projection

Table of Contents

Abstract	i
Abbreviation	ii
Publication List	v
1. Introduction	1
1.1. Literature as data and the concept of bibliomics	1
1.2. Recent shift of BioNLP toward deep learning	1
1.3. Challenges in knowledge-graph utilization	2
1.4. Context dependence of GRNs	3
1.5. Knowledge utilization gap in mechanistic modeling	4
1.6. Core proposal: Literature-derived context-dependent GRNs	5
1.7. Evaluation perspectives of context-dependent GRNs	5
1.8. Positioning and contributions of this thesis	6
2. Methods	8
2.1. Data preparation and preprocessing	8
2.2. Context-dependent GRNs for downstream tasks	9
2.3. Gene embedding of context-dependent GRNs	11
2.4. Automated construction of mechanistic models and evaluation	14
3. Results	18
3.1. Overall design and evaluation strategy	19
3.2. Scoring context dependency using BERT model	19
3.3. Context-dependent GRNs capture disease-specific network structure and mechanisms	24
3.4. Embeddings derived from context-dependent networks align with spatial clustering of DEGs and reflect biological phenomena	31
3.5. Context-dependent gene embeddings improve drug–target prediction from gene expression 33	
3.6. Automated ODE model construction integrated with an LLM	35
3.7. Chapter Summary	43
4. Discussion	44
4.1. Contributions of this study	44
4.2. Biological relevance of literature-derived context-dependent GRNs	44

4.3. Recall-oriented retrieval and continuous weights: Controlling influence rather than eliminating noise	45
4.4. Implementation bottleneck: Multi-context queries and re-computational cost.....	46
4.5. Effectiveness of context-dependent knowledge and the challenge of integrating experimentally curated networks.....	46
4.6. Automated ODE model construction anchored by context-dependent GRNs and remaining challenges	47
4.7. Summary: Context-dependent GRNs as a promising paradigm for connecting literature knowledge with experiments and models	48
5. References	49
Acknowledgments.....	53

Publication List

1. **Tsutsui M.** and Okada M. DynProfiler: a Python package for comprehensive analysis and interpretation of signaling dynamics leveraged by deep learning techniques.
Bioinformatics Advances, Volume 4, Issue 1, 2024, vbae145
2. **Tsutsui M.**, Arakane K. & Okada M. Literature-derived, context-aware gene regulatory networks improve biological predictions and mathematical modeling.
Bioinformatics. Under review (revised manuscript submitted)

1. Introduction

1.1. Literature as data and the concept of bibliomics

Understanding how tens of thousands of genes in the human genome cooperate—and which breakdowns in this coordination lead to disease—is indispensable for designing effective therapeutic strategies. I believe this problem should be approached not merely as an accumulation of knowledge about individual molecules, but as an organized system of interactions, namely gene regulatory networks (GRNs). In recent years, large-scale datasets such as genomics and proteomics have played a decisive role as a foundation for understanding biological systems. At the same time, scientific outputs have continued to grow explosively, making it unrealistic for individual researchers to track the whole picture manually. PubMed alone now contains more than 39 million publications (National Library of Medicine, 2025b), and research articles themselves have reached a scale at which the literature itself has become a primary, machine-readable source of biomedical knowledge.

Against this backdrop, I have focused on the idea of “bibliomics”, which extends the “omics” framework beyond molecular measurements to the systematic analysis of the entire body of literature (the bibliome)(Grivell, 2002). They early emphasized the need for computational infrastructure to search for and extract useful information from the rapidly expanding literature and pointed to literature mining as a practical way forward.

Subsequently, organizational foundations such as MeSH indexing(National Library of Medicine, 2025a), along with information extraction technologies for life-science texts, have advanced, and the literature has increasingly been regarded not only as something to cite, but also as an object of analysis.

1.2. Recent shift of BioNLP toward deep learning

I regard the transition of BioNLP (biomedical natural language processing) from an era centered on simple co-occurrence to one centered on machine learning and deep learning as a major turning point. In other words, in pre-training, models first learn general patterns

from large-scale text and are then adapted to specific tasks. In particular, biomedical language models pre-trained on the literature, such as BioBERT (Lee *et al.*, 2020), substantially improved named entity recognition (NER) and relation extraction (RE), opening the way to identify gene names and their interactions from text with high accuracy. Concretely, a system can recognize mentions such as “EGFR” and “ERK” as standardized gene/protein names (NER) and then extract a directed statement such as “EGFR activates ERK” (RE). This makes it possible to turn narrative findings into structured, computable relations at a scale that is difficult to achieve by manual curation alone.

Moreover, domain-specific pretraining represented by BiomedBERT (Gu *et al.*, 2021) directly addresses distribution-shift issues and further boosts performance on biomedical text processing. In addition, PubTator 3.0 (Lai *et al.*, 2023; Lee *et al.*, 2020) provides a large-scale, pre-annotated resource in which PubMed articles are already tagged with recognized and normalized entities (e.g., genes) and extracted relations, including directed regulatory statements such as inhibition and activation, thereby making literature-derived knowledge readily usable at scale.

1.3. Challenges in knowledge-graph utilization

I view recent progress in biomedical literature information extraction as having accelerated the construction and curation of large-scale structured resources, including GRN-related databases such as OmniPath (Türei *et al.*, 2016) and STRING (Szklarczyk *et al.*, 2023), dramatically expanding both the volume and scope of structured knowledge accessible to researchers. Diverse types of knowledge—including gene regulation, protein–protein interactions, and relationships between diseases and drugs—have been integrated in an interoperable manner across resources (e.g. OmniPath, STRING, and PrimeKG (Chandak *et al.*, 2023)), and in recent years have increasingly been accumulated and maintained at scale as knowledge graphs.

What I am particularly conscious of, however, is that systematic efforts to leverage such well-curated knowledge graphs for specific downstream tasks have only begun to intensify very recently. For example, frameworks that propagate drug-related information over knowledge graphs to propose candidate disease treatments (Huang *et al.*, 2024), as

well as approaches that inject structured knowledge into neural networks as an inductive bias—so the model is guided by known biological relationships—supporting more interpretable classification. (Hartman *et al.*, 2023), illustrate a rapidly expanding direction of using structured knowledge as a prior for some downstream tasks.

At the same time, I consider a fundamental limitation of this trend to be the fact that many knowledge graphs treat edges as boolean (present/absent), largely ignoring both the confidence of regulatory relationships and their context dependence. In the current era of explosive growth in published findings, it is intuitively implausible to treat a frequently reported gene regulation and a regulation reported only very rarely as equally reliable. However, the number of reports does not automatically imply higher confidence, because reporting frequency is shaped by field bias and scientific attention and therefore may not reflect reliability. Accordingly, I see a research opportunity in connecting knowledge graphs to downstream tasks by explicitly representing uncertainty and context dependence, instead of treating edges as equally reliable.

1.4. Context dependence of GRNs

I considered context dependence to be essential because GRNs cannot be treated as a single, universal network: the strength and functional relevance of a regulatory interaction can differ substantially across diseases, tissues, cell types, perturbations, and timescales. Consequently, relationships with strong overall evidence in the literature can still be weak or irrelevant in a particular biological context.

For instance, while the p53–MDM2 axis is positioned as a central regulatory relationship in cancer—playing key roles in the DNA damage response and cell-cycle control—it may not act as an equally dominant factor in other domains such as neurological or autoimmune diseases. The key point here is not whether an interaction exists, but how strongly it matters in that context—its quantitative, causal, and controlling influence. A GRN is most useful when it supports context-specific prediction and intervention, not simply by listing interactions. Networks that ignore interaction strength therefore provide weaker guidance in downstream analyses.

Nevertheless, many conventional literature-mining approaches to network construction still treat interactions as binary links or weight them only by citation counts,

and thus fail to adequately capture context differences (Gill *et al.*, 2024). For these reasons, I aimed to build context-specific priors by assigning context-dependent weights to regulatory edges from the literature, and to test their utility quantitatively in downstream analyses.

1.5. Knowledge utilization gap in mechanistic modeling

I consider the problem of context dependence to be particularly critical in mechanistic mathematical modeling (e.g., ODE models) for understanding disease mechanisms and predicting drug responses. Model construction typically begins with determining the network structure (reactions and interactions). However, this process inevitably involves multiple layers of decisions—selecting which papers to consult, deciding which interactions to include or exclude, and incorporating experimental context such as disease, cell type, and stimulation—making it both labor-intensive and highly dependent on human expertise. Consequently, it is challenging to incorporate context in a systematic way when specifying mechanistic model structure.

This challenge has long been recognized in systems biology (Fröhlich *et al.*, 2018; Hill *et al.*, 2017). Recent frameworks address it by integrating pathway resources or prior knowledge networks with omics data to infer and refine network structure (Ruscone *et al.*, 2025; Rodriguez-Mier *et al.*, 2025). These approaches represent important progress in that they treat knowledge not merely as a reference, but as a principled guide that constrains and directs structure inference, thereby improving the reproducibility and efficiency of structure determination.

Nevertheless, in many existing frameworks, context specificity is entrusted primarily to omics data: prior knowledge is treated as context-agnostic, and context is imposed by selecting a subnetwork that best fits the measurements. This makes context modeling constrained by data availability and limits applicability in data-scarce settings. Moreover, although the literature provides rich contextual descriptions that could be used to shape prior knowledge directly, systematic, scalable ways to do so remain underexplored.

For this reason, I argue that important procedures are still missing—namely, context-aware tailoring of prior knowledge, explicit quantification of context-dependent

relevance for individual edges, and conversion into downstream-ready representations.

1.6. Core proposal: Literature-derived context-dependent GRNs

In this thesis, I propose a context-adaptive literature-mining framework that constructs gene regulatory networks (GRNs) from large-scale biomedical literature while explicitly conditioning prior knowledge on biological context. Specifically, I represent the target biological context as a query (e.g., a disease, cell type, or experimental condition) and treat regulatory relationships extracted from the literature as evidence whose strength depends on how relevant the source publication is to the query. In this formulation, the same regulatory edge is not assumed to be uniformly valid across all settings; instead, its contribution adapts to the target context through quantitative, context-dependent weighting.

A prerequisite is to define “semantic relevance” between a free-text query and each publication. Using a BERT-based model, I represent both the query and each paper as fixed-length numeric vectors that capture their meaning (sentence embeddings). This makes it possible to quantify how well a paper matches the query with a similarity measure such as cosine similarity.

The framework consists of three components. First, given a query describing a disease, cell type, or experimental condition, gene regulatory relationships are comprehensively extracted from literature relevant to that context. Second, the semantic relevance between the query and each publication is quantified and used to assign probabilistic weights to the extracted relationships, yielding a weighted GRN tailored to the query-defined context. Third, network embedding is applied to the resulting weighted network to obtain continuous representations that are directly usable in downstream statistical analyses, machine learning, and knowledge integration.

1.7. Evaluation perspectives of context-dependent GRNs

While the proposed framework—weighting literature-derived regulatory relations by semantic context–literature relevance while also utilizing differences in reporting frequency—appears intuitively promising, it is not self-evident how much it helps, for

what purposes, and under which evaluation criteria. I therefore set out to characterize its practical value in a task-oriented manner, rather than assuming usefulness from the design alone.

To assess the effectiveness of the proposed framework, I evaluate it from three perspectives. First, biological validity: I test whether the literature-derived networks are consistent with experimentally observed expression changes. Specifically, using approximately 2,500 transcriptomes across 68 diseases in DiSignAtlas (Zhai *et al.*, 2024), I evaluate whether differentially expressed genes become more proximal and form tighter clusters in the context-specific embedding space, and demonstrate an advantage over networks constructed without introducing context.

Second, utility as prior knowledge: I evaluate the impact on a drug-target gene prediction task. Using L1000 expression profiles (Subramanian *et al.*, 2017), I assess whether context-specific embeddings improve the recovery of true target genes of drugs among top-ranked candidates (e.g., recall within the top 5%) compared with alternative approaches based on context-independent knowledge bases such as FROGS (Chen *et al.*, 2024) and STRING (Szklarczyk *et al.*, 2023).

Third, support for mechanistic modeling: I examine whether context-specific GRNs can contribute to semi-automated construction of ODE models. A key issue here is the formal gap between GRNs and executable mechanistic models: GRNs encode signed regulatory relations (activation/inhibition), whereas ODE models require executable reaction specifications (species, stoichiometry, and rate laws). To address this gap, I design a framework in which the context-specific GRN serves as an anchor to retrieve and prioritize candidate reactions from BioModels (Glont *et al.*, 2020). I further employ a large language model (LLM) to reconcile name variations and notational inconsistencies and to integrate heterogeneous sources into a biologically coherent model. Using breast cancer-specific ErbB receptor signaling as a case study, I demonstrate that an interpretable ODE model can be generated with minimal manual effort.

1.8. Positioning and contributions of this thesis

The central contribution of this thesis is to explicitly introduce the previously underemphasized axis of context dependence into BioNLP research, and to transform vast

literature corpora into disease- and condition-specific prior knowledge that can be used consistently across network analysis, machine learning, and mechanistic mathematical modeling. Rather than accumulating literature knowledge as a mere collection of edges, I reconstruct it so that edge weights adapt to a query-defined context, and I quantitatively validate both its biological plausibility and its downstream utility. Through this approach, I aim to provide a foundation for making knowledge integration and model construction—processes that often rely on labor-intensive curation—more objective and scalable. Ultimately, I hope to present an implementable form of bibliomics that can accelerate disease understanding and the design of therapeutic strategies.

2. Methods

2.1. Data preparation and preprocessing

Query-literature similarity definition

I define a query as a short natural-language description representing a biological context of interest (e.g., disease, cell type, or experimental condition). Because the query is used for semantic retrieval, any textual description can serve as a context (including more complex statements such as “non-response to drug X under condition Y”); to define a new context, users provide a concise textual query and, optionally, constraints (e.g., time window, species). Specifically, for disease contexts, the query consisted of the MeSH-normalized disease name (National Library of Medicine, 2025a) and its associated description. For cell line contexts, I used textual descriptions provided by the LINCS project (Subramanian *et al.*, 2017) corresponding to each cell line. Similarity was then calculated against a corpus of PubMed titles and abstracts created from the annual baseline of 2024, which includes articles published up to December 2023. Sentences within the literature were processed using a sliding window approach where the window spanned two consecutive sentences and was shifted by one sentence at a time. Each window was encoded using SentenceTransformers (Reimers and Gurevych, 2019, 2020) library with the S-PubMedBert-MS-MARCO model (Deka *et al.*, 2022), which was selected based on its domain-specific knowledge acquired during the pre-training of its base model, BiomedBERT (Gu *et al.*, 2021). Sentence embeddings were normalized using z-score standardization across the entire literature corpus (Chen *et al.*, 2020). Query-document similarity was calculated as the maximum cosine similarity between the normalized query vector and any sentence vector within the document (Arnulf *et al.*, 2014).

