OUKA
The University of Osaka
Institutional Knowledge Archive

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

Form 3

Abstract of Thesis

Name （ 筒井真人 ）	
Title	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)

Abstract of Thesis

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—which in this study I focus on as ordinary differential equation (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. Nevertheless, there has been little work on defining context dependence in a comprehensive and quantitative manner, or on demonstrating how such context-aware knowledge can be systematically used in downstream analyses. In response, this study focuses on maximizing the use of the vast and continuously growing biomedical literature, and investigates practical ways to define context dependence and to evaluate it quantitatively at scale.

In this study, I propose a framework that assigns quantitative, context-dependent weights to gene regulations extracted from the literature. Concretely, I first collect candidate regulatory statements (e.g., “gene A activates gene B” or “protein X inhibits pathway Y”) from scientific articles and record their provenance. I then represent a target biological context—such as a disease, cell type, and/or experimental condition—as a textual query. Using pretrained biomedical language models, I embed both the query and the relevant literature describing each gene regulation into a shared semantic space, and compute their similarity (e.g., cosine similarity). This similarity score is used as a continuous weight, so that regulations supported by literature semantically close to the target context are up-weighted, while those supported by distant contexts are down-weighted. In this way, the resulting GRN becomes not a single fixed graph, but a context-adapted, weighted network tailored to the user-specified query.

To make this context-specific GRN readily usable by computers and downstream statistical or machine-learning methods, I transform the network into compact numerical representations, referred to as “embeddings”. Specifically, I apply network embedding to the weighted GRN to produce node embeddings, which serve as context-dependent prior knowledge for downstream analyses.

I validated the proposed approach in two large-scale settings. I first show 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. Interestingly, I discovered that the signals introduced by context dependence and by literature reporting frequency are quantitatively associated with gene-expression–based statistics observed in large-scale transcriptome data. This constitutes a novel finding and indicates that context-aware, literature-derived regulatory evidence can capture biologically relevant structure. Furthermore, by verifying accuracy improvements across multiple machine-learning tasks using the context-dependent GRN, I provide quantitative support that the networks produced by my framework encode meaningful biological information.

Finally, 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.

様式 7

論文審査の結果の要旨及び担当者

氏　名	（　筒井 真人 ）	
	（職）	氏　名
論文審査担当者	主 査　教 授	岡田　眞里子
	副 査　教 授	水口　賢司
	副 査　教 授	上田　昌宏

論文審査の結果の要旨

本論文は、生命科学文献に散在する遺伝子間の相互作用に関する情報を、自然言語処理の手法を用いて、生物学的コンテクスト（疾患・細胞種など）に応じて抽出可能な事前知識として変換し、遺伝子ネットワークや薬剤応答予測などの下流解析へ接続する計算的な枠組みを提案・検証したものである。特に、同じ遺伝子間相互作用であっても、コンテクストに応じて遺伝子の重要度が文献情報により変わり得る点を、ネットワーク上の重みとして表現し、下流解析に取り込み、その精度を向上させた点に特徴がある。

具体的には、ヒト疾患を例とした解析では、疾患名をコンテキストに利用することで、網羅的遺伝子発現データなどのオミクス実験データの解釈性や薬剤標的の予測精度を向上させることが可能であった。また、乳がんのシグナル伝達系の例においては、大規模言語モデルとの併用により、実行可能な数理モデルを半自動で構築し、シミュレーションすることが可能になった。

審査では、提案枠組みの新規性と評価設計の妥当性に加えて、「文献情報から制御情報が正しく抽出可能か」と「疾患ネットワークの新規性」という観点を中心に検討した。制御構造に関しては、現状として問題点はあるものの、クエリーの工夫や抜き出したい情報に対する言語処理によって向上可能なことが議論された。一方で、得られた重み付きネットワークを用いて、差異発現遺伝子のまとまりや薬剤標的予測の性能向上を示したことから、解きたい問題によっては、本手法が有効であると考えられた。疾患ネットワークの新規性に関しては、既存のデータベース等を用いることの限界が議論された。一方で、コンテクスト依存ネットワークを手掛かりに、実行可能な数理モデルを自動生成する枠組みが示されたことから、データ解析の上でも、今後の生命科学への応用は期待できるとの意見がなされた。

以上を踏まえると、本論文の主要な貢献は、文献由来の知識をコンテクストに応じて連続的に重み付け可能な形へ整理し、その妥当性評価と応用までを一貫した設計として提示した点にある。これは、従来十分に評価されてこなかった文献情報といった生物データの領域に実証的基盤を与えるものとして評価できる。また、今後の発展や応用展開も十分期待できるものである。これらの議論の結果、本論文は、方法論の提案に加えて、ヒト疾患での検証と応用可能性を示しており、博士論文として求められる独創性と研究的貢献を備えていると判断した。