Accuracy evaluation of BERT-based literature retrieval

To evaluate the accuracy of literature retrieval for disease-related queries, we employed official MeSH descriptors (Rogers, 1963) as queries (2,941 terminal “Category C” diseases) and designated literature associated with these terms as positive examples, and

all others as negative examples. The descriptors were supplemented with description annotation from the MeSH Descriptor Data (National Library of Medicine, 2025a).

Retrieval performance was summarized by the Area Under the Receiver Operating Characteristic curve (AUROC), with an optimal threshold of 0.21 identified, leading us to define literature with similarity scores exceeding 0.2 as relevant literature. For comparison of retrieved literature counts, we employed MeSH tags and the PubMed API. In this study, “MeSH-tag” denotes records explicitly indexed with the target MeSH descriptor, whereas “PubMed API” denotes results obtained by submitting the MeSH term as a plain-text query (no field qualifier). Because the PubMed API applies Automatic Term Mapping and field expansions, this query mode integrates MeSH-based matching with keyword/text search and can also return in-process citations not yet MeSH-indexed. Accordingly, the PubMed-API retrieval is expected to include the MeSH-tagged set.

Construction of gene regulatory networks

Gene regulatory relationships were obtained from the pre-annotated PubTator3 table (relation2pubtator3.gz; accessed 2024-12-18), restricting to human and mouse records. Mouse genes were mapped to human orthologs using Ensembl, resolving synonyms via HGNC. These regulations were treated as undirected graphs in our analysis for analytical simplicity. The reporting frequency f_{ij} of each gene pair regulation (g_i, g_j) is defined as the number of distinct publications (distinct PMIDs) in which the relation between g_i and g_j is mentioned. We then normalize f_{ij} per million publications using the total number of publications N , then log-transformed to serve as edge weights:

$$w_{ij} = \log \left(1 + \frac{f_{ij} \times 10^6}{N} \right)$$

I scale by 10^6 so that, for typical contexts with 10^4 - 10^5 relevant papers, edge-level values fall in the 10-100 range and, after log transform, in a numerically convenient 1-10 range.

2.2. Context-dependent GRNs for downstream tasks

Construction of context-dependent GRNs

Using the similarity scores defined in the “Query-Literature Similarity Definition section” as weights, we aggregated the reporting frequencies of each gene pair, normalized per

million publications, and applied log transformation:

$$w_{ij}^{(c)} = \log \left(1 + \frac{\sum_{d \in D} (\text{sim}(c, d) \times f_{ij}^{(d)}) \times 10^6}{\sum_{d \in D} \text{sim}(c, d)} \right)$$

where c represents the context (such as disease), D denotes the complete literature collection, and $f_{ij}^{(d)}$ indicates the reporting frequency of gene pair (g_i, g_j) in document d .

Post-processing of context-dependent GRNs

For every one of the 2,941 terminal MeSH diseases I merged 880,469 literature-derived gene regulations (20,824 genes) into a context–relationship matrix:

$$M \in \mathbb{R}^{2941 \times 880469}$$

whose entries are the log-transformed context-specific weights $w_{ij}^{(c)}$ for gene pair (g_i, g_j) in disease context c . This high-dimensional matrix was compressed to 200 dimensions using Principal Component Analysis (PCA) and visualized using Uniform Manifold Approximation and Projection (UMAP).

Gene importance was computed as degree-penalized PageRank (Brin and Page, 1998; Feng *et al.*, 2018) with penalty coefficient $\beta = 1.0$ by default); sensitivity analyses were performed at $\beta = 0$ and $\beta = 1.5$. Gene Set Enrichment Analysis (GSEA) (Subramanian *et al.*, 2005) was performed using gene importance scores to annotate the disease-specific gene interaction networks.

Disease category prediction

Logistic regression was performed using 200-dimensional PCA features ($\mathbf{x}_d$) to predict major categories of the MeSH tree for the 2,941 diseases. Accuracy was evaluated through 5-fold cross-validation.

Drug repurposing accuracy evaluation

The PrimeKG dataset (Chandak *et al.*, 2023) was utilized to establish correspondences between diseases and compounds, with train, validation, and test splits following the methodology described in the TxGNN paper (Huang *et al.*, 2024). Disease features were extracted as corresponding row vectors $\mathbf{m}_d$ from the context-dependency matrix and compressed to 200 dimensions using PCA:

$$\mathbf{x}_d = PCA(\mathbf{m}_d) \in \mathbb{R}^{200}$$

For compounds, co-occurrence frequency matrices $\mathbf{F} \in \mathbb{R}^{N_c \times N_g}$ between compounds and genes across all literature were constructed using the PubTator3 database (Wei *et al.*, 2024), where N_c and N_g represent the number of compounds and genes, respectively:

$$\mathbf{F}_{cg} = \text{co-occurrence frequency}(c, g)$$

Each compound's row vector $\mathbf{f}_c$ was compressed to 200 dimensions using PCA:

$$\mathbf{x}_c = PCA(\mathbf{f}_c) \in \mathbb{R}^{200}$$

According to the PrimeKG dataset, relationships between diseases and compounds $t \in \{\text{indication, contraindication, off-label use}\}$ were presented as 3-dimensional one-hot vectors $\mathbf{x}_t$. The input features for LightGBM (Ke *et al.*, 2017) consisted of concatenated 403-dimensional vectors combining disease, compound, and relationship type information:

$$\mathbf{x}_{input} = [\mathbf{x}_d; \mathbf{x}_c; \mathbf{x}_t] \in \mathbb{R}^{403}$$

The task was formulated as binary classification, assigning label $y = 1$ to positive examples (actual disease-compound-relationship triplets) and $y = 0$ to negative examples. LightGBM was employed to predict probabilities $p(y = 1 \mid \mathbf{x}_{input})$ using default hyperparameters. Negative example generation followed the TxGNN methodology. The Area Under the Precision-Recall Curve (AUPRC) was used as the evaluation metric, as it is more appropriate than AUROC for imbalanced datasets where positive examples (actual drug-disease pairs) constitute a small fraction of the total data.

2.3. Gene embedding of context-dependent GRNs

Node Embedding Procedures with Degree-Penalized Transition Matrix

The networks employed in this study exhibit scale-free properties, which pose challenges for conventional random walk-based network embedding methods such as DeepWalk (Perozzi *et al.*, 2014) and Node2vec (Grover and Leskovec, 2016), as these approaches suffer from strong bias toward hub nodes and insufficient learning of low-degree node information. To address this limitation, I implemented the degree-penalized transition matrix (Feng *et al.*, 2018).

Using the degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, and common neighbor matrix $\mathbf{C}$, I constructed the penalty matrix $\mathbf{W}$ as follows:

$$\mathbf{W} = (\mathbf{D}^{-\beta})^T (\mathbf{C} + \mathbf{A}) (\mathbf{D}^{-\beta})$$

where β is a hyperparameter controlling the strength of suppression for high-degree nodes, set to $\beta = 1.0$ in this study. The resulting matrix $\mathbf{W}$ was normalized, and transition probabilities from 1-step to K -step were accumulated to create a multi-step transition matrix $\mathbf{P}$:

$$\mathbf{P} = \sum_{k=1}^K \mathbf{T}^k \text{ (with } K = 3\text{)}$$

where $\mathbf{T}$ represents the row-normalized transition matrix derived from $\mathbf{W}$. For each node, one positive example node was sampled based on $\mathbf{P}$, and five negative example nodes were sampled based on inverse probabilities to create positive-negative pairs.

The embedding model employed a skip-gram-format dot product model learning 256-dimensional embedding vectors with binary cross-entropy loss and Adam optimizer with an initial learning rate of 0.001 and minimum learning rate of 10^{-6} .

Evaluation of node embedding similarity in differentially expressed genes

I evaluated whether disease-specific differentially expressed gene (DEG) sets mapped coherently within the embedding space. DEG lists were derived from GEO-sourced transcriptome datasets curated in the DiSignAtlas database (Zhai *et al.*, 2024). In total, 2,553 datasets covering 68 distinct diseases were analyzed. For each dataset, I calculated DEGs against their respective disease controls, retained genes whose adjusted P-value was < 0.05 , assigned each of these genes the mean of (i) its rank by $-\log_{10}(\text{adjusted P})$ and (ii) its rank by $\log_2 |\text{fold-change}|$, and then defined DEG sets by taking the top-k genes (with several k values examined) for use in the subsequent cosine-similarity analysis.

As the embedding spaces from which gene vectors were drawn, I compared two graph-based representations: (1) unweighted graphs, assuming uniform weights for all edges, and (2) conditional graphs based on disease context (detailed in the Context-Dependent GRNs section). For each dataset, the corresponding DEG set G_{DEG} was extracted from the embedding space, and the average pairwise cosine similarity was calculated:

$$\bar{s}_{DEG} = \frac{2}{|G_{DEG}|(|G_{DEG}| - 1)} \sum_{i < j} \frac{\mathbf{e}_i^T \cdot \mathbf{e}_j}{\|\mathbf{e}_i\| \cdot \|\mathbf{e}_j\|}$$

Z-scores were computed using statistics from cosine similarities across all gene pairs. As a control experiment, gene sets of identical size to the DEGs were randomly selected, and

the same metrics were calculated. This methodology confirmed that embeddings derived from conditional graphs exhibited significantly higher spatial consistency for disease-specific DEG sets.

Drug-target predictive model construction and evaluation

To evaluate whether context-dependent GRNs can improve predictive accuracy in drug target identification, I implemented a drug target prediction model using the L1000 dataset (Subramanian et al., 2017). Recent study has demonstrated that incorporating gene relations from GO (The Gene Ontology Consortium et al., 2021) as prior knowledge can enhance prediction performance in drug discovery tasks (Chen et al., 2024). Building on this foundation, I hypothesized that context-dependent gene representations tailored to specific cell types would provide superior prior knowledge compared to non-context-dependent approaches. Following Chen et al.'s model architecture, I replaced only the gene embedding vectors while keeping the rest of the network unchanged. The model utilizes a Siamese neural network architecture where compound and gene expression signature pairs serve as inputs to predict the probability of target relationships. The network processes both input embedding vectors through shared fully connected layers $f(\cdot)$, computes their element-wise product, and feeds the result to a final classifier:

$$p(target|c, g) = \sigma \left(W \cdot \left(f(\mathbf{x}_c) \odot f(\mathbf{x}_g) \right) + b \right)$$

where $\mathbf{x}_c$ and $\mathbf{x}_g$ represent compound and gene signatures, $\odot$ denotes element-wise multiplication, and σ is the sigmoid function. Binary cross-entropy was employed as the loss function. For each compound-gene pair (c, g) from our dataset of 2,340 compound-gene pairs (1,438 compounds and 499 targets), multiple gene expression signatures existed across 83 different cell lines, so only compound and shRNA/cDNA pairs acquired from identical cell types were used as inputs.

Then, I treated cell type information from the LINCS project (Subramanian et al., 2017) as context and utilized independently learned gene embedding vectors for each cell type. Gene embeddings were generated from our context-dependent GRNs using Node2Vec with degree-penalized sampling to account for scale-free network properties, thereby reflecting cell type-specific differences in gene function within the embeddings. Evaluation was conducted through 5-fold cross-validation, comparing the proportion of

known targets ranking within the top-k positions (recall@top-k).

Prediction scores for identical (c, g) pairs across multiple cell types were integrated using rank-based aggregation, adopting the minimum normalized rank for each gene as the representative score, following the evaluation strategy used in the prior work (Zhai *et al.*, 2024). All model architectures and training conditions were standardized to ensure that only differences in embedding vector information representation affected the results.

2.4. Automated construction of mechanistic models and evaluation

Breast cancer-specific mathematical model construction

To demonstrate automated model construction, I developed a pipeline that integrates literature-derived networks, experimental data, and fragmented equations from BioModels (Glont *et al.*, 2020) using AI agents to address notation inconsistencies and pathway connectivity gaps.

PageRank centrality scores and 256-dimensional gene embedding vectors were obtained from the breast cancer-specific context-dependent GRN constructed using the aforementioned methodology. I used the transcriptome data of MCF-7 breast cancer cells stimulated with a growth factor, Heregulin, targeting ErbB receptor (Nagasato-Ichikawa *et al.*, 2024). DEGs were identified using edgeR with criteria of $\log_2|FC| > 2$ and adjusted p-value < 0.001 .

Reaction equation retrieval and prioritization from BioModels

All equations were retrieved from BioModels Parameters (Glont *et al.*, 2020), and gene symbols were mapped to components of each reaction equation. Three scores were calculated for each reaction equation r :

Gene importance score:

$$S_1(r) = rank\left(\frac{1}{|G_r|} \sum_{g \in G_r} PR(g)\right)$$

Inter-gene similarity score:

$$S_2(r) = rank\left(\frac{2}{|G_r|(|G_r| - 1)} \sum_{i < j} \cos(\mathbf{e}_i, \mathbf{e}_j)\right)$$

DEG-related score:

$$S_3(r) = \text{rank} \left(\frac{1}{|G_r| \cdot |G_{DEG}|} \sum_{g \in G_r} \sum_{d \in G_{DEG}} \cos(\mathbf{e}_g, \mathbf{e}_d) \right)$$

where G_r represents the gene set contained in reaction equation r , $PR(g)$ is the PageRank score of gene g in the contextualized network, and $\text{rank}(\cdot)$ denotes ascending rank transformation. The comprehensive score was calculated as a weighted average of scores 1-3, with score 3 weighted 10 times higher:

$$S_{total}(r) = \frac{S_1(r) + S_2(r) + 10 \cdot S_3(r)}{12}$$

The top 30 reaction equations based on the integrated score were selected.

Parameter fitting and simulation analysis

The mathematical model constructed using the agent-based framework was fitted to experimental data to assess its ability to recapitulate signaling dynamics.

The experimental data was obtained from previous studies that investigated the phosphorylation dynamics of several signaling molecules within MCF-7 cells upon Heregulin and EGF stimulation, including the phosphorylation of EGFR and HER2 (Nagashima *et al.*, 2007), and Shc, MEK, ERK, AKT (Birtwistle *et al.*, 2007). Since the experimental data in these studies were only provided as figures, I used the coordinates of the datapoints in the plots to estimate the experimental values. The estimated experimental values were subsequently normalized using the maximum value for each species before being used for the estimation of the model parameters.

To test the model's predictive powers, I plotted the simulated phosphorylation dynamics of several molecules alongside experimental data that was not used during the parameter estimation. Here, experimental data for phosphorylated HER3 (Nagashima *et al.*, 2007) and RSK and FOS (Nakakuki *et al.*, 2010) were obtained and similarly normalized using the maximum values for each species within the simulated timeframe.

To associate the model species to their corresponding experimental data, I defined model observables accordingly (Table 1).

The parameter estimation was conducted using BioMASS (Imoto *et al.*, 2020), with which a total of 30 parameter sets were obtained by differential evolution with the

objective to minimize the mean square error loss between the simulated and experimental values. Further information on the experimental setup for this section can be found within the GitHub repository of this paper (<https://github.com/okadalabipr/context-dependent-GRNs>).

Table 1. Definition of observables for the LLM-assisted model.

Phosphorylated-EGFR	$\frac{2 \times [\text{EGF_EGFR_p}] + 2 \times [\text{EGFR_EGFR_p}] + [\text{EGFR_HER2_p}]}{[\text{EGF}] + [\text{EGF_EGFR}] + 2 \times [\text{EGF_EGFR_EGF_EGFR}] + 2 \times [\text{EGF_EGFR_p}] + 2 \times [\text{EGFR_EGFR_p}] + [\text{EGFR_HER2_p}]}$
Phosphorylated-HER2	$\frac{[\text{Heregulin_HER3_HER2_p}] + [\text{EGFR_HER2_p}]}{[\text{HER2}] + [\text{Heregulin_HER3_HER2}] + [\text{Heregulin_HER3_HER2_p}] + [\text{EGFR_HER2}] + [\text{EGFR_HER2_p}]}$
Phosphorylated-Shc	$\frac{[\text{Shc_p}]}{[\text{Shc}] + [\text{Shc_p}]}$
Phosphorylated-MEK	$\frac{[\text{MEK_p}]}{[\text{MEK}] + [\text{MEK_p}]}$
Phosphorylated-ERK	$\frac{[\text{ERK_p_cytoplasm}] + [\text{ERK_p_nucleus}]}{[\text{ERK}] + [\text{ERK_p_cytoplasm}] + [\text{ERK_p_cytoplasm}]}$
Phosphorylated-AKT	$\frac{[\text{AKT_p}]}{[\text{AKT_inact}] + [\text{AKT_p}]}$
Phosphorylated-HER3	$\frac{[\text{Heregulin_HER3_HER2_p}]}{[\text{HER3}] + [\text{Heregulin_HER3}] + [\text{Heregulin_HER3_HER2}] + [\text{Heregulin_HER3_HER2_p}]}$
Phosphorylated-RSK	$[\text{RSK_p}]$
Phosphorylated-FOS	$[\text{cFOS_p}]$

LLM-based multi-agent system for equation integration

While the top 30 selected equations contained relevant biological components, they were fragmented and contained inconsistent notation (e.g., ErbB2 vs HER2) and lacked pathway connectivity. To address these limitations, I employed a multi-agent LLM system to integrate equations into executable models. A four-agent system was implemented based on GPT4-o3, with each agent equipped with dedicated system prompts and web

search functionality (Tavily Search API). Reaction equations were described in the Text2Model format (Imoto *et al.*, 2022). The format enables the definition of a biochemical reaction network using a standardized, human-friendly manner (e.g., $A + B \rightarrow C \mid k=0.1$), which can be automatically converted into an executable mathematical model. Although the format was originally developed to assist biologists when constructing mathematical models, it was utilized as an interface for LLMs in our study by explicitly specifying the syntax in each agent's system prompt. The system comprises the following four stages:

- (i) proposing major signaling pathways and readouts using a LangGraph-based workflow, following the empirical use of the framework (Wang and Duan, 2025), where the LLM receives 30 pre-selected reaction equations and generates four search queries. These queries are executed using the Tavily Search API, and the LLM summarizes the retrieved content to identify key signaling pathways and phenotypic readouts relevant to the given reactions.
- (ii) normalizing component names using the LLM (e.g., ErbB2, HER2 $\rightarrow$ ERBB2), connecting reaction equations using the pathways and readouts from stage (i) as anchors, integrating duplicate reactions, and identifying isolated reaction groups as subnetworks
- (iii) automatically detecting input-output components of each subnetwork, with the LLM inferring connection points between them and generating intermediate reactions as needed or removing subnetworks to construct an integrated model with complete connectivity.
- (iv) generating for search queries related to feedback mechanisms (i.e., regulatory loops where pathway outputs influence upstream components) and crosstalk (i.e., inter-pathway communication and signal integration). Based on the retrieved results, the LLM added reactions that reflect these mechanisms, which are critical for robust cellular responses in systems biology.

3. Results

This chapter presents the results of the computational framework for literature-derived context-dependent gene regulatory networks (context-dependent GRNs), which is the central topic of this study, in a stepwise manner from five perspectives: (i) context-matched literature retrieval, (ii) network construction and incorporation of disease mechanisms, (iii) consistency with gene-expression patterns in the network-embedding space, (iv) utility for drug-target prediction from gene expression, and (v) automated construction of ordinary differential equation (ODE) models integrated with a large language model (LLM). An overview of the entire framework is shown in Figure 1.

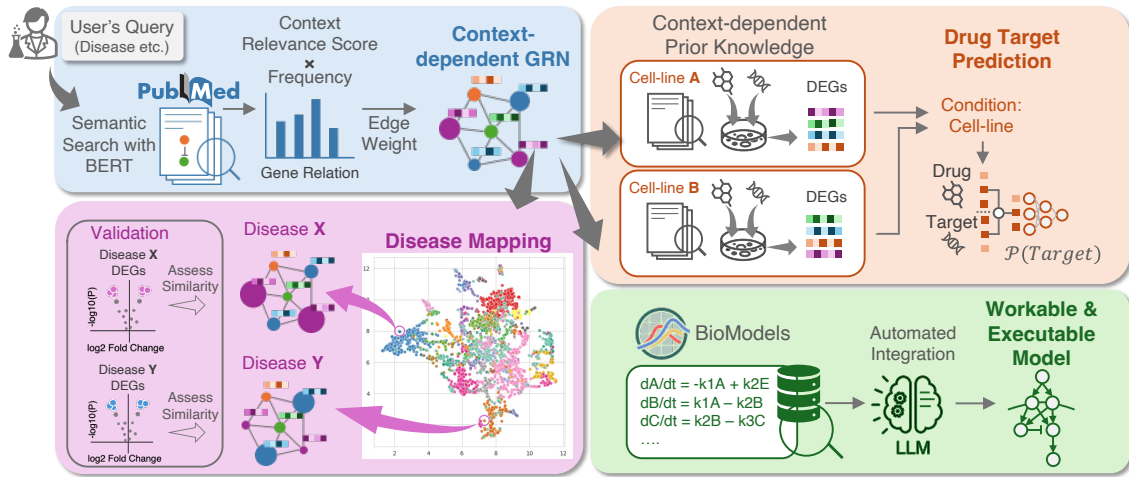


Figure 1. Overview of context-dependent GRN framework and applications

Overview of the comprehensive framework for constructing context-dependent GRNs from literature and their applications to diverse biological tasks: (1) validation of biological relevance through analysis of differentially expressed genes in embedding spaces, (2) drug target prediction using the network as prior knowledge, and (3) automated mathematical model construction by integrating context-dependent networks with large language models.

3.1. Overall design and evaluation strategy

My aim is not merely to extract gene relationships from the literature and assemble a network, but to represent—within a unified framework—how the importance of the same regulatory relationship can change with biological context, and to make this representation directly usable in downstream tasks such as expression-data interpretation, drug-target inference, and mechanistic model construction.

In this thesis, “context” denotes a set of conditions that define a biological setting, including disease, tissue or cell type, stimulation or environment, and observation timescale. Because these conditions can alter both how consistently a relation is reported and which pathways dominate, I treat context differences as a core factor in network construction. Accordingly, I quantify context–literature relevance, use it to weight edges in a GRN, and derive features such as embeddings and centrality from the resulting weighted network for downstream analyses.

3.2. Scoring context dependency using BERT model

BERT (Bidirectional Encoder Representations from Transformers) is a Transformer-based language model that interprets words simultaneously from both left and right contexts in a sentence, enabling representations that capture polysemy as well as word order and syntactic relations. In recent years, sentence embeddings built on BERT have been widely used to map sentences or paragraphs into fixed-length vectors, allowing us to calculate the semantic similarity between texts without relying on exact word overlap.

3.2.1 Defining “Ground Truth” using MeSH tags

To evaluate how well the proposed approach identifies literature relevant to a given context, I used disease-related MeSH tags assigned to PubMed articles as an approximate ground-truth label for context-matched papers. Specifically, I ranked papers by cosine similarity between the disease-query embedding and each paper’s sentence embedding, and examined whether papers carrying the corresponding MeSH tags were concentrated toward higher ranks.

When using disease descriptions and MeSH terms as queries, papers annotated with the corresponding MeSH tags consistently exhibited higher mean normalized similarity

scores (typically around 0.4) than papers associated with other MeSH tags (Figure. 2). Importantly, the method can distinguish between closely related subtypes such as “breast cancer” and “triple-negative breast cancer (TNBC),” while producing larger score gaps for mechanistically distant diseases such as “type 2 diabetes.” This pattern suggests that the approach leverages semantic proximity in the embedding space, rather than simple keyword matching.

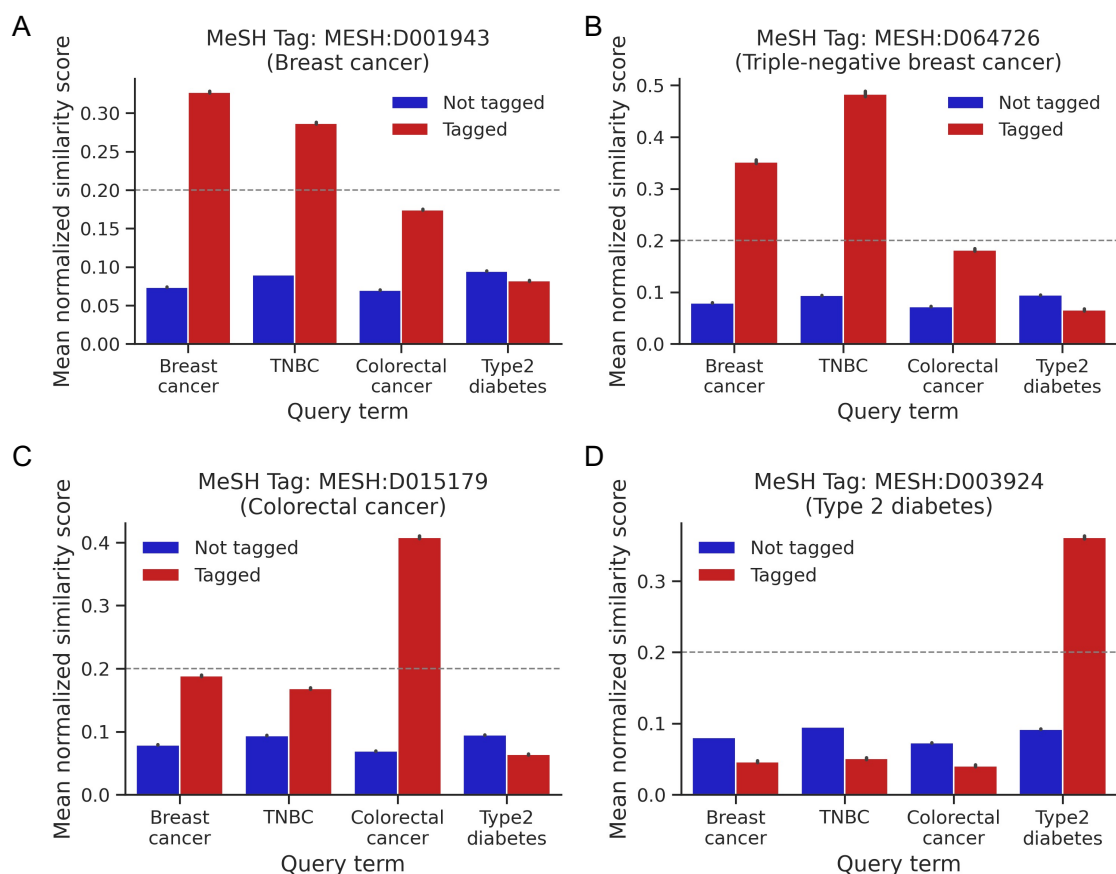


Figure 2. Context-specific literature retrieval using BERT embeddings for specific diseases (A-D) Average similarity scores for documents with (red) and without (blue) corresponding MeSH tags across different query diseases: (A) Breast cancer, (B) Triple-Negative Breast cancer, (C) Colorectal cancer, (D) Type 2 diabetes. X-axis shows query diseases used for similarity calculation.

3.2.2 Large-scale generalization across 2,941 diseases

To examine the generality of the proposed approach, I analyzed 2,941 diseases corresponding to the leaf nodes of MeSH disease categories and compared the similarity-score distributions between relevant and non-relevant literature. The two distributions were clearly separable (Figure. 3A), and the resulting classification performance reached AUROC = 0.96 (Figure. 3B). The optimal operating point (the point closest to the upper-left corner of the ROC curve) was around 0.21. Using a threshold of 0.2, BERT-based method detected more relevant documents than traditional MeSH-tag or PubMed API searches (Figure. 3C), likely due to its ability to capture semantic relationships and context beyond exact keyword matching.

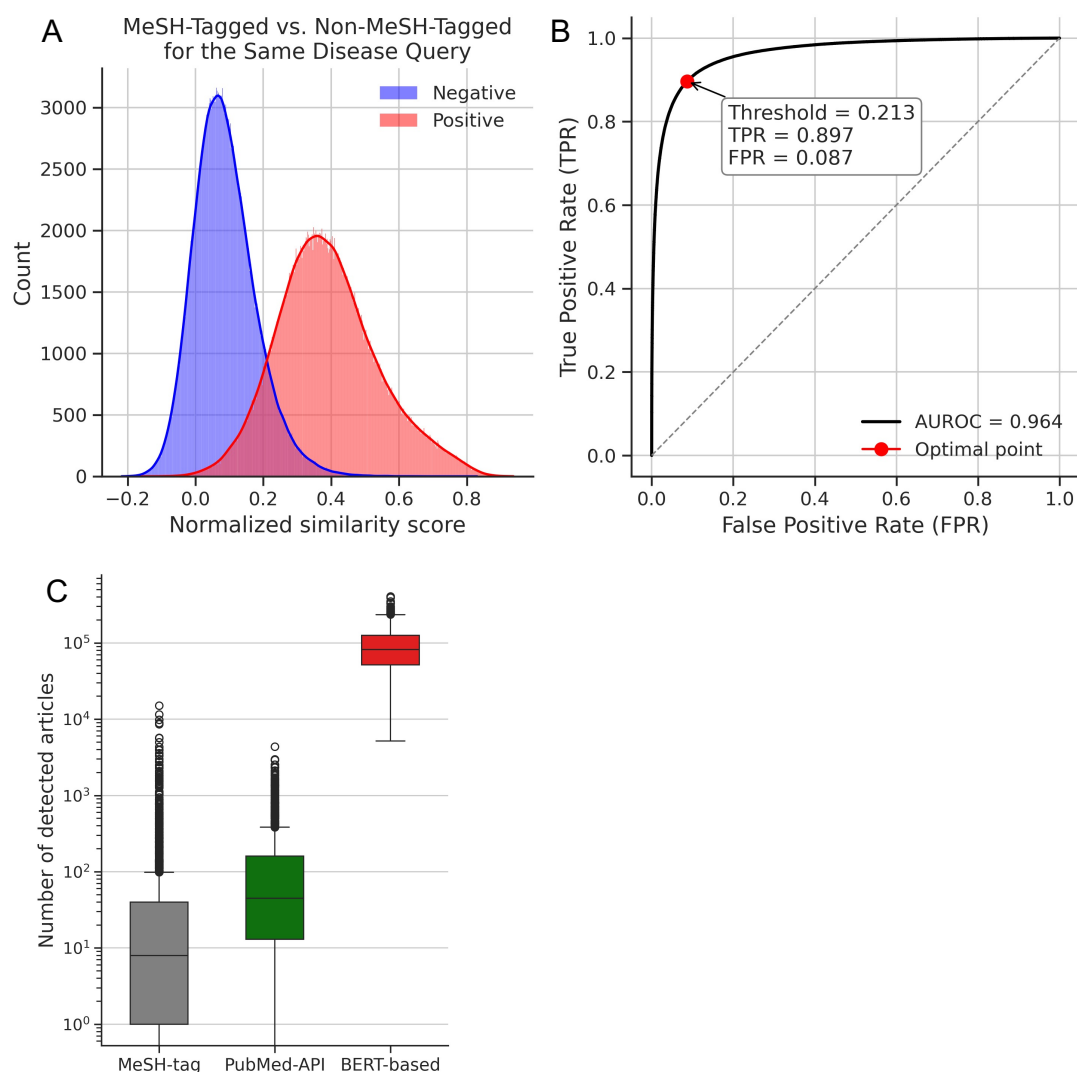


Figure 3. Context-specific literature retrieval using BERT embeddings for comprehensive set of diseases

(A) Distribution of similarity scores for documents with (red) and without (blue) corresponding MeSH tags across all diseases analyzed.

(B) ROC curve showing True Positive Rate versus False Positive Rate based on similarity scores. The optimal operating point (red dot) indicates the threshold where TPR - FPR is maximized.

(C) Number of documents retrieved by different search methods: MeSH Tag search, PubMed API search, and BERT-based approach.

3.2.3 Importance of domain-specific pretraining: superior performance to a general-purpose BERT

In the analyses above, I used a biomedical domain-specialized sentence-embedding model (S-PubMedBert-MS-MARCO) to compute query-publication similarity. To justify this choice, I compared it with a general-purpose model (multi-qa-mpnet-base-dot-v1) under comparable embedding/model scale. The general-purpose model achieved a lower performance (AUROC = 0.93; Fig. 4), supporting the practical importance of biomedical domain pretraining for literature retrieval. In the following sections, I use similarity scores from S-PubMedBert-MS-MARCO as the context-compatibility signal for edge-weighting in context-dependent GRNs.

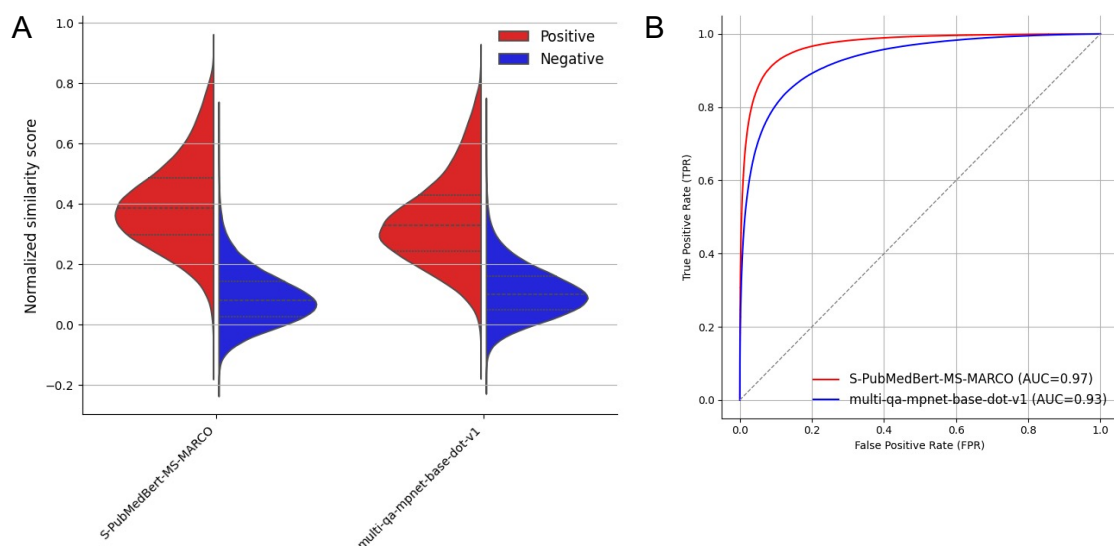


Figure 4. Validation of domain-specific BERT-based similarity scoring approach

(A) Comparison of similarity score distributions between domain-specific BERT model (S-PubMedBert-MS-MARCO) and general BERT model (multi-qa-mpnet-base-dot-v1). Both models have identical parameter size and embedding dimensions.

(B) ROC curve comparison between domain-specific BERT model (S-PubMedBert-MS-MARCO, red) with AUROC = 0.97 and general BERT model (multi-qa-mpnet-base-dot-v1, blue) with AUROC = 0.93. The superior performance of the domain-specific model highlights the importance of biomedical domain pre-training for literature retrieval tasks

3.3.Context-dependent GRNs capture disease-specific network structure and mechanisms

3.3.1 Constructing GRNs and topological differences in representative diseases

As a basis for gene regulatory relationships extracted from the literature, I used gene–gene relations annotated by PubTator3 as the underlying data source, and assigned context relevance weights based on the BERT-derived similarity scores introduced in the previous section. After normalization and log transformation, the weights were mapped onto edges, yielding GRNs in which the weight of the same gene pair can vary across contexts.

When comparing “breast cancer” and “type 2 diabetes,” the subnetwork induced by the top 10 genes ranked by PageRank centrality differed markedly: for example, TP53 was prominently positioned as a central node in the former, whereas Insulin/INS was central in the latter, consistent with known disease mechanisms (Figure. 5A, C). Moreover, GSEA based on centrality identified PI3K signaling and p53-related pathways in breast cancer and pancreatic β -cell–related pathways in type 2 diabetes, confirming that the core of each network aligns with established disease biology (Figure. 5B, D).

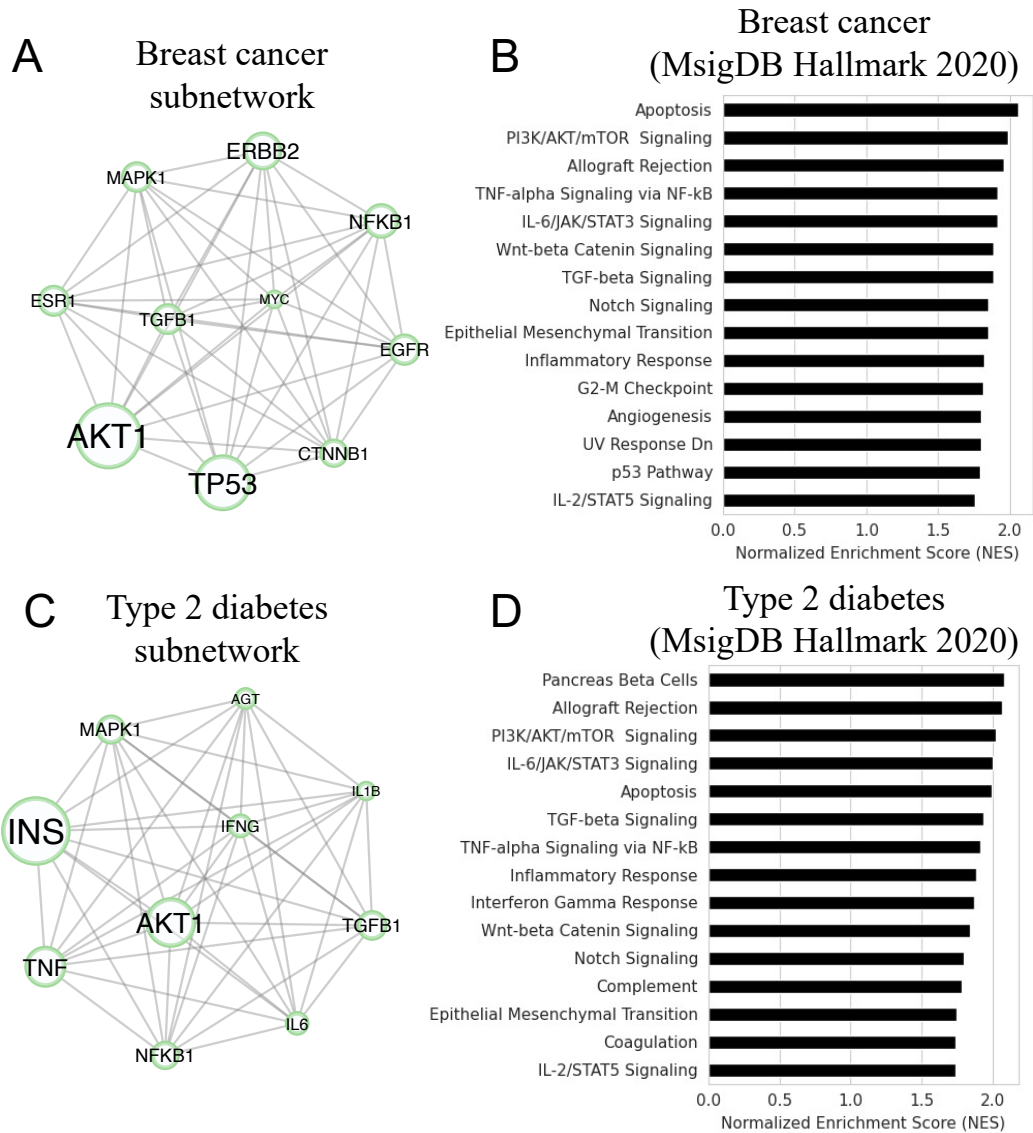


Figure 5. Context-dependent GRNs capture disease-specific characteristics

(A, C) Disease-specific subnetworks composed of the top 10 PageRank centrality genes for (A) breast cancer and (C) type 2 diabetes. Node size represents PageRank centrality scores.

(B, D) Gene set enrichment analysis results using PageRank centrality scores for (B) breast cancer and (D) type 2 diabetes.

3.3.2 Mitigating hub bias: Degree penalties clarify disease specificity

Because biological interaction networks are widely observed to exhibit scale-free topology—characterized by a small number of hub genes with very high degree—network evaluation and interpretation must explicitly account for the influence of such hubs. Indeed, both OmniPath, a molecular interaction database, and the literature-derived GRNs exhibited power-law distributions (Figure. 6), suggesting that networks built from literature share structural characteristics commonly observed in biological networks.

In such networks, centrality measures can become overly dominated by hubs, potentially diluting disease specificity. To address this, I introduced a degree-penalty term ($\beta = 0, 1.0, 1.5$) when computing centrality, and compared the overlap of GSEA results between diseases (Figure. 7). As β increased, the number of significantly enriched gene sets shared between breast cancer and type 2 diabetes decreased. Conversely, disease-specific terms—such as G2–M Checkpoint (breast cancer) and Pancreas Beta Cells (type 2 diabetes)—were extracted more clearly (Figure. 7), indicating that controlling hub bias can quantitatively improve disease specificity when connecting network analysis outputs to downstream interpretation. In the following analyses, I use $\beta = 1.0$.

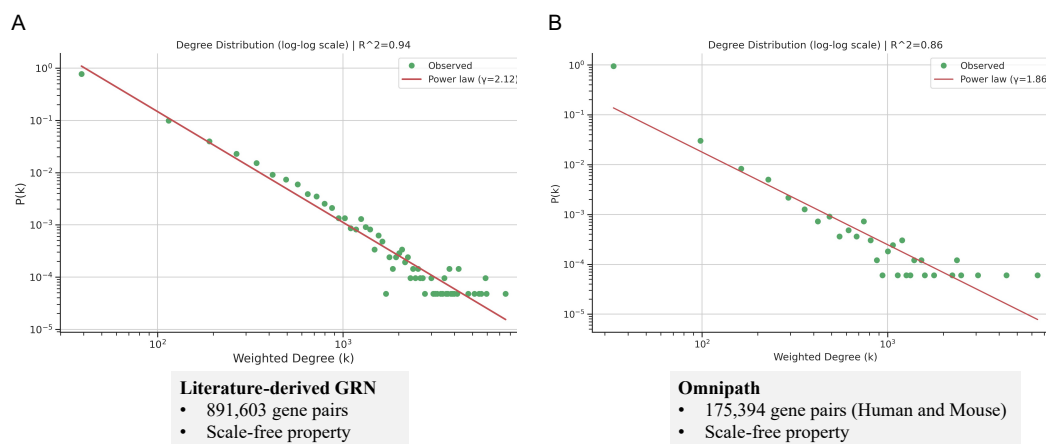


Figure 6. Scale-free properties of biological networks

Degree distribution plots demonstrating scale-free characteristics across different GRNs. **(A)** Literature-derived GRN from our approach. **(B)** OmniPath database network. X-axis represents node degree, Y-axis represents the fraction of nodes with that degree. While our literature-derived network and OmniPath exhibit power-law distributions characteristic of scale-free biological networks.

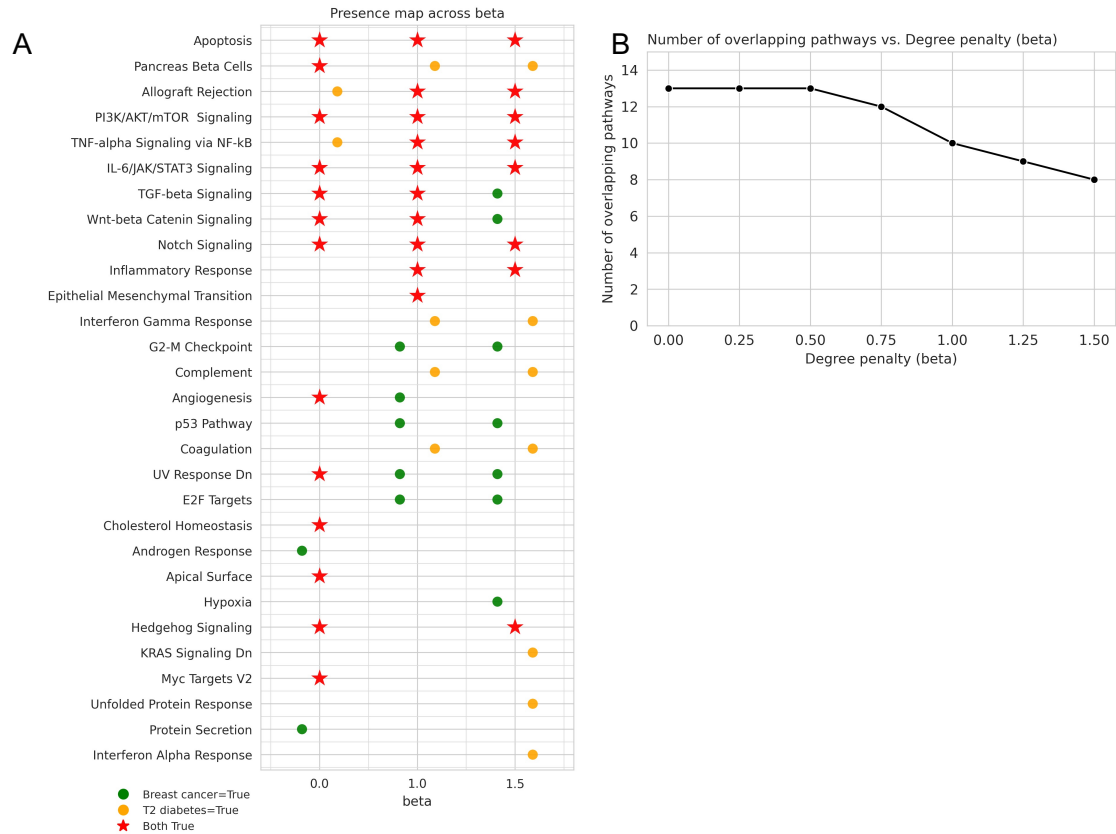


Figure 7. Degree-penalty reduces hub bias and clarifies disease-specific enrichment.

We compared degree-penalized centrality with $\beta = 0, 1.0$, and 1.5 . GSEA used PageRank centrality computed on context-dependent GRNs constructed separately for each β (see Methods).

(A) Gene-set terms significant at each β are plotted with β on the x-axis and individual terms on the y-axis. Terms significant only in Breast cancer are marked with green circles; terms significant only in Type 2 diabetes (T2D) with yellow circles; terms significant in both diseases with red stars. Higher β yields lower overlap and clearer disease specificity (e.g., G2–M Checkpoint for Breast cancer, Pancreas Beta Cells for T2D).

(B) Overlap ratio of the Top-15 significantly enriched terms between Breast cancer and T2D as a function of β . Increasing the penalty (larger β) reduces cross-disease overlap, and accentuates disease specificity—for example, G2–M Checkpoint emerges for Breast cancer, whereas Pancreas Beta Cells emerges for T2D.

3.3.3 Large scale disease map using context-dependent GRNs

I then expanded the construction to all diseases and aggregated networks for 2,941 diseases, obtaining 20,824 genes and 880,469 gene pairs in total. For consistency, this set of 2,941 diseases is the same as that used in Section 3.1.3 for the large-scale evaluation of literature retrieval performance. The resulting disease–gene-pair matrix was extremely sparse, with 88.1% zeros, reflecting that most regulatory relationships are not discussed universally but are reported in a context-dependent manner.

Dimensionality reduction of the disease–gene-pair matrix revealed clustering patterns consistent with MeSH disease categories (Figure. 8). Furthermore, a logistic regression classifier predicting disease categories from PCA-derived features achieved 73.6% accuracy (data not shown). I found Misclassifications were more frequent for certain categories (e.g., occupational diseases, which may be less defined by shared mechanisms) (data not shown), indicating that literature-derived networks preserve biologically meaningful disease structure while also reflecting the inherent difficulty of categorical disease definitions.

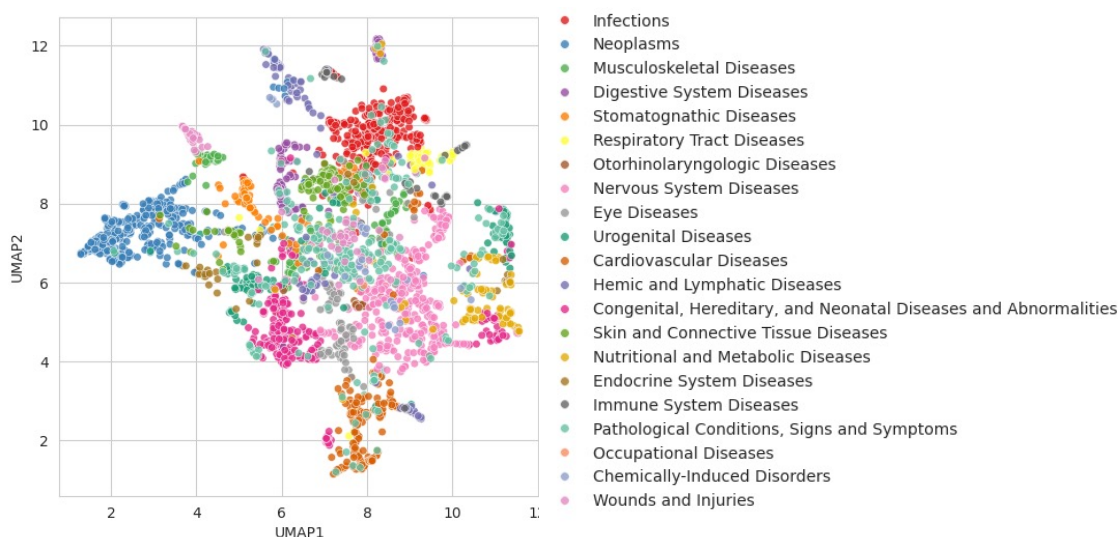


Figure 8. UMAP visualization of disease-gene pair matrix across 2,941 diseases
Each point represents a disease; colors indicate MeSH “C” disease categories.

3.3.4 Zero-shot prediction of drug indications (repurposing)

Based on the observations above, I hypothesized that if the disease state captured by context-dependent GRNs can be encoded as features, then drug indications can be predicted even for diseases with few known indications. For evaluation, I used PrimeKG (Chandak *et al.*, 2023), a knowledge graph comprising 2,054 diseases, 2,074 compounds, and 85,262 relationships (indication / contraindication / off-label). To mimic a realistic repurposing setting, I adopted a zero-shot disease split, in which all information related to test-set diseases is completely excluded from the training data (Figure. 9).

For disease features, I extracted representations from the GRN matrix and compressed them to 200 dimensions using PCA; for drug features, I similarly compressed a literature co-occurrence matrix to 200 dimensions. Using the resulting 400-dimensional concatenated features, I trained a LightGBM classifier and achieved AUPRC = 0.96 under the zero-shot condition, outperforming a previously reported approach (Huang *et al.*, 2024) (Figure. 9). This result supports the view that context-dependent GRNs retain disease-specific mechanistic information in a form that contributes to prediction.

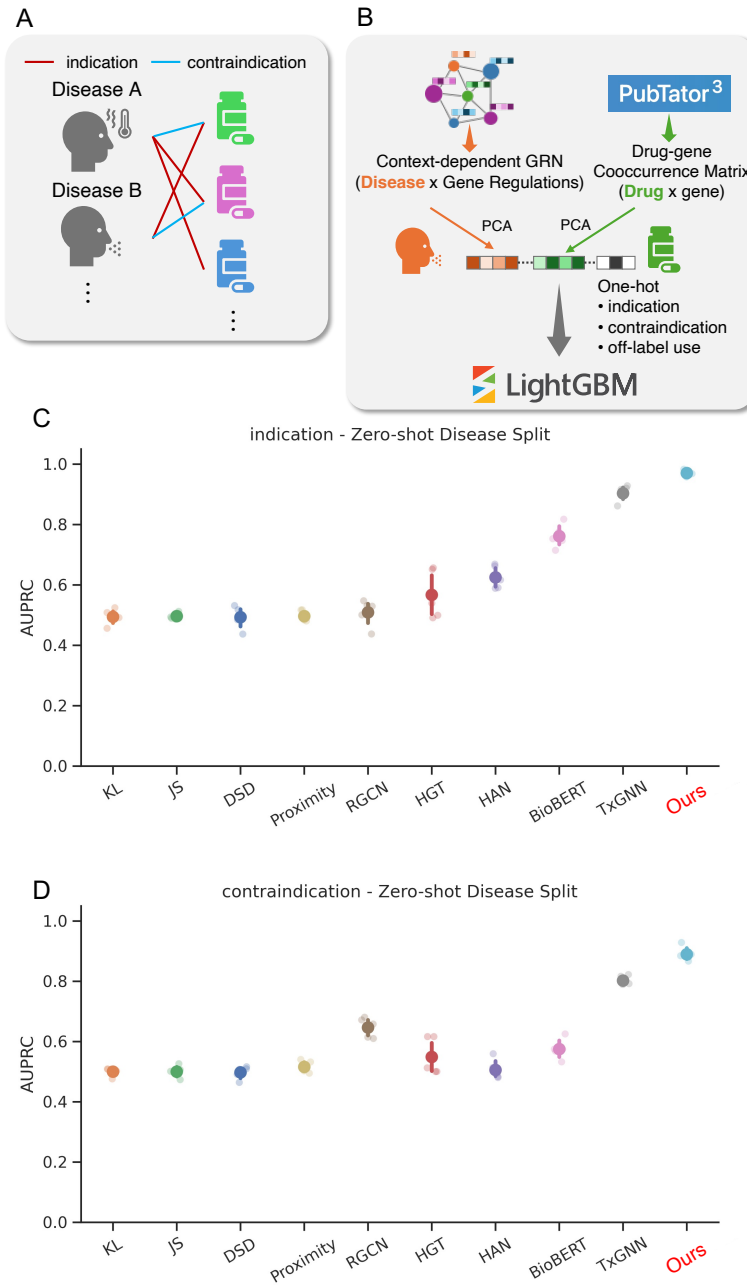


Figure 9. Drug repurposing prediction using context-dependent GRNs

(A) Schematic illustration of the PrimeKG dataset. (B) Overview of feature construction and model training pipeline. Disease features derived from context-dependent GRNs and drug features derived from literature co-occurrence matrices are concatenated and used to train LightGBM classifier.

(C, D) AUPRC comparison for drug indication (C) and contraindication (D) prediction across different models. Our approach demonstrates superior performance in capturing disease-specific mechanisms for both prediction tasks.

3.4. Embeddings derived from context-dependent networks align with spatial clustering of DEGs and reflect biological phenomena

3.4.1 Gene embedding with degree-penalized DeepWalk

As discussed in Section 3.2.2, hub bias is not only an issue for centrality measures but also for network embeddings that convert network structure into feature vectors. In this study, I embedded each gene using a random-walk-based approach; however, in scale-free networks, random walks can be disproportionately influenced by hubs. To address this, I adopted degree-penalized DeepWalk, aiming to suppress hub dominance while preserving local network structure.

3.4.2 DEGs cluster most strongly in the context-dependent GRN for the matched disease context

To evaluate consistency with gene expression data, I used 68 diseases and 2,553 transcriptome datasets from DiSignAtlas (Zhai *et al.*, 2024). Diseases with particularly large numbers of datasets included COVID-19 (140), SLE (111), Influenza (110), Crohn’s disease (103), and Asthma (102). For each disease, I constructed DEG sets under multiple significance thresholds, and quantified DEG coherence in embedding space using the mean pairwise cosine similarity among DEGs (Figure. 10A).

Across analyses, DEG sets showed significantly higher similarity than random gene sets, and more significant DEGs exhibited stronger clustering (Figure. 10B). Importantly, when comparing three conditions for the same DEG set—(i) an unweighted GRN, (ii) a weighted GRN from a different disease context, and (iii) a weighted GRN from the matched disease context—condition (iii) consistently produced the strongest clustering (Figure. 10B; red curve vs. others). In addition, embeddings derived from different disease contexts still outperformed the unweighted baseline. I interpret this pattern as reflecting a denoising effect of weighting. By up-weighting well-supported relations and down-weighting sparsely supported ones, the embeddings are less dominated by context-irrelevant edges and better capture coherent functional modules shared across diseases.

Taken together, these results quantitatively demonstrate that assigning higher weights to literature-frequent, context-consistent relations brings genes that are co-

expressed under the corresponding disease condition (DEGs) closer together in embedding space.

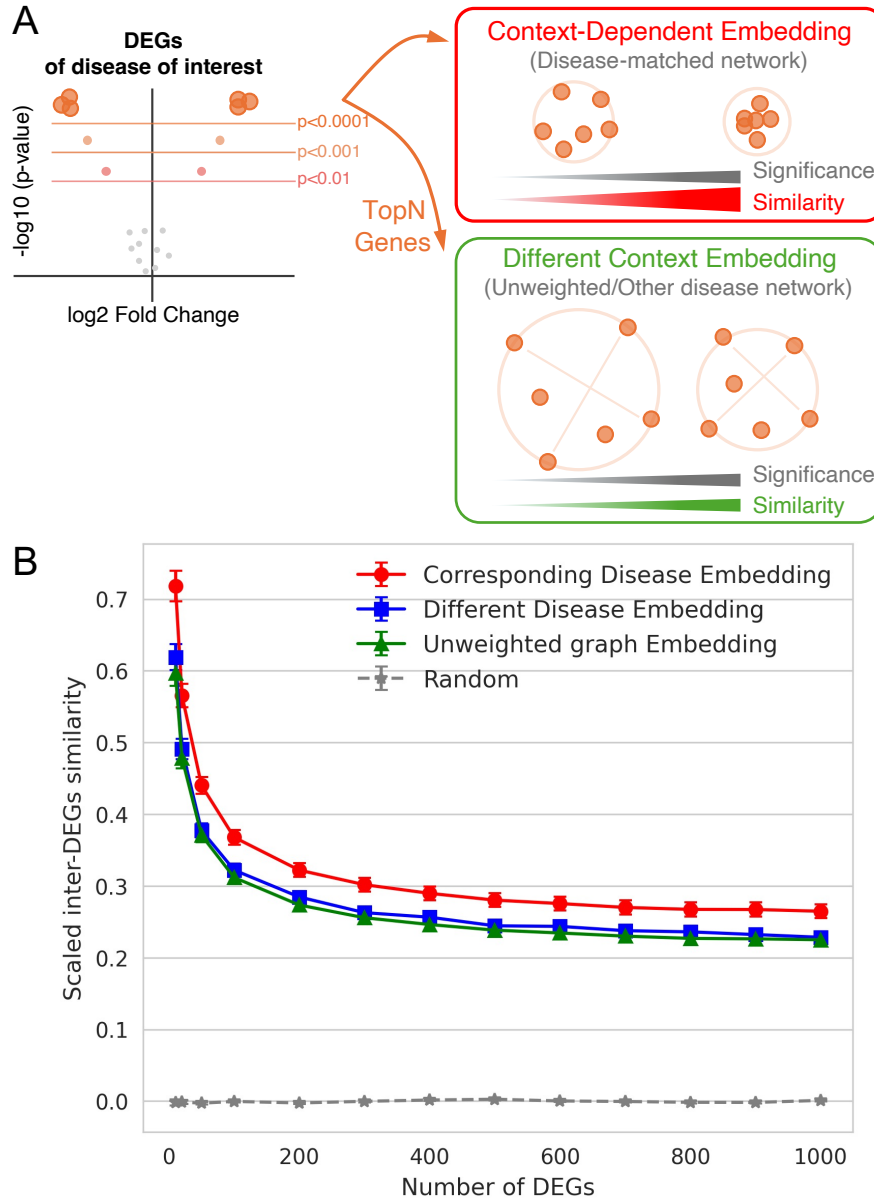


Figure 10. Literature-derived networks correlate with gene expression patterns in embedding space

(A) Schematic of DEG selection from volcano plot and projection into embedding spaces. Top (red box): DEGs clustered tightly in disease-matched context-dependent embedding. Bottom (green box): Same DEGs dispersed in different disease context or unweighted network embedding.

(B) Average pairwise cosine similarity versus the number of top DEGs (mean $\pm$ s.e.m. across datasets). Colors: red, disease-matched context-dependent; blue, different-disease context; green, unweighted; grey dashed, random baseline.

3.5.Context-dependent gene embeddings improve drug–target prediction from gene expression

3.5.1 Comparison with context-independent embeddings using the L1000 dataset

Next, I tested whether context-dependent gene embeddings, used as prior knowledge, improve expression-based drug–target prediction on the L1000 dataset. Following the standard setup in prior work (Chen *et al.*, 2024), I matched compound- and gene-perturbation signatures measured in the same cell line and inferred targets by learning from their expression relationships. Because expression signatures and regulatory programs are cell-line dependent, I hypothesized that gene embeddings derived from cell line–specific context-dependent GRNs would provide more appropriate priors than context-independent embeddings. To isolate the effect of the prior, I kept the model and training protocol identical to the baseline and changed only the gene embeddings (Figure. 11).

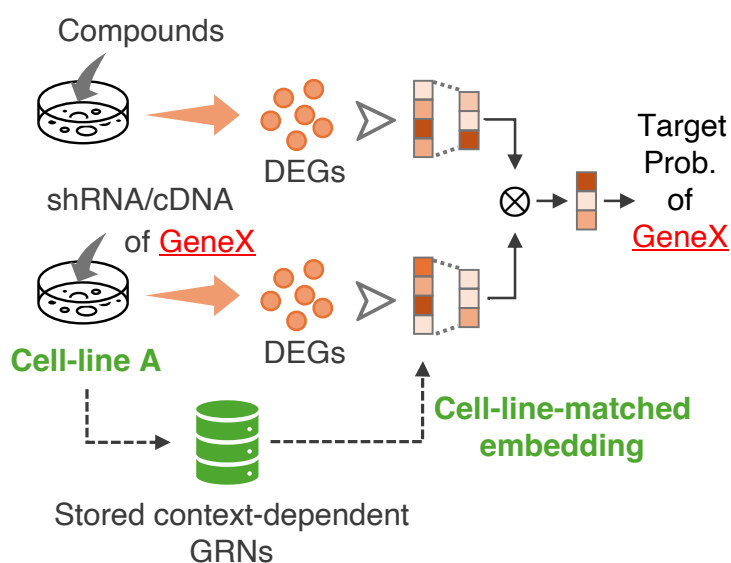


Figure 11. Schematic overview: Using context-dependent GRN-derived priors for drug-target prediction.

3.5.2 Visualization of cell-line context dependency

First, visualization of the networks constructed for individual cell lines showed clustering patterns consistent with tissue-of-origin and tumorigenic status. This indicates that cell-line-specific context from the literature is reflected in the resulting networks (Figure. 12).

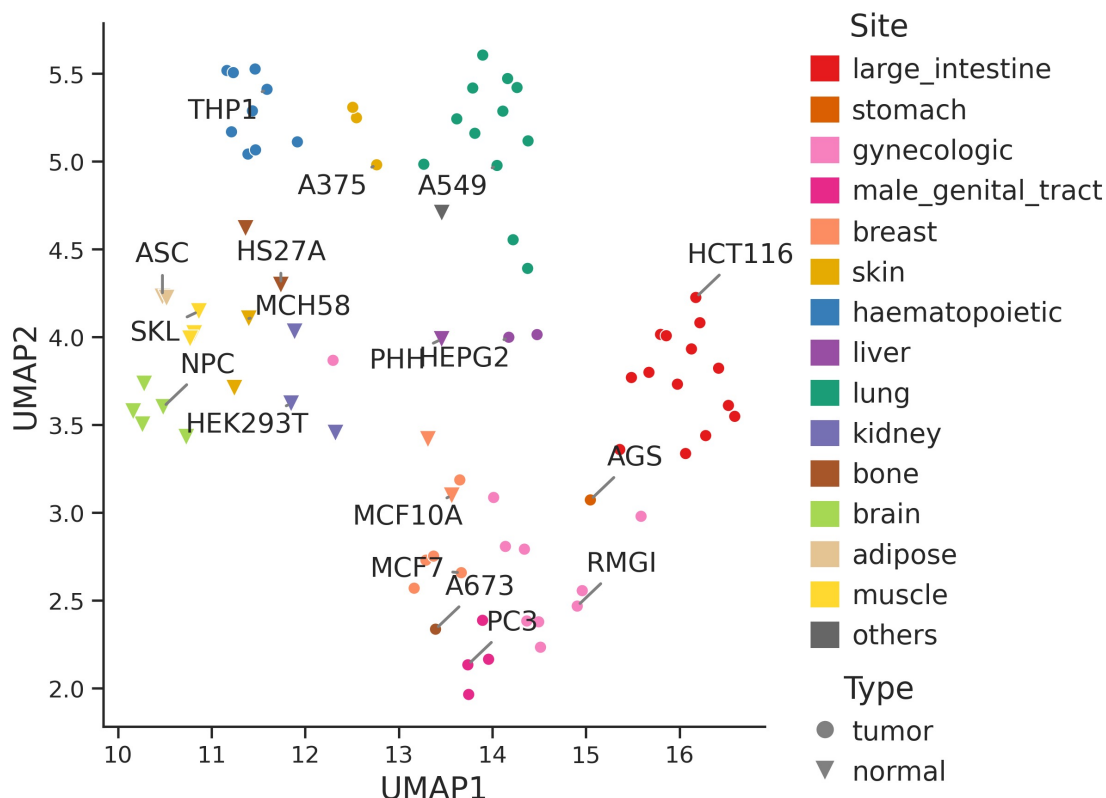


Figure 12. UMAP visualization of cell line-specific GRNs. Colors represent primary site and shapes represent tumor status.

3.5.3 Context-dependent GRNs outperformed existing representations

Using five-fold cross-validation on 83 cell lines and 2,340 compound–gene pairs (1,438 compounds and 499 targets), the context-dependent embeddings achieved $\text{recall@5\%} = 54.6\%$, outperforming the baseline embedding used in the existing method (FRoGS) by approximately five percentage points (Figure. 13). I also included embeddings derived from OmniPath and STRING in the comparison: OmniPath performed similarly to FRoGS, while STRING performed somewhat better (51.9%), possibly because STRING

integrates not only literature evidence but also experimentally supported interactions. Importantly, my approach explicitly introduced context dependence and achieved the best performance among these alternatives, demonstrating that context-dependent embeddings can serve as effective prior knowledge for interpreting expression data and improving drug–target prediction.

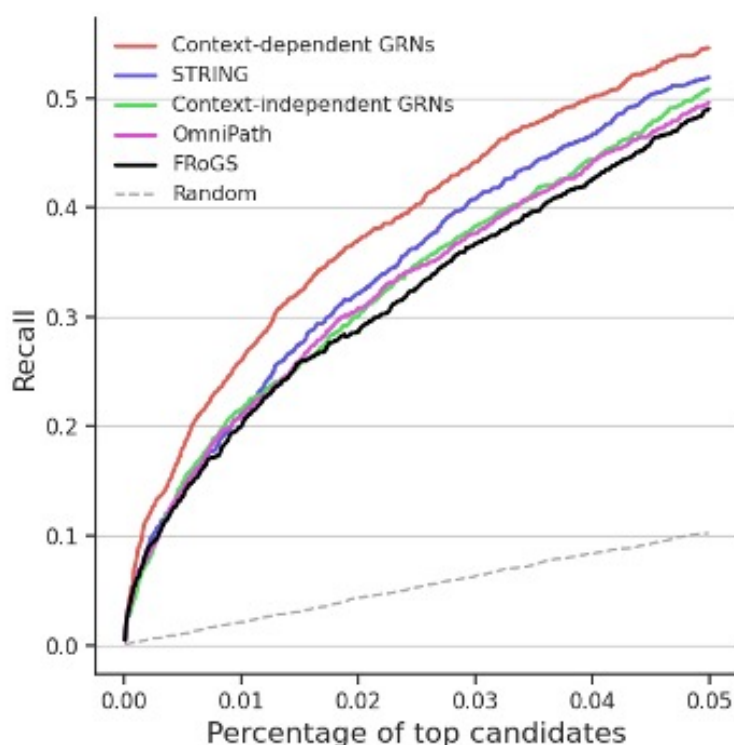


Figure 13. Context-dependent embeddings improve drug-target prediction from gene expression data

Recall@top-k results up to top 5%. Context-dependent GRN (red), context-independent (static) GRN (green), STRING (blue), OmniPath (pink), FROGS (black).

3.6. Automated ODE model construction integrated with an LLM

3.6.1 Connecting literature-derived networks to mechanistic models

While GRNs provide a map of biological relationships, ordinary differential equation (ODE) models are effective for quantitatively explaining mechanisms and predicting dynamics. However, building ODE models is labor-intensive and heavily dependent on

expert judgment, particularly because it requires (i) deciding which set of reaction equations to adopt, (ii) integrating fragmented knowledge into a coherent system, and (iii) ensuring executability (e.g., a closed reaction system, appropriate initial conditions, and conserved quantities).

To address this, I designed an automated pipeline that uses a context-dependent GRN as a navigation layer to select relevant equations from existing model repositories such as BioModels, and then employs LLM assistance to integrate, reconcile, and refine them into an executable mechanistic model (Figure. 14).

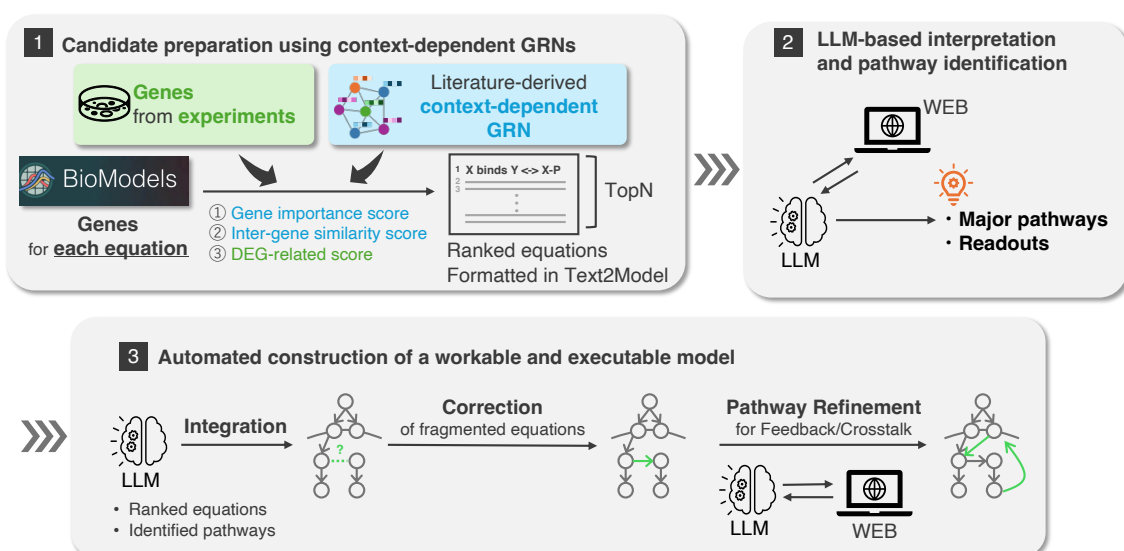


Figure 14. Conceptual overview of automated mathematical model construction workflow

Conceptual workflow for automated mathematical model construction by integrating context-dependent GRNs with large language models (LLMs). The system selects relevant equations from existing databases such as BioModels, followed by integration and correction of fragmented equations to generate functional disease-specific models. Within the LLMs inputs and outputs, the structure of the mathematical model is represented in the Text2Model format (Imoto *et al.*, 2022), a structured, textual representation of biochemical reactions. This enables the LLMs to manipulate its structure and the automatic conversion into an executable format.

3.6.2 Equation selection: integrating literature-based and experimental signals

To prioritize candidate equations (model components), I combined (a) literature-derived signals—such as network centrality and gene–gene similarity in embedding space—and (b) experimental signals—such as similarity to differentially expressed genes (DEGs). Candidates were ranked by a weighted average of these two signals.

For breast cancer–specific model construction, I used the disease query “breast cancer” and, as the experimental signal, DEGs from MCF-7 cells at 2 hours after heregulin stimulation, setting the experimental weight to 10 (See Methods). Figure 15 shows how the ranking changes as the weighting is varied. As the contribution of experimental information increases (from weight 0, reflecting literature signals only), reactions related to the EGF–MEK–ERK pathway and FOS move toward the top of the ranking, whereas reactions associated with TP53 and MYC move downward. This trend is consistent with the interpretation that the 2-hour post-stimulation time point primarily reflects an ERK-driven immediate-early transcriptional response (e.g., the FOS/AP-1 program) (Nakakuki *et al.*, 2010). In other words, increasing the weight of the experimental signal preferentially elevates reactions that align with the transcriptional program actively induced under the specific condition (EGF–MEK–ERK and FOS-related reactions), while pathways that are broadly important in the literature but less prominent under this condition (e.g., TP53- or MYC-mediated stress responses and proliferation control) become relatively deprioritized. This weighting therefore provides an operational mechanism to shift emphasis between “generally important disease factors” and “context-specific dominant factors” under the target experimental setting, enabling equation selection that is better aligned with the condition of interest.

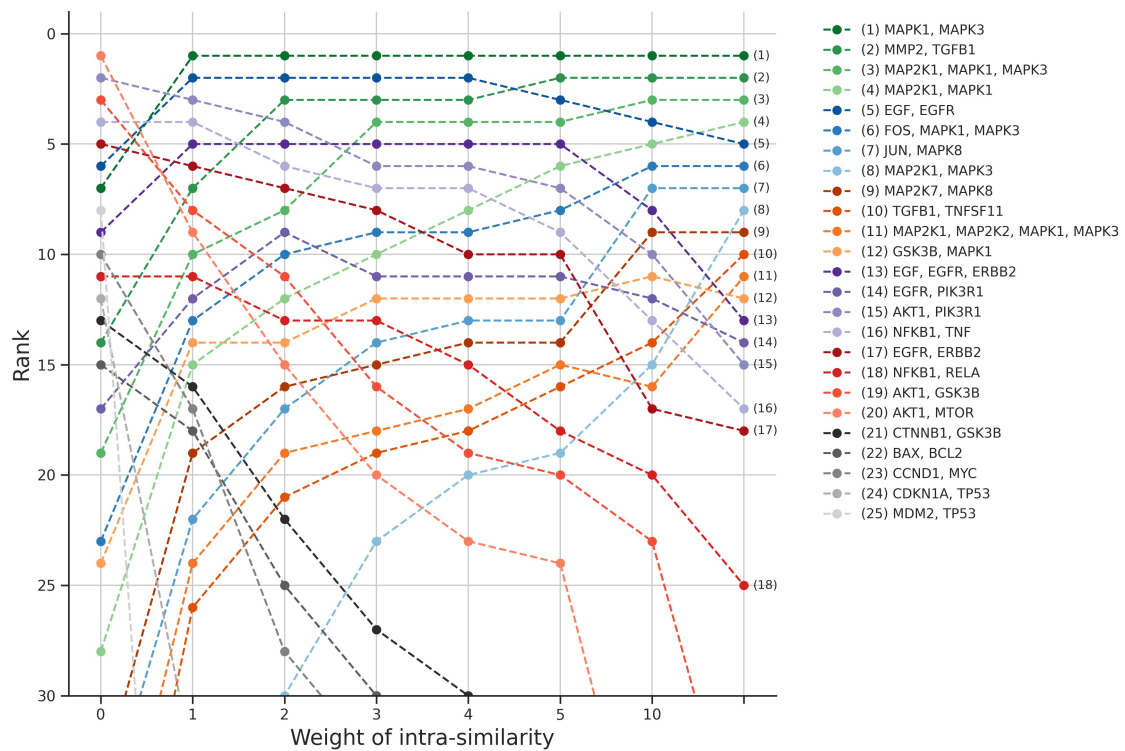


Figure 15. Effect of experimental metric weighting on equation selection

Ranking changes of BioModels equations as a function of experimental metric weight for breast cancer-specific model construction. Top 30 ranked equations (zoomed view). X-axis represents the weight assigned to experimental metrics (similarity to differentially expressed genes), Y-axis shows the ranking of each equation. Different colored lines represent individual equations, labeled by their associated gene sets. The plot demonstrates how balancing literature-derived metrics (centrality and inter-gene similarity) with experimental evidence affects equation prioritization, allowing flexible integration of publication-derived and experimental information.

3.6.3 Case study: ErbB signaling

KEGG enrichment analysis of the DEGs again indicated involvement of the MAPK and PI3K/AKT pathways (Figure. 16A, B). However, this information alone does not directly translate into an ODE model that specifies reaction granularity and explicit readouts. Therefore, I attempted to construct an ODE model starting from the high-priority reactions identified in Section 3.5.2 that were consistent with the experimental data (e.g., top-ranked reactions centered on the ERK–FOS axis). In practice, however, candidate equations were distributed across multiple fragmented model components, and model construction required addressing the following issues:

- 1) Integrating fragments: resolving inconsistencies in inputs and outputs across fragments and connecting them into a single reaction network.
- 2) Reconciling redundancy and gaps: consolidating duplicated representations of the same process and adding missing reactions and molecular species.
- 3) Resolving naming inconsistencies: identifying differences in the notation of reaction species (molecules, complexes, modification states) and harmonizing their identities.
- 4) Removing unnecessary elements: excluding off-target reactions or species to prevent uncontrolled model expansion.
- 5) Explicitly defining and supplementing readouts: explicitly placing readouts that correspond to observed data (e.g., pERK, c-FOS) within the model and supplementing observation equations and intermediate species missing from existing fragments.

Because performing these procedures manually would require a high iterative effort, I developed an LLM-assisted AI agent that automatically identifies candidate connections between fragmented equation sets, selects and supplement's key reaction systems and readouts based on literature evidence, and ultimately reconstructs an executable ODE model (Figure. 14).

The automatically generated model captured the overall structure from the ErbB/HER cascade through MAPK and PI3K/AKT to RSK and FOS, and I confirmed that its architecture was similar to previously reported models (Imoto *et al.*, 2022; Nakakuki *et al.*, 2010) (Figure. 16C).

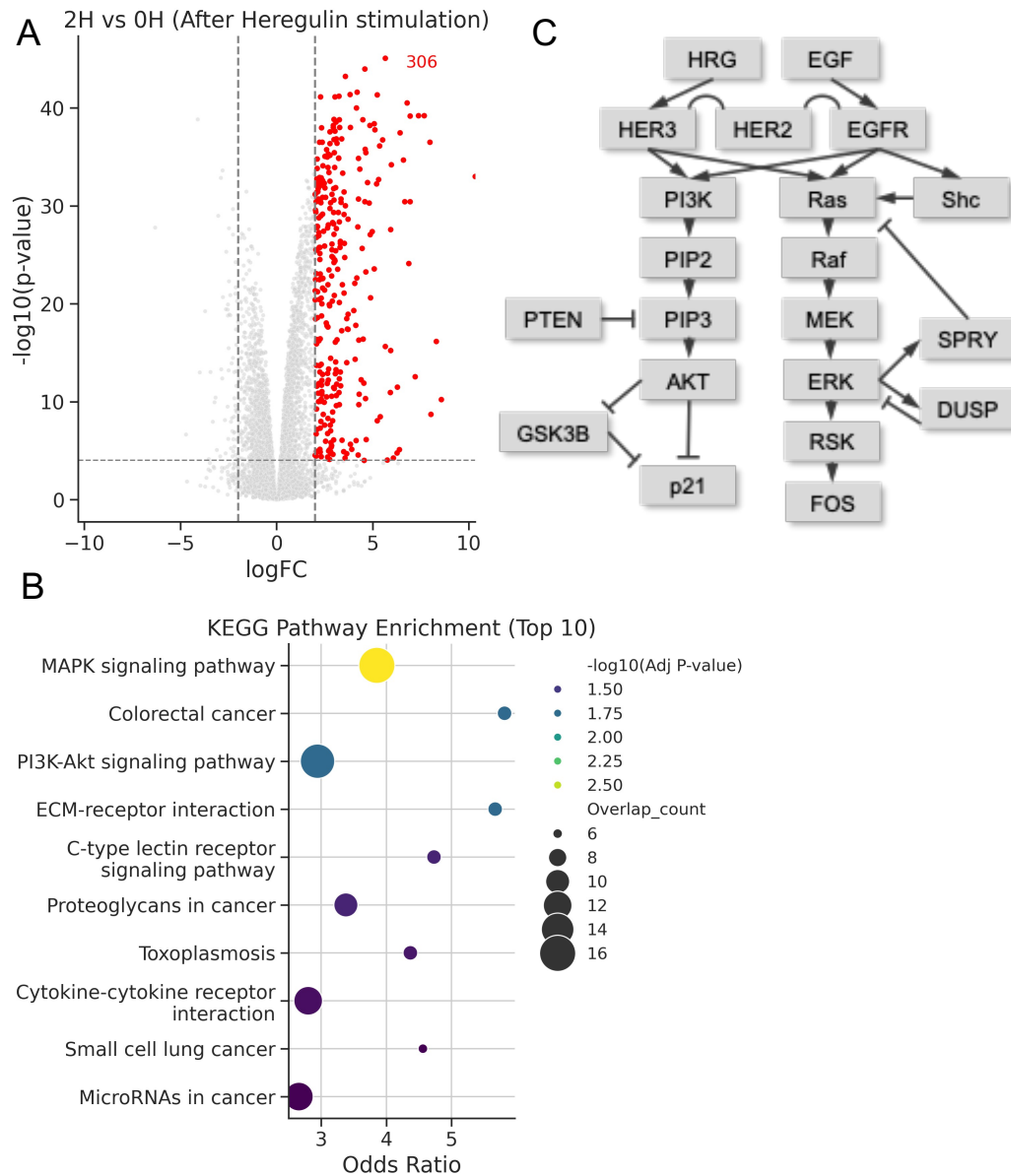


Figure 16. DEG analysis and automatically generated model structure

(A) Volcano plot of DEGs in MCF-7 cells treated with Heregulin for 2 hours. Significantly upregulated genes (red; $\log_{FC} > 2$, $p\text{-value} < 0.0001$) were used for subsequent analysis.

(B) KEGG pathway enrichment analysis results for upregulated DEGs.

(C) Illustration of the automatically generated mathematical model.

3.6.4 Executable but insufficient: Missing FOS-related reactions and limited downstream reproducibility

Although the generated model was executable, it lacked reactions related to FOS production and dephosphorylation; I therefore improved the model by manually adding several missing reactions. After incorporating experimental datasets for phosphorylated HER3 (Nagashima *et al.*, 2007) and for RSK and FOS (Nakakuki *et al.*, 2010) and performing parameter estimation, the model qualitatively reproduced ligand-dependent EGFR–ERK dynamics, whereas reproduction for downstream RSK and FOS dynamics was limited (Figure. 17). I interpret this result as reflecting not only the difficulty of parameter identification, but also the likelihood that incomplete network structure itself constrained downstream reproducibility. Specifically, I concluded that key regulatory structures proposed by Nakakuki et al.—such as a feed-forward AND gate and a FOS-mediated negative feedback loop—were likely not sufficiently recovered in the current automatically generated model.

These results demonstrate both the capabilities and limitations of the framework to automatically generate functional mathematical models. Since the LLM-based model modification framework allows continuous refinement using natural language, it may be possible to further extend our approach to overcome these limitations by providing the LLMs with more context and information on the model structure and simulation results.

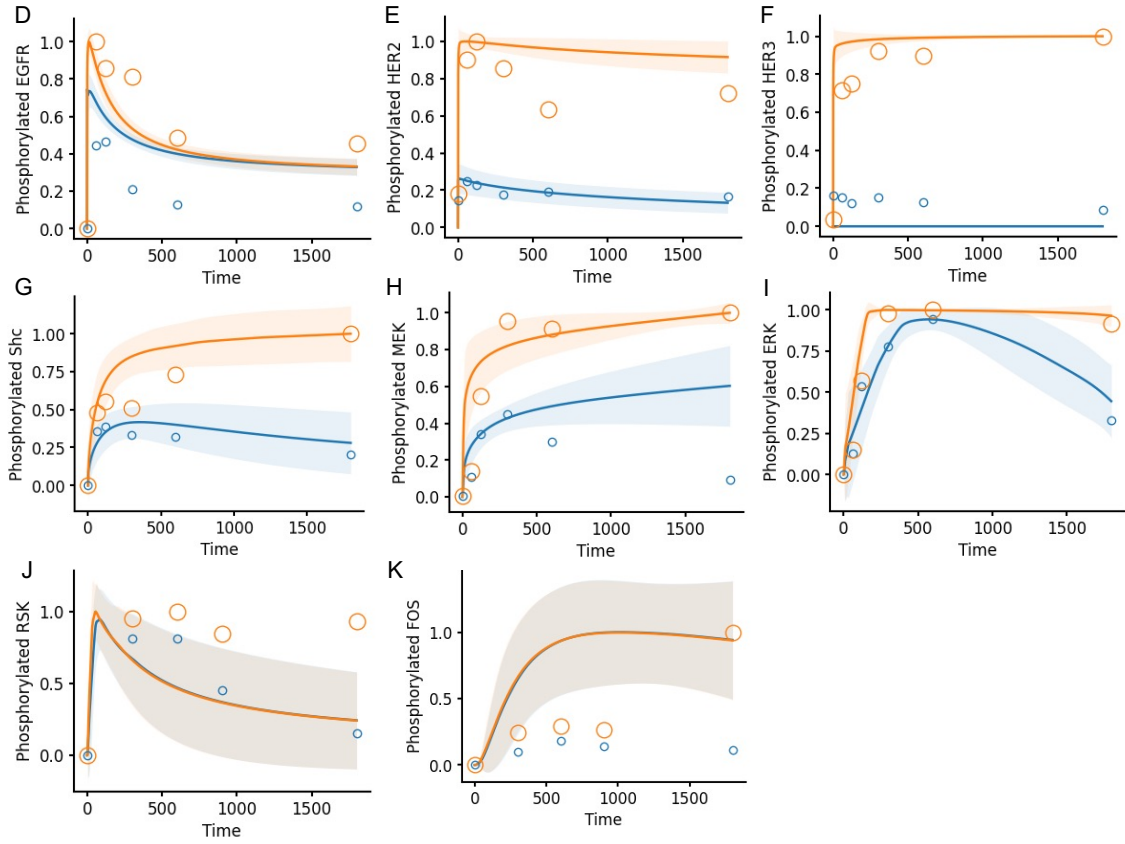


Figure 17. Simulation results using automatically generated Erb model

(D-K) Time-course simulation of phosphorylation dynamics for key signaling components following stimulation with Heregulin (orange) or EGF (blue). Phosphorylation dynamics of EGFR (D), HER2 (E), HER3 (F), Shc (G), MEK (H), ERK (I), RSK (J), and FOS (K) are shown. Circles indicate experimental data points, and solid lines show simulation results averaged across 30 parameter sets. The experimental data were obtained from previous studies and subsequently normalized using the maximum value for each species within the simulated timeframe. The experimental data for HER3, RSK, and FOS were not used for the parameter estimation.

3.7. Chapter Summary

In this chapter, I (1) accurately retrieved context-relevant literature using BERT embeddings; (2) constructed context-dependent GRNs by using the resulting scores as edge weights, thereby recapitulating disease-specific centrality patterns, pathway enrichment, and disease-category structure; (3) demonstrated on large-scale data that DEGs cluster most strongly in the embedding space derived from the matched context; (4) improved performance on a real downstream task, drug-target prediction; and (5) showed that, while LLM integration enables the automatic generation of executable ODE models, structural insufficiencies remain that limit the reproducibility of downstream dynamics.

Together, these results indicate that context-dependent GRNs are not merely an accumulation of literature knowledge. Rather, by mapping literature-derived prior knowledge into a unified representation (networks and embeddings), they provide consistent input information that can be shared across distinct computational biology tasks, including interpretation of expression data, machine-learning-based prediction, and mechanistic modeling. At the same time, the observed limitations in structural revision during automated model construction highlight clear directions for future improvement, such as designing more effective feedback mechanisms for LLM-guided refinement.

4. Discussion

4.1. Contributions of this study

In this study, I present a framework that transforms gene-regulatory knowledge scattered across the life-science literature into weighted gene regulatory networks (GRNs), explicitly assuming that the importance of a regulatory relationship varies with biological context (e.g., disease, cell type, stimulation conditions, and time). While attempts to extract gene–gene regulatory relationships from text and assemble them into networks have been widely pursued, quantitative measures that link each extracted relationship to which contexts it pertains to and how strongly it is supported by evidence have not been sufficiently established. To address this gap, I score the alignment between a given context and each publication, and then normalize this signal by also accounting for factors such as mention frequency. This enables context-dependent weighting of regulatory relationships.

I demonstrate that context-dependent GRNs can serve as a common foundation that connects expression-data interpretation, accurate predictive modeling, and mechanistic model construction. To my knowledge, this is the first study to demonstrate that context-aware weighting—based on semantic query–literature relevance and reporting frequency—can produce prior knowledge that measurably benefits downstream tasks. Moreover, the fact that these analyses share a unified representation (a weighted network and its embeddings) suggests a practical foundation for future literature-driven biological research.

4.2. Biological relevance of literature-derived context-dependent GRNs

This study first quantifies, for a freely described context, how well each publication semantically matches that context as a continuous measure. This allows the framework to handle publications that would be easily missed by strict keyword matching, including paraphrases and variations in experimental conditions.

Importantly, I show that this context-based score is not merely an information-

retrieval metric, but can function as prior knowledge that captures biologically meaningful structure in downstream analyses. Specifically, in embedding spaces derived from context-dependent networks, the corresponding differentially expressed genes (DEGs) form tighter clusters, and this tendency becomes stronger as DEG significance increases. This suggests that reporting biases in the literature are not simply noise; at least in part, they may reflect context-dependent biological variation.

Prior work has discussed study-to-study and gene-level research biases (Stoeger *et al.*, 2018) and proposed heuristics to down-weight frequently mentioned genes (Oba and Nakato, 2024). In contrast, what this thesis emphasizes is that a literature-derived network weighted along the axis of context alignment can be quantitatively validated at disease scale by examining how strongly it corresponds to transcriptomic patterns—and that such validation supports its effectiveness as actionable, context-dependent prior knowledge.

4.3. Recall-oriented retrieval and continuous weights: Controlling influence rather than eliminating noise

In this study, literature retrieval is performed by computing sentence-level similarity and assigning each publication a score via max-over-sentences. This design prioritizes recall (minimizing missed relevant papers), but it can also introduce off-target publications with low similarity, which would be treated as false positives. Rather than focusing on reducing false positives at the retrieval stage, I prioritized assessing how much information from low-similarity publications actually contributes to downstream tasks. Concretely, by using context–literature similarity as a continuous edge weight (rather than a binary decision), the contribution of low-similarity publications can be down-weighted while retaining broad coverage.

This design is particularly practical for downstream tasks that rely on embeddings (e.g., drug-target prediction). If a gene is insufficiently covered in the network and no embedding can be obtained, that gene is excluded from the analyses, which can bias the evaluation itself. Continuous weighting provides a pragmatic compromise: it keeps low-evidence genes in the network while suppressing their influence.

At the same time, evaluation of retrieval accuracy still faces the challenge of defining “ground-truth” labels. I used MeSH assignment as an approximate gold standard for

relevance, but publications that are highly relevant yet lack MeSH annotation would be counted as false negatives. Therefore, future work should move beyond treating a single MeSH-based binary label as an absolute ground truth, expand to evaluation designs that combine multiple evidence sources, and introduce score calibration that accounts for differences across contexts.

4.4.Implementation bottleneck: Multi-context queries and re-computational cost

Because the workflow in this study constructs a large GRN for each context and recomputes gene embeddings, computational cost becomes dominant as the number of contexts increases. While this may be acceptable for small analyses in research settings, it limits practicality for exploratory use cases that require comparing many contexts (e.g., cross-disease comparisons or iterative model construction).

A key future direction is to move away from fully precomputing everything and instead shift toward functions that dynamically adjust network structure or vector representations at query time, localizing re-computation. For example, possible implementation strategies include: (i) maintaining a base “generic network/embedding” and rapidly redistributing weights according to context, (ii) approximating updates via localized re-computation restricted to top-ranked publications or edges, and (iii) reducing update magnitude through caching of past queries and meta-learning across similar contexts.

4.5.Effectiveness of context-dependent knowledge and the challenge of integrating experimentally curated networks

In the drug-target prediction benchmark, context-dependent embeddings improved recall compared with context-independent approaches, demonstrating the value of incorporating contextual information—such as cell type—into prior knowledge. The key point is that, because gene expression changes and drug responses are inherently context dependent, it is reasonable for the prior knowledge used to interpret them to be context conditioned as well.

At the same time, PPI resources that integrate large-scale experimentally supported interactions, such as STRING, achieved high recall even without explicit context modeling. I interpret this mainly as a consequence of scale: the number of experimentally supported interactions is extremely large and substantially exceeds the number of literature-derived regulatory edges extracted in this study (by roughly 10-fold). Therefore, an important next challenge is to determine at what granularity we should best fuse (i) literature-derived knowledge, which is context-aware and interpretable, with (ii) experimentally integrated networks, which are highly comprehensive and often more reliable. A key advantage of context-dependent GRNs is that they provide query-conditioned prior knowledge, which is difficult to derive from experimentally aggregated networks alone. The next step is to integrate the two in a principled manner that respects their complementary roles.

4.6. Automated ODE model construction anchored by context-dependent GRNs and remaining challenges

In this study, I implemented a pipeline that automatically collects, normalizes, and integrates reaction equations and model fragments scattered across the literature and model repositories by combining semantic retrieval with LLM-based assistance, ultimately generating an executable ODE model for a target context. A major advantage of this approach is that it turns a sequence of steps that previously depended heavily on manual expert effort into a semi-automated procedure. In addition, using a context-dependent GRN as an anchor for reaction retrieval allows the system to avoid unconstrained search and to focus exploration on high-priority candidates.

However, from the perspective of reproducing experimental dynamics, I also found that structural insufficiencies can remain. Specifically, in analyses of phosphorylation dynamics following EGF/herregulin stimulation in MCF-7 cells, the generated model could explain part of the major behavior, but it had limited ability to reproduce downstream responses in a fully consistent manner, suggesting that missing reaction-network structure may be constraining reproducibility. This limitation is not easily resolved by parameter estimation alone; rather, it indicates the need to use mismatches with observed data as clues to identify missing mechanisms and to update the model

structure itself.

Accordingly, an important future direction is to incorporate a “closed loop” that iteratively interprets data, diagnoses sources of mismatch, generates hypotheses for missing reactions or regulatory structures, adds them to the model, re-estimates parameters, and re-validates. That said, the primary aim of this study is to demonstrate that context-dependent GRNs can serve as anchors for integrating heterogeneous knowledge sources and for automatically constructing executable models from fragmented information. Closed-loop optimization for high-fidelity dynamic reproduction and comprehensive robustness assessment remain important topics for future work.

4.7. Summary: Context-dependent GRNs as a promising paradigm for connecting literature knowledge with experiments and models

In summary, this study shows that gene-regulatory knowledge extracted from the literature can be weighted in a context-dependent manner and consistently connected to experimental patterns, predictive tasks, and mechanistic model construction through LLM integration.

At the same time, several issues remain essential for maturing the framework into a robust research infrastructure, including limitations of evaluation labels (MeSH), publication bias and hub effects, computational cost, and iterative structural refinement of mechanistic models using GRN-guided knowledge. By addressing these challenges, context-dependent GRNs could evolve into a flexible prior-knowledge infrastructure that bridges vast literature information to precision medicine and computational biology, enabling broader applications.

5. References

- Arnulf,J.K. *et al.* (2014) Predicting Survey Responses: How and Why Semantics Shape Survey Statistics on Organizational Behaviour. *PLoS ONE*, **9**, e106361.
- Birtwistle,M.R. *et al.* (2007) Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Mol. Syst. Biol.*, **3**, 144.
- Brin,S. and Page,L. (1998) The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.*, **30**, 107–117.
- Chandak,P. *et al.* (2023) Building a knowledge graph to enable precision medicine. *Sci. Data*, **10**, 67.
- Chen,H. *et al.* (2024) Drug target prediction through deep learning functional representation of gene signatures. *Nat. Commun.*, **15**, 1853.
- Chen,Q. *et al.* (2020) BioConceptVec: Creating and evaluating literature-based biomedical concept embeddings on a large scale. *PLOS Comput. Biol.*, **16**, e1007617.
- Deka,P. *et al.* (2022) Improved methods to aid unsupervised evidence-based fact checking for online health news. *J. Data Intell.*, **3**, 474–505.
- Feng,R. *et al.* (2018) Representation learning for scale-free networks. In, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI’18/IAAI’18/EAAI’18. AAAI Press, New Orleans, Louisiana, USA, pp. 282–289.
- Fröhlich,F. *et al.* (2018) Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model. *Cell Syst.*, **7**, 567-579.e6.
- Gill,J.K. *et al.* (2024) Large language model based framework for automated extraction of genetic interactions from unstructured data. *PLOS ONE*, **19**, e0303231.
- Glont,M. *et al.* (2020) BioModels Parameters: a treasure trove of parameter values from published systems biology models. *Bioinformatics*, **36**, 4649–4654.

- Grivell,L. (2002) Mining the bibliome: searching for a needle in a haystack? *EMBO Rep.*, **3**, 200–203.
- Grover,A. and Leskovec,J. (2016) node2vec: Scalable Feature Learning for Networks. In, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 855–864.
- Gu,Y. *et al.* (2021) Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans Comput Healthc.*, **3**, 2:1-2:23.
- Hartman,E. *et al.* (2023) Interpreting biologically informed neural networks for enhanced proteomic biomarker discovery and pathway analysis. *Nat. Commun.*, **14**, 5359.
- Hill,S.M. *et al.* (2017) Context Specificity in Causal Signaling Networks Revealed by Phosphoprotein Profiling. *Cell Syst.*, **4**, 73-83.e10.
- Huang,K. *et al.* (2024) A foundation model for clinician-centered drug repurposing. *Nat. Med.*, **30**, 3601–3613.
- Imoto,H. *et al.* (2020) A Computational Framework for Prediction and Analysis of Cancer Signaling Dynamics from RNA Sequencing Data—Application to the ErbB Receptor Signaling Pathway. *Cancers*, **12**, 2878.
- Imoto,H. *et al.* (2022) A text-based computational framework for patient -specific modeling for classification of cancers. *iScience*, **25**, 103944.
- Ke,G. *et al.* (2017) LightGBM: a highly efficient gradient boosting decision tree. In, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17. Curran Associates Inc., Red Hook, NY, USA, pp. 3149–3157.
- Lai,P.-T. *et al.* (2023) BioREx: Improving biomedical relation extraction by leveraging heterogeneous datasets. *J. Biomed. Inform.*, **146**, 104487.
- Lee,J. *et al.* (2020) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, **36**, 1234–1240.
- Nagasato-Ichikawa,A. *et al.* (2024) ErbB2/HER2 governs CDK4 inhibitor sensitivity and timing and irreversibility of G1/S transition by altering c-Myc and cyclin D function.
- Nagashima,T. *et al.* (2007) Quantitative Transcriptional Control of ErbB Receptor Signaling Undergoes Graded to Biphasic Response for Cell Differentiation*. *J.*

- Biol. Chem.*, **282**, 4045–4056.
- Nakakuki,T. *et al.* (2010) Ligand-Specific c-Fos Expression Emerges from the Spatiotemporal Control of ErbB Network Dynamics. *Cell*, **141**, 884–896.
- National Library of Medicine (2025a) National Library of Medicine. Medical Subject Headings – Descriptor Data, 2025 edition. https://nlmpubs.nlm.nih.gov/projects/mesh/MESH_FILES/xmlmesh/.
- National Library of Medicine (2025b) PubMed overview. <https://pubmed.ncbi.nlm.nih.gov/about/>.
- Oba,G.M. and Nakato,R. (2024) Clover: An unbiased method for prioritizing differentially expressed genes using a data-driven approach. *Genes Cells*, **29**, 456–470.
- Perozzi,B. *et al.* (2014) DeepWalk: Online Learning of Social Representations., pp. 701–710.
- Reimers,N. and Gurevych,I. (2020) Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In, Webber,B. *et al.* (eds), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, pp. 4512–4525.
- Reimers,N. and Gurevych,I. (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In, Inui,K. *et al.* (eds), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, pp. 3982–3992.
- Rodriguez-Mier,P. *et al.* (2025) Unifying multi-sample network inference from prior knowledge and omics data with CORNETO. *Nat. Mach. Intell.*, **7**, 1168–1186.
- Rogers,F.B. (1963) Communications to the Editor. *Bull. Med. Libr. Assoc.*, **51**, 114–116.
- Ruscione,M. *et al.* (2025) NeKo: A tool for automatic network construction from prior knowledge. *PLOS Comput. Biol.*, **21**, e1013300.
- Stoeger,T. *et al.* (2018) Large-scale investigation of the reasons why potentially important genes are ignored. *PLOS Biol.*, **16**, e2006643.
- Subramanian,A. *et al.* (2017) A Next Generation Connectivity Map: L1000 Platform and

- the First 1,000,000 Profiles. *Cell*, **171**, 1437-1452.e17.
- Subramanian,A. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.*, **102**, 15545–15550.
- Szklarczyk,D. *et al.* (2023) The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.*, **51**, D638–D646.
- The Gene Ontology Consortium *et al.* (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.*, **49**, D325–D334.
- Türei,D. *et al.* (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods*, **13**, 966–967.
- Wang,J. and Duan,Z. (2025) Empirical Research on Utilizing LLM-based Agents for Automated Bug Fixing via LangGraph.
- Wei,C.-H. *et al.* (2024) PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Res.*, **52**, W540–W546.
- Zhai,Z. *et al.* (2024) DiSignAtlas: an atlas of human and mouse disease signatures based on bulk and single-cell transcriptomics. *Nucleic Acids Res.*, **52**, D1236–D1245.

Acknowledgments

I would like to begin by expressing my deepest gratitude to my supervisor, Dr. Mariko Okada, for giving me the opportunity to pursue this research. Throughout my doctoral study, Dr. Mariko Okada consistently provided thoughtful guidance and insightful feedback from a broad perspective. In particular, I greatly appreciated Dr. Mariko Okada's advice on research direction and the encouragement to connect ideas across different fields. Conducting research with a mindset of bridging disciplines was both enjoyable and highly rewarding, and it shaped the way I think as a scientist.

I would also like to sincerely thank all members of our laboratory for their support and for creating an environment where I could focus on my research. I am especially grateful to Dr. Keita Iida and Mr. Kiwamu Arakane for the countless discussions we shared in our daily work. I also thank Dr. Ayaka Ichikawa for kindly providing the gene expression data that contributed to this study. In addition, I would like to express my appreciation to our former member, Dr. Hiroaki Imoto, whose insightful comments and research perspectives gave me valuable direction at important moments.

My sincere thanks also go to my close friends and my colleagues at work. Your kindness, understanding, and encouragement gave me peace of mind and the time and space to immerse myself in research wholeheartedly. I am grateful for the many moments of support—both spoken and unspoken—that helped me continue during challenging periods.

Finally, I would like to thank my family—my mother and my brother—and my father in heaven. Thank you from the bottom of my heart for raising me, believing in me, and supporting me in every step of my journey. I am truly grateful for everything you have done for me